Patrick Frank
Earth and Space Science Manuscript 2017EA000256
Response to Review #5

The response begins with a comment. Reviewer anonymity is in place, in part, for the good reason of removing personality from the process. Ethical withdrawal of anonymity requires prior mutual agreement. The unagreed nullification of reviewer anonymity was not welcome.

Summary Response:

1. The reviewer mistakes manuscript eqn. 6 to concern climate physics rather than the observed behavior of climate models; items 1.1, 1.2.1, 4.4, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, and 5.9.4. These evidence a misinterpretation of the entire manuscript analysis.
2. The reviewer's recommended major revisions are misconceived and, if followed, would leave nothing publishable; items 1.6, 1.8, 4.1, 5.7, 5.9.2, 7 (all) and 8 (all).
3. The reviewer asked for a more extensive analysis than in [*P. Frank*, 2008], despite the much more extensive analysis provided in the manuscript; item 5.9.4.
4. A new description of the origin and meaning of the cloud error from [*Lauer and Hamilton*, 2013] is provided in Sections 2.3 and 2.4.1. The derivational logic is also now provided in new SI Section 6.2.
5. The reviewer incorrectly supposed an author assumption that the satellite datasets are perfect and without any systematic biases; item 6.3.
6. The choice of the long wave cloud forcing error of [*Lauer and Hamilton*, 2013] is clear, not obscure; items 7.3.1.1 through 7.3.2.3.2.
7. The reviewer misunderstood the significance of [*Jiang et al.*, 2012]; items 7.4.2 and 7.4.3.
8. The reviewer is mistaken about how uncertainty enters an anomaly; item 7.4.4.
9. The reviewer misunderstood how systematic error is propagated and what is being propagated; items 7.5, 7.6.1, 7.6.2, and 7.7.
10. The reviewer is mistaken that the manuscript uncertainties derive from erroneous air temperatures; item 7.7.
11. The reviewer's supposition that the manuscript error propagation is invalid lacks analytical rigor and is wrong; items 7.4.1 through 7.7, see especially 7.4.5, 7.5, and 7.6.1.
12. In view of the above, the recommendation that the manuscript conclusions be set aside is rendered without force.

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

1. *Overview:*

1.1 *The author attempts to estimate a lower bound for the uncertainty error bars associated with the global temperature projections of current Global Climate Models (GCMs). To do this, he first presents a simple, semi-empirical analytical model for the global temperature response to greenhouse gas concentrations.*

> 1.1 The opening statement is not correct. Manuscript Equation 6 has nothing whatever to do with the global temperature response to greenhouse gas concentrations. Rather, it emulates the behavior of climate models. The reviewer has mistaken a critically central point right at the start.

This focus on model behavior as opposed to climate is given immediately in Abstract sentence two. Introduction lines 124-127 and 131-132 explicate that the focus is on model behavior, and not on climate physics.

Introduction lines 141-144 emphasize this distinction through use of italics. Lines 150-152 make the point again.

Section 2.2, line 313 introduced eqn. 6 as, "*a simple emulation model of how modern GCMs project the impact of increasing greenhouse gases on global averaged air temperatures.*" (underline added)

Line 324: "*Eqn. 6 is a surmise that GCMs project the GSASAT as a linear extrapolation of fractional wve GHG forcing*."

However, despite repetition, the reviewer apparently missed this critical point. Eqn. 6 is about climate models. It is most definitely not about climate or, "*the global temperature response to greenhouse gas concentrations*."

1.2 *According to this model, global temperature trends are linearly related to greenhouse gas concentrations.*

1.2.1 Again, the reviewer is not correct.

The model shows that GCMs project temperature as a linear function of GHG forcing. It conveys nothing about the relation of global temperatures and GHG concentrations.

1.2.2 Note also that forcing is a log function of GHG concentration, not a linear function as the reviewer has it.

1.3 *The author shows that the output from this simple linear model is a reasonable approximation of the global temperature projections of current GCMs. He then uses this observation to argue that the appropriate framework for determining the error bars is to use "linear propagation of errors".*

1.3 The analysis shows that, to the extent GCMs use standard forcings, manuscript eqn. 6 exactly emulates their projections of global air temperatures.

Linear propagation of error follows rigorously from linear combinations of results [*Bevington and Robinson*, 2003; *Taylor and Kuyatt.*, 1994; *Vasquez and Whiting*, 2006].

1.4 *Specifically, he argues that the uncertainty errors for each year of the projection should be added to the uncertainty errors of the preceding years. This approach leads to a fairly rapid increase in the magnitude of the error bars.*

1.4.1 There is no such thing as, "uncertainty errors." There are uncertainties, which are statistical measures of reliability. There are errors, which typically are the physically meaningful difference magnitudes between predictions and observables, or between measurements and calibration standards.

1.4.2 In any case, the uncertainties are not added, but combined in quadrature, according to eqns. 1 & 2.

1.5 *He then estimates a lower bound for the uncertainty associated with a single year from reviewing some of the literature assessing the modelling of clouds in GCMs. When this lower bound is plugged into his error propagation approach, the error bars rapidly increase to encompass a range of [plus minus]15[degree sign]C by the end of a century. This proposed uncertainty range is so large that it would make all the global temperature projections of current GCMs effectively worthless. The author made a similar argument in an earlier paper which he references, i.e., Frank, 2008.*

1.5 The author generally agrees with this conclusion, with the exception that the "error bars" are actually uncertainty bars.

1.6 *In my opinion, his use of the "linear propagation of errors" is inappropriate and invalid. His proposed error bars are unrealistically large and lead to incorrect conclusions.*

1.6 As shown below, the reviewer's opinion is analytically unfounded.

1.7 *Having said that, I do believe that there are several good points and arguments made in the paper which should be of interest and relevance for the climate modelling community as well as those using the results of climate models - e.g., the distinction between accuracy & precision; the fact that climate model uncertainties are typically described in terms of intra- and inter-model comparisons; the fact that there has been relatively little effort made into quantifying climate model uncertainties in terms of observation-based measurements; the differences between modelled and observed clouds.*

1.7 The author thanks the reviewer for agreeing these points are worth communicating.

1.8 If the author were to drop his use of "the linear propagation of errors" and restructure the article to focus on these more justifiable points and arguments, I believe there is potentially a paper which could be important and worthy of publication. However, as it stands, with the focus on these unrealistically large and inappropriate error bars, the paper is not publishable.

1.8 The uncertainties are neither unrealistic nor inappropriate, as shown below. The reviewer's suggestion, to remove the entire error analysis, would effectively obliterate the work.

1.9 *I appreciate that the author has been promoting this argument about using "the linear propagation of errors" for GCM projections since at least Frank, 2008. So, he might not be prepared to change tack, as I suggest. However, if he does, I think with some restructuring and refocusing of the material, and some additional discussion of the literature on comparisons between models and observations (such as the references included in this review), there could be enough material for a good paper.*

With that in mind, I recommend the article should be "returned to the author for major revisions".

1.9 The author appreciates the reviewer's sensitivity to past work. However, it is important to observe that the concerns expressed to this point are the reviewer's subjective opinions only. None of them is thus far objectively grounded.

Below are some more detailed comments on each of the sections in the paper.

The remainder of the review is lengthy and often lacks analytical focus. Therefore salient points are extracted in italics and author responses are provided.

2. *Comments on citation style*

2.1 *Throughout the paper there are several inconsistencies in how references are cited. ...*

2.1 The citation style is invariably AGU standard, formatted within and by Endnote.

2.2 *In the references section itself (Section 5), there are also inconsistencies,...*

2.2 Thank-you. All journal titles are now properly abbreviated.

2.3 *I would include urls for some of the references which don't have a doi, e.g., Frank, 2008 and Haynes, 2012.*

2.3 The AGU style guide does not include a URL field for journal articles.

3. *Comments on the Table of terms or acronyms*

3.1 *The author repeatedly assigns the term "general circulation (climate) model" to the acronym "GCM" throughout the paper, including in the glossary. This used to be a common usage, but it is now a rather outdated assignation, as the GCM acronym is now usually given for the term "Global Climate Model".*

3.1 According to the American Meteorological Society "**Standard technical abbreviations and acronyms**,"

https://www.ametsoc.org/ams/index.cfm/publications/authors/journal-and-bams-authors/author-resources/list-of-acronyms-and-abbreviations/

"GCM" refers to General Circulation Model. Apparently, it is not outdated.

At the same AMS URL, under "**Climatic, meteorological, oceanographic, and other models**" the acronym "GCM" appears within the name of specific climate models, e.g., "CGCM3.1 CCCma Coupled Global Climate Model, version 3.1."

It is further defined under that heading to mean "general circulation model (OR global climate model)" with "general circulation model" as the primary entry.

3.2 *Also, the author uses the term "global average surface air temperature" (GASAT). ... I would probably recommend using GMST.*

3.2 In fact, GASAT is defined as, 'global averag**ed** surface air temperature' throughout. This small change pays attention to the important point that the global temperature record average is a statistical average, [*Essex et al.*, 2007] and not the mean of a physical temperature. GMST is misleading in implying an average of readings of a single temperature. A sentence has been added to paragraph three in the Introduction to draw attention to this fact.

3.3 *The author uses the term "greenhouse gas" to refer to "CO2, CH4, N2O, various chlorofluorocarbons, etc." What about H2O? Or is he intentionally using the term to refer to non-H2O greenhouse gases? If so, this should be mentioned.*

3.3 Water is now specifically excepted from the list.

3.4 *Also, in the explanation for RCP, he uses "5AR" to refer to the 5th Assessment Report instead of the "AR5" acronym.*

3.4 Fixed, thank-you.

3.5 *Some of the terms in the table seem unnecessary to me - does the reader gain much benefit by having to refer back to the glossary for "Lag-1", "theory-bias" or "wve"? These are only used a few times in the paper and it would probably be easier for the readers to have them repeated or explained whenever they are used.*

3.5 These terms, including lag-1, were added because reviewers of prior versions asked after them when left undefined.

The author defers to the editor to decide whether the glossary table should be retained or removed.

4. *Comments on Section 1*

4.1 *... As I will discuss in "Comments on Section 2.4", this argument is inappropriate and invalid for estimating the uncertainties associated with global temperature projections.*

Hence, the discussion of this argument in the Introduction, e.g., most of lines 81-105, should be removed.

4.1 Respectfully, the reviewer is not correct in this evaluation, as shown below. Therefore these sentences have been left intact.

4.2 *Line 49: Instead of saying "by about 3 Celsius", use the degree symbol, i.e., 3[degree sign]C, and maybe give a range of predicted values too.*

4.2 Celsius does not take the degree sign. Range is unimportant to the argument.

4.3 Line 52: *"...is central to the question of causality..." is quite ambiguously phrased - causality of what, exactly? Maybe say something like "...is central to this prediction..."*

4.3 "Causality" takes its context from the immediately prior sentences of the paragraph within which its own sentence is embedded. Paragraphs typically deal with an extended thought. "Causality, then, must refer back to the warming climate. It is difficult to understand the reviewer's confusion.

Nevertheless, in deference to the reviewer's concern, sentence one of the paragraph has been revised to, "*... could cause an increase in...*," providing a direct inferential connection with the message of the third sentence.

In science, "prediction" is a causal deduction from falsifiable theory. "Causal" is therefore primary.

4.4 *Lines 124-127: The author claims that he won't attempt to survey or address the physics of climate, or to model the terrestrial climate in any way. Instead, he will focus "strictly on the behaviour and reliability of climate models alone". This seems a reasonable approach to keeping the article short, focused and on topic. Yet, in Section 2.1, he then proceeds to describe a simple model of global mean surface temperature.*

4.4 The author does no such thing in Section 2.1 (or anywhere else) as to, "*describe a simple model of global mean surface temperature.*" Section 2.1 describes an emulator of GCM air temperature projections. The foundational study of [*Manabe and Wetherald*, 1967] is assessed to extract the fractional CO_2 forcing <u>appropriate to climate models</u>. See sentence 1 of Section 2.1: "*... as relevant to general circulation climate models*."

This has nothing to do with describing or deriving, "*a simple model of global mean surface temperature*." It is used to derive an emulator of climate model behavior.

The same reviewer misperception is present in items 1.1 and 1.2 above, and appears again later in the review.

4.5 *Lines 129-182: In my opinion, the outline of the structure of the paper is probably too long. A short outline of just a paragraph or two should be sufficient. He doesn't need to go through all the details of what happens in the rest of the paper (this isn't a review or thesis!). The reader will find out these details in a few pages & it makes the points quite repetitive.*

4.5 The Introductory outline was added because reviewers of prior versions required it. Their criticism was that the manuscript was difficult to follow without a roadmap. However, the Introduction has been shortened somewhat to try to meet the reviewer's concerns.

5. *Comments on Sections 2.1 and 2.2*

5.1 *The author presents a simple, semi-empirically derived model of the expected global SAT increase from an increase in GHG.*

5.1 Again, the author does no such thing. The model emulates the behavior of climate models. To emphasize: ***it does not*** provide, "*the expected global SAT increase from an increase in GHG*."

As indicated in item 1.1, this point is repeatedly made in the manuscript. The reviewer even took note of this point in review item 4.4, but here again set it aside.

5.2 *This is the same model he presented earlier in Frank, 2008. He then claims (or assumes?) that the physics underlying his model are essentially the same as the physics built into GCMs ...*

5.2 No such claim or assumption about physics is made anywhere in the manuscript, or in the Supporting Information. The opening sentence of Section 2.1 introduces eqn. 6 as, "*...a simple emulation model of how modern GCMs project...*" not as a model reproducing GCM physics. It is then demonstrated through multiple examples that eqn. 6 successfully emulates GCM air temperature projections, identifying them as linear extrapolations of GHG forcing. No claim of physical congruence is anywhere made.

Absent any such claim of physical identity anywhere in any of the author's work, one is left wondering from where the reviewer obtained this idea. The reviewer provided no example.

5.3 *... (and in Frank, 2008 he also suggested that his theoretical model were reasonable substitutes for GCMs in terms of estimating global SAT projections without involving the considerable computational expense of a GCM).*

5.3 That suggestion did not appear in [*P. Frank*, 2008]. The point of congruence with climate model air temperature projections was made in the talk the author gave at the 2016 meeting of the Doctors for Disaster Preparedness, available online since 29 July 2016, here: https://www.youtube.com/watch?v=THg6vGGRpvA.

In that talk, it was merely pointed out that the demonstration of successful emulation meant the PWM can be used to produce the analogous time-temperature lines without needing a climate model. No claim was made concerning a congruence of physics.

5.4 *This appears to be a key assumption in his argument, yet it is flawed.*

5.4 There is no such assumption anywhere in the manuscript or in the author's argument.

5.5 *There are a number of different theoretical models which have been developed to estimate the expected global SAT response to GHG. ..., etc.*

5.5 The reviewer's entire disquisition here on GCM physics is irrelevant because manuscript eqn. 6 emulates GCM outputs, **not GCM physics.**

It seems necessary, once again, to observe that the manuscript repeatedly made this distinction (see item 5.8 below). Nevertheless, the reviewer has quite evidently not grasped this analytical essential.

Manuscript Figures 2, 3, 4, 8 and 9, along with SI Figures S1 and S3 through S8, thoroughly demonstrate the successful emulation. Successful emulation demonstrates that, regardless of internal physics, GCM air temperature projections are mere linear extrapolations of GHG forcing.

5.6 *For this reason, the author should not assume that the physical implications suggested by his model necessarily also apply to the GCMs.*

5.6.1 Once again, no such assumption was made. No climate physics is anywhere implied by manuscript eqn. 6. No climate physics is anywhere imputed to manuscript eqn. 6.

Review section 5 reveals that the reviewer has entirely misunderstood the critically central point of the study, no matter that it was reiterated throughout the manuscript (see item 5.8 below). Review section 5 has no relevance to the manuscript analysis.

5.6.2 Manuscript eqn. 6 demonstrates that GCM air temperature projections are linear extrapolations of GHG forcing; that demonstration and nothing more.

5.7 *Having said all of that, from what I can tell, the main relevance of Section 2.1 to the article seems to be his finding in Section 2.2 that the global SAT projections are reasonably well approximated as simple linear extrapolations of the increase in [GHG]. If that is the main point that is being made, then this could presumably also be demonstrated by applying a linear least squares fitting between the GCM projections and the changes in GHG forcing. In my opinion, this would be a simpler and more straightforward approach to reaching the same conclusion.*

5.7.1 The reviewer is not correct in this conclusion. Fitting a projection merely demonstrates something about that particular output. General emulation demonstrates something about the source of all the outputs.

Manuscript Figure 2 shows the general, not specific, relevance of emulation eqn. 6 to GCM projections. The remaining Figures 3, 4, 8 and 9, along with SI Figures S1 and S3 through S8 then demonstrate specific applications of the general case.

5.7.2 The reviewer is also invited to consult Section 3 of the Supporting Information, "*The generalized PWM used to test individual SRES and RCP projections*," where the method of determining the coefficients for equation 6, applicable to individual GCM projections, is described.

5.8 *Indeed, in lines 124-127 of Section 1, the author explains how he will not try to survey or address the physics of climate, or to model the terrestrial climate in any way. Yet, this seems to be exactly what he does with Section 2.1!*

5.8 The author does no such thing in Section 2.1. The several examples below demonstrate the reviewer's perceptual mistake; note especially the bolded phrases.

The reviewer is contradicted by the very first sentence in Section 2.1: "*The analysis begins in this section with an estimate of the fraction of the terrestrial greenhouse temperature produced by the wve forcing of CO2, **as relevant to general circulation climate models**.*"; i.e., ***not*** as relevant to climate physics.

Section 2.1, line 195: "*[Manabe's results] ... can be used to estimate the fractional baseline GH warming induced by wve CO_2 forcing **as deployed in climate models**.*"; i.e., ***not*** as rendered in climate physics.

Line 290, Section 2.1.3, title: "*The fractional GH Temperature due to wve CO_2 forcing, **as relevant to climate models**.*"; i.e., ***not*** as relevant to climate physics.

Section 2.1.3, line 297-300: "Thus the asymptotic temperatures estimate the greenhouse warming due to water vapor alone under clear or cloud-covered skies, respectively, **as represented in the model used by Manabe and Wetherald.**"; i.e., ***not*** as represented in climate physics.

Line 313, Section 2.2, Title: "*A General Emulation of the GASAT Projections of Climate Models*"; ***not*** the physical response of air temperature to GHG concentration.

Line 313, first sentence of Section 2.2 "*The foregoing is sufficient to inform a simple emulation model of **how modern GCMs project the impact** of increasing greenhouse gases on global averaged air temperatures.*"; i.e., ***not*** how the physical climate responds to increasing GHGs.

Line 324: "*Eqn. 6 is a surmise that GCMs project the GASAT as a linear extrapolation of fractional wve GHG forcing.*"; i.e., ***not*** a representation of the physics within GCMs.

Line 340: "*Equation 6 is termed the "Passive Warming Model" (PWM) because it represents the increasing GASAT projections of GCMs to follow directly and linearly from the wve forcing due to changing GHGs.*"; i.e., ***not*** because it represents anything about climate physics.

These examples should be sufficient to demonstrate, beyond any doubt, that the physics of climate is ***not*** addressed, surveyed, or assumed ***anywhere*** in the manuscript.

Further, these examples demonstrate that the specific focus on the behavior of climate models is noted over and yet over again in the manuscript. The physics of climate does not appear anywhere.

The reviewer is here again exhibiting a thorough and fundamental misunderstanding of the analytical focus, despite its clear and repeated expositions in the manuscript.

5.9.1 *Therefore, I recommend the author considers how important his equation 6 model is to this study:*

5.9.1 The author has so considered and recognizes eqn. 6 as critically central to the study. Eqn. 6 does, after all, demonstrate the linearity of GCM air temperature projections with respect to GHG forcing. The remainder of the manuscript analysis follows directly from that demonstration.

The reviewer has shown throughout to have thoroughly misunderstood manuscript Section 2.1 and the significance of eqn. 6. Therefore the reviewer's suggestion may safely be set aside.

5.9.2 *If he is just using it for the linearity result, then I would remove Section 2.1 and replace his analysis in Section 2.2 with a statistical analysis of the relationship between GCM projections and GHG forcing using, e.g., linear least squares fitting.*

5.9.2 The reviewer has misunderstood manuscript Section 2.1. Therefore the reviewer suggestion item 5.9.2 is ungrounded.

See also response item 5.7.2.

5.9.3 *If he wants to use this opportunity to point out that this is consistent with his model, he could do this with a simple sentence, e.g., "This is consistent with an earlier analysis by the author [Frank, 2008]" or something similar.*

5.9.3 The present manuscript is more rigorous and more complete than the Skeptic paper, including the derivation from the work of [*Manabe and Wetherald*, 1967], the mean free path analysis justifying the utility of Figure 1b, the extension to CMIP3 and CMIP5 models, and the new inclusion of the RCP scenarios.

5.9.4 *If he wants to use this article to highlight his 2008 model and argue that its global SAT projections are a reasonable substitute for those of the GCMs, then in my opinion, he should carry out a far more detailed compare/contrast assessment of the similarities and differences of his model to the GCMs (as well as to other models such as Benestad, 2016). Indeed, I would argue that this would probably be better carried out as a separate study, and should include a more detailed analysis than the original Frank, 2008 paper.*

5.9.4.1 Reviewer suggestion 5.9.4 is a perfect demonstration of the misunderstanding that pervades the review. Manuscript eqn. 6 does not produce substitutes for GCM projections. This point is extensively documented above.

Throughout the manuscript, the distinction was emphasized that eqn. 6 is about emulation but **not** about physics. Nevertheless, the reviewer has failed to grasp this essential point, despite that it was repeatedly presented.

5.9.4.2 The reviewer apparently did not consult the Supporting Information, where the extremely extensive ("*a far more detailed compare/contrast assessment*") comparison of the emulations from eqn. 6 with the outputs of 30 air temperature projections from 21

CMIP3 GCMs is presented, along with 12 projections from six CMIP5 GCMs.

All this is new in addition to, in the manuscript, comparison with 19 projections from 16 CMIP2+ models, with 18 projections from six CMIP3 GCMs, and with 4 projections from two more CMIP5 GCMs, and with the IPCC ensemble average.

Therefore, the more extensive analysis desired by the reviewer is already provided.

5.9.5 *At any rate, regardless of whether his semi-empirical model or a linear least-squares approach (or something else) is used, he should add some analysis of the residuals from the linear fits. Indeed it is quite ironic that the author is criticising the prior discussion of GCMs for not adequately assessing the residuals between the model output and observations, yet he does not include any such assessment of his claim of linearity. Even just by eyeballing Figures 2 and 3, it seems to me that there are noticeable differences between his linear PWM projections and the GCM projections.*

5.9.5.1 The emulations used the standard SRES and Meinhausen RCP forcings. Where the GCMs apparently adhered to these forcings, the emulations were exact, e.g., Figure 3, Figure 4a,b and Figure 8, and the figures in the SI.

However it appears that some GCM SRES or RCP projections departed from the standard forcings. Illustrating this case, where the emulation departs from a given GCM projection, other projections from the same model are exactly emulated. For example in Figure 3 the CSIRO mk3.0 SRES A1B projection shows a departure, while the B1 and A2 CSIRO emulations are exact. The GFDL GCM apparently adhered to the standard SRES forcings, and the emulations are exact.

The same considerations apply to the emulations displayed in the Supporting Information. Indeed, virtually every case of a departure was restricted to an A1B SRES scenario. However, some A1B SRES scenarios were exactly emulated. This shows that the problem lay in the GCM departing from the standard forcings, and not with the approach to emulation. Were the identically-used forcings available in the departure cases, the emulation would almost certainly have been exact.

5.9.5.2 Most of the GCM simulations include significant climate noise, e.g., Figure 2, Figure 3, Figure 4, and Figure 9, as well as Supporting Information Figure S3 through Figure S8. This climate noise will dominate the emulation residuals. Following the reviewer's recommendation, therefore, would yield a mostly meaningless difference metric.

5.9.5.3 The author does not criticize prior work for, "*not adequately assessing the residuals between the model output and observations.*" The author criticizes prior work for not providing physically valid uncertainty bars. Criticized also is the related and prevalent confusion of precision with accuracy.

5.10 *I still think the deviations from linearity are probably small enough to justify saying something like, "...the GCM projections are reasonably approximated in terms of a linear relationship..." But, the deviations from linearity (i.e., residuals with regards to the linear model/fit) should still be discussed and statistically analysed. Also, on lines 403-5, he shouldn't* *claim the projections "...are just linear extrapolations..."*

5.10.1 The unfailing success of emulation eqn. 6 demonstrates that GCM air temperature projections are indeed linear extrapolations of GHG forcing. Pointing out this relation is thus entirely justified.

5.10.2 For the reviewer's benefit, [*Andrews et al.*, 2012; *Gregory et al.*, 2004] show the linear relation between GHG forcing and air temperature expressed within GCMs. I.e., that $N = F-\alpha\Delta T$, 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 successfully emulate the air temperature projection of any CMIP5 model, when the GCM-specific value of α is known.

These papers fully justify the strong conclusion the reviewer wishes to set aside.

Additionally, the linear relation between forcing and projected air temperature is candidly admitted by the IPCC in box 1.3 of [*Pyle et al.*, 2016], i.e., $\Delta T_s = \lambda\Delta F$, where λ is model climate sensitivity.

This admission by the IPCC, in and of itself, again fully justifies the strong conclusion made within the manuscript.

5.10.3 Response item 5.9.2 resolves the question of an emulation difference metric.

6. *Comments on Section 2.3*

6.1 *In this section, the author uses the analysis of several other studies (in particular Jiang et al., 2012 and Lauer & Hamilton, 2013) to conclude that there are systematic biases in the current GCMs' simulated total cloud fraction estimates.*

6.1 The reviewer is not correct. The author's own analysis demonstrated the presence of systematic cloud error by means of the data and calculations presented in Figure 2, Figure 3, and Table 1, and the discussion in the associated text.

[*Jiang et al.*, 2012] reported the cloud error and [*Lauer and Hamilton*, 2013] quantified the long wave cloud forcing (LWCF) error that derived from the error in simulated cloud cover. However, neither of these studies were themselves used in the manuscript analysis showing the systematic nature of cloud error. This latter demonstration is unique to the present manuscript.

6.2 *I think the author does a reasonable job of arguing that there appear to be systematic biases in the current GCM cloud modelling. This observation isn't new - after all, he is basing his analysis on the observations of several previous studies.*

6.2 The conclusion of systematic cloud error that is strongly correlated among CMIP5 models, is indeed a new result first reported in this manuscript.

6.3 *I would make some changes, though:*

The author seems to be implicitly assuming that the satellite datasets are perfect and without any systematic biases themselves. This is a big assumption which underlies much of the ensuing discussion.

6.3 There is no such assumption. The author presents LWCF error as a lower-limit of uncertainty, not as a complete accounting. See line 164 of the introduction, and the title of Section 2.4: " *A lower limit of uncertainty in the modeled global average annual thermal energy flux.*" See also manuscript lines 517, 518, and 722.

It is true that there are uncertainties in satellite observations of cloud cover, typically about ±10%. A more complete accounting of simulation uncertainty would include the calibration observational uncertainty added in quadrature to the simulation-observational difference. Such a study would be very valuable, and the reviewer is invited to make it.

6.4 *Personally, I think it is a reasonable working assumption to make, but it should be made explicit.*

6.4 As noted in item 6.3, the working context is the lower-limit of uncertainty, not that observational error is absent. This uncertainty model is explicitly stated.

6.5 *With that in mind, the sentence on line 446 saying "That is, the simulation is inaccurate", and also the "...without an improvement in TCF verisimilitude" on lines 495-6 should be toned down or else removed.*

6.5.1 To meet the reviewer's concern, the sentence has been changed to, 'the simulation is incomplete.' Nevertheless, in view of the fact that the observed cloud cover was the target of the simulation hindcasts, it is difficult to understand how a simulation minus observed error metric does not show that the simulation was inaccurate.

6.5.2 The reviewer has not indicated where the analytical context of prior lower correlation among AMIP1 model simulation errors coupled with similar average error does not justify the conclusion that the verisimilitude has not improved with time.

6.6 *Also, doesn't Jiang et al., 2012 argue that "...more than half of the models show improvements from CMIP3 to CMIP5 in simulating column-integrated cloud amount..."? This seems to be slightly at odds with the discussion on lines 490-496 on the comparison between AMIP1 and CMIP5.*

6.6.1 It is not necessarily at odds, because the manuscript analysis concerns the pairwise correlations between the model errors, not individual model assessments.

[*Jiang et al.*, 2012] discuss comparative CMIP3, CMIP5 changes in simulated individual water paths, e.g., liquid water path (LWP) or ice water path (IWP), within model families. The manuscript assesses the pairwise correlation of total cloud error among models.

[*Jiang et al.*, 2012] do not discuss pairwise correlation of model cloud error. Thus, such improvements as noted by [*Jiang et al.*, 2012] have no bearing on trends in cross-model correlation of cloud error.

6.6.2 Relevant to the reviewer's focus, however, from [*Jiang et al.*, 2012] Figure 5, of the five of twelve CMIP5 models that showed improved LWP only two, MIROC and GFDL, also appeared among the seven showing improved IWP.

Other CMIP5 models were unchanged or worse relative to their CMIP3 equivalents. That is, CMIP5 models with improved LWP could show poorer IWP, and vice-versa. The two GISS models are a notable example of this trade-off.

The manuscript assessment of cloud error concerns total cloud fraction, not individual paths. The trade-off in model individual path error noted above qualitatively explains the small overall improvement in simulated total cloud fraction found in the manuscript analysis.

6.7 *Figure 5 is difficult to properly assess without Figure S10 in the SI. Therefore I recommend moving Figure S10 back into the main article, so that the residuals can be directly compared to the model simulations and satellite observations themselves.*

6.7 Given the issues of space, the author leaves this decision to the editor.

6.8 *Lines 485-488: As far as I know, the main difference between the GISS e2r simulations and the GISS e2h were to do with the ocean models used (Russell and Hycom ocean models respectively - see https://data.giss.nasa.gov/modelE/cmip5/) and had nothing to do with the atmospheric components. So, it is not surprising that they produced similar results with regards to cloud modelling. Therefore, I don't think this analysis adds to the discussion and lines 485-8 should probably be removed.*

6.8 The fact that the two GISS models are both so similar in their physics, as the reviewer noted and as noted in the manuscript, and produced the most highly correlated error most strongly demonstrates the point of systematic error.

Were model cloud simulation errors random, one would expect models deploying a similar physics to produce random errors as commonly uncorrelated, as they are uncorrelated with the errors of models with less similar physics. The fact that the physically most similar models produced the most highly correlated error demonstrates the source of the simulation error derives from systematic theory error. That is, the GISS model result provides the strongest proof of the point.

This assessment has been added to the text to clarify the point.

6.9 *In lines 419-425, when discussing the Jiang et al., 2012 study, the author states, "CMIP5 TCF hindcasts comprising 25-year (1980-2004) annual projection means were compared to the corresponding A-train observations". When reading this, I mistakenly assumed that "corresponding" meant the same 1980-2004 period, and I suspect many readers would also. It was only when checking Section 6 of the SI (which begins on page 13 of the SI, not 12 as* *the Table of Contents claims) that I realised this wasn't the case. The reason why Jiang et al., 2012 used non-overlapping periods was reasonably well-justified, i.e., the models didn't have any significant trends in clouds or water vapour over those periods. However, this has implications for the author's analysis in Section 2.4 (see below). So, the use of "corresponding" is misleading, and this description should be rephrased.*

6.9.1 The reviewer's point is taken. "the corresponding" has been removed.

6.9.2 It is shown below that the reviewer's expressed implications regarding cloud cover are the opposite of what he inferred.

6.10 *Also, on a minor point, on line 413, should it read "...have been carried out" and on line 415, is the use of the word "acute" a typo? Maybe he meant "accurate"?*

6.10 Line 413 is fixed, thank-you. On line 415, "acute" was meant to convey 'penetrating' or 'revealing.' "Penetrating" has been substituted for acute.

7. *Comments on Section 2.4*

7.1 *I disagree with much of the analysis in this section. In my opinion, much of this section needs to be variously removed, rewritten and/or replaced. As the article is currently drafted, the most dramatic conclusions of the paper are based on this analysis. However, as I mentioned at the start, I do believe there is much of value in the paper, and if the author is prepared to redraft the paper with a different emphasis, there is probably a paper worthy of publication in here.*

7.1 Clearly, the author disagrees with the reviewer's opinion. Shown below is that the reviewer's concerns regarding error propagation are not well-founded and should be set aside.

To be clear, following the reviewer's suggestion to set aside the error propagation makes also pointless the manuscript analysis of [*Manabe and Wetherald*, 1967], the radiative mean-free-path analysis, seven of the nine Figures and all the associated text, most of the Supporting Information, and virtually all the discussion section.

Left only would be a recapitulation of known cloud simulation error and a noting of its correlation among climate models. What would be the point of publication?

The reviewer's suggestion here is remarkably unrealistic.

7.2 *With that in mind, here are my main problems with Section 2.4 as it currently stands:*

The author is basically making the same argument that he has already made in Frank, 2008, except using an "annual LWCF error" of [plus minus]4 W/m2 instead of his earlier estimate of [plus minus]2.8 W/m2, and updating his analysis from CMIP3 to CMIP5. I strongly disagree with the conclusion of Frank, 2008 that the uncertainty limits associated with GCM projections are a very dramatic [plus minus]100[degree sign]C/century. However, if the analysis in Frank, 2008 were correct, then its conclusions were already so dramatic that increasing the error bars *even further wouldn't make much difference. That is, IF the analysis in Frank, 2008 were valid, then this update would be unnecessary and as a result not worth publishing. But, as I will discuss below, I actually don't think that analysis was correct, and so recommend it be dropped from the paper instead of updated.*

7.2 The reviewer's point can be condensed to this: if [*P. Frank*, 2008] is correct then the present analysis is a repetition unworthy of publication.

The reviewer has ignored that the manuscript analysis of [*Manabe and Wetherald*, 1967] is new. The result of that analysis puts equation 6 and thus the entire study on firm analytical grounds.

The mean-free-path argument is new, and the finding that Beer's Law applies only below 1 ppm CO_2 is apparently new. These new findings together justify finding that 1 ppm CO_2 is of negligible forcing. This latter finding is also apparently new.

The [*P. Frank*, 2008] analysis was restricted to CMIP2+ models. The extensions to CMIP3 and especially to CMIP5 models are eminently worthy. They bring the study to high contemporary significance.

The reviewer's dismissal ignores all of this. See also response item 5.9.4.2.

7.3.1 *With regards to his estimate of the "annual LWCF error", his rationale for specifically choosing [plus minus]4 W/m2 is very poorly explained or justified. I had to re-read Lauer & Hamilton, 2013 several times before realising that he was probably referring to this brief sentence in the text - "For CMIP5, the correlation of the multimodel mean LCF is 0.93 (rmse = 4 W m-2)..." (p3833, Lauer & Hamilton, 2013). Is that rmse value the basis for the author's "annual LWCF error" estimate? If so, then the author is being far too cryptic in his rationale. He is also being somewhat misleading, as most readers (including me) would have assumed from his matter-of-fact citation of Lauer & Hamilton, 2013 that this was one of their key findings. If this is indeed the basis for his estimate, he should make it clearer exactly where in the paper he took his estimate from and let the reader know that this was only a passing comment in Lauer & Hamilton, 2013. If it's not the basis for his estimate, then I couldn't find anything else, so he definitely needs to be more explicit in his reasoning!*

7.3.1.1 The reviewer suggests that the rationale for choosing the long wave cloud forcing error from is poorly explained or justified. Yet, the first sentence in Section 2.4.1 (line 502) says, "*CMIP5 TCF error entrains an error in simulations of tropospheric thermal energy flux.*" Tropospheric thermal energy flux is the source of tropospheric air temperature. Surely, this connection should come readily to mind.

Following this, paragraph 2 and paragraph 3 in Section 2.4.1 (lines 514-529) go on to provide a detailed rationale for the choice of LWCF error as the resolution metric governing model simulations of air temperature.

Sections 2.3 and 2.4 of the revised manuscript have been revised and expanded to remove all possible ambiguity.

7.3.1.2 Manuscript Section 2.3 focuses on total cloud fraction error, to prepare for the encounter with LWCF error in Section 2.4.

Total cloud fraction and TCF error are again introduced right at the beginning of Section 2.4 and after some discussion are further related to the work of [*Lauer and Hamilton*, 2013].

The mean model uncertainty in total cloud fraction was not merely, *"a passing comment*" in [*Lauer and Hamilton*, 2013] but one of their reported analytical findings, and the obvious metric of analytical focus.

Extracting and referencing this result from among the other results reported in [*Lauer and Hamilton*, 2013] is entirely appropriate to the analytical focus of the manuscript. The reviewer apparently had to work a bit to find the relevant discussion in [*Lauer and Hamilton*, 2013] and has chosen to complain about it. The need to put in a certain amount of work to understand a scientific paper does not seem reason for complaint.

Nevertheless, a new explanation has been added to Sections 2.3 and 2.4.1, and the derivational logic of LWCF error statistic is now presented in new Section 6.2 of the Supporting Information.

7.3.2.0 *Assuming that was the basis for his estimate, ...*

7.3.2.0 Why is an assumption necessary? The source of the LWCF error metric was clearly made in manuscript lines 506-512.

7.3.2.1 *I am not sure that this is actually an appropriate value. From what I can tell, the rmse of 4 W m^{-2} refers to a spatial correlation over the entire globe.*

7.3.2.1 The ±4Wm^{-2} rms LWCF error derives from the errors in simulated cloud fraction. See new SI Section 6.2 It does not refer to correlation. The geospatial correlations in [*Lauer and Hamilton*, 2013] are between observed and simulated cloud cover. They do not refer to the uncertainty in LWCF.

7.3.2.2 *From Figure 2 of Lauer & Hamilton, 2013, it can be seen that the values for some regions are overestimated, while other regions are underestimated. However, since the author is estimating the uncertainty associated with *global* SAT, this estimate of the regional variability in the biases is less relevant.*

7.3.2.2.1 Regional errors across the globe are combined in quadrature to produce a global uncertainty. This is not hard to grasp.

The significance of the LWCF rms error is understood from [*Lauer and Hamilton*, 2013] equation (1) and the accompanying text. Simulated cloud cover error is appraised per grid-point across each year for each model. Twenty-year mean errors are derived for each model, and the rms error is calculated. The rms error has global significance, and indicates a representative simulation uncertainty generally applicable to CMIP5 climate models.

The final LWCF rms error reflects uncertainty in the simulated atmospheric thermal energy flux. In this light, it is very difficult to understand how the reviewer can conclude an uncertainty in simulated global atmospheric thermal energy flux does not impact the certainty in simulated global air temperature.

7.3.2.2.2 In review item 7.3.2.1, the reviewer stated that the LWCF rms error referenced the entire globe. This was correct. In 7.3.2.2 the reviewer then re-supposes the LWCF rms error has only regional significance. Review item 7.3.2.2 thus contradicts review item 7.3.2.1.

7.3.2.3 *I might be wrong in this interpretation of the 4 W m-2 figure, since I am not very familiar with the satellite-based cloud datasets, but if so this would again highlight how the author has failed to *clearly* explain the rationale for his estimate.*

7.3.2.3.1 See response item 7.3.1.1. The author's rationale was clearly explained.

7.3.2.3.2 The reviewer's "*His estimate*:" the ±4 Wm^{-2} rms LWCF error is not the author's estimate. It is the analytical result from [*Lauer and Hamilton*, 2013]. This was clearly explained in manuscript lines 506-523.

To make this point more clearly, a new explanation has been added to the opening of revised manuscript Sections 2.3 and 2.4. There it is now pointed out that the [*Lauer and Hamilton*, 2013] appraisal of model versus observations is a model calibration, with the observational record as the calibration standard. The LWCF rms error is a calibration error intrinsic to, derives from, and is produced by all CMIP5 models. All simulations derived from these models are subject to that error.

In addition, the logic entering [*Lauer and Hamilton*, 2013] equation (1) is now derived in new Section 6.2 in the Supporting Information, with a summary in the new opening to Section 2.4. It is hoped these additions resolve the reviewer's concerns.

7.4.1 *Even assuming that the estimate is correct, this does not in itself mean that this applies to the uncertainties over global mean surface temperature trends, which is what he is attempting to estimate.*

7.4.1 Response items 7.3.1.1 through 7.3.2.3 should have made it clear that the ±4 Wm^{-2} rms LWCF simulation error does indeed apply to uncertainty in simulated global air temperature.

If they do not, then the author hopes that revised Sections 2.3 and 2.4 and the new derivation in SI Section 6.2 provide the needed resolution. LWCF error is a global annual average uncertainty metric in simulated forcing. It represents a lower limit of CMIP5 resolution of the tropospheric thermal energy flux, and thus of the thermal impact of anthropogenic GHG forcing.

Again the ±4 Wm^{-2} rms LWCF error is not the author's estimate.

7.4.2 *First, as he notes in the SI, Jiang et al., 2012 didn't find any significant trends in the models* *for clouds or water vapour over the 1980-2004 period. If the models didn't show significant trends in clouds over this period, then the associated error due to clouds on GMST shouldn't have changed much either.*

7.4.2 The SI noted that [*Jiang et al.*, 2012] found no significant trend in observed cloud cover over the evaluation period. The statement does not concern the error in simulated cloud cover over that period.

Figure 2 in [*Lauer and Hamilton*, 2013] show the twenty-year average simulation errors in various aspects of cloud cover, across the globe. Whether global average cloud cover changed or didn't change much during this time has no bearing on the fact of simulation error. It is difficult to see how the reviewer thinks there is critical content in this.

7.4.3 *Second, the models still showed trends in GMST over the same period (e.g., Fyfe et al., 2013), i.e., the GMST trends of the models over that period were not heavily influenced by the trends in clouds. Regardless of whether you consider this a problem with the current models or not, this indicates that the 4 W m-2 uncertainty in LWCF only at best indirectly influences the model GMST trends.*

7.4.3 This point has been addressed in [*Kiehl*, 2007], who noted that models are tuned to reproduce global air temperature. The errors arising from incorrect cloud cover are thereby hidden.

It is also noted that the reviewer is judging model accuracy in a circular fashion; i.e., by the coherence in model outputs that are not impacted by modeled structure.

The manuscript analysis derives the uncertainty in the simulated global air temperature. It does not specify anything about the magnitude or trend in global air temperature. Statistical uncertainties do not influence model expectation values. Reviewer item 7.4.3 confuses a statistical uncertainty with a physical perturbation. This is a very naïve mistake.

7.4.4 *Third, as Palmer, 2016 notes in his Figure 1, there is considerable disagreement between the various CMIP5 models on the absolute global mean surface temperature, but the models are relatively consistent in the trends for global mean surface temperature anomalies. Perhaps one factor could be related to this uncertainty in LWCF, e.g., the different LWCF estimates alter the model's estimates for the absolute GMST.*

7.4.4 This point is resolved in response item 7.4.3. Similarity of trends among models is a direct result of model tuning plus the shared assumptions concerning the impact of GHG forcing built into the model physics.

7.4.5 *But, in that case, the associated uncertainty would apply to the absolute GMST, not the GMST trends.*

7.4.5 There is a standard way of combining uncertainties into a difference. The uncertainty in a difference (the anomaly) is the uncertainty in the mean added in quadrature to the uncertainty in the individual absolute temperature used for the anomaly *cf.* Chapter 3 in [*Bevington and Robinson*, 2003]..

The uncertainty in the temperature mean is the rms of the individual temperatures entering the mean. The anomaly uncertainty is therefore invariably larger than the uncertainty in any of the individual temperatures.

This standard bit of uncertainty propagation into a difference is thoroughly ignored throughout climate science, and the reviewer has ignored it here.

The reviewer is also implicitly assuming that model error is a known constant offset that subtracts away in an anomaly. This assumption is also widely shared in the climate modeling community and is thoroughly discussed in SI Section 7. The assumption of constant error has never been empirically tested.

7.5 *Additionally, if there is a systematic bias in the way the GCMs model clouds, as opposed to the biases being merely stochastic in nature, then the errors are definitely non-Gaussian in nature and the errors should be treated as systemic, not stochastic.*

7.5 The manuscript approach follows the method of propagation of systematic error in linear physical models thoroughly discussed in [*Vasquez and Whiting*, 2006]. They provide equivalent summations to propagate both systematic and random error.

For systematic error, this is $\pm u_{sys} = [\sum_{i=1}^{m} \varphi_{s,i}^{2}]^{1/2}$, where *i* indexes each source of error and φ_s is the systematic error within the i^{th} error source. This is exactly the approach taken in the present manuscript.

[*Vasquez and Whiting*, 2006] go on to say that this equation is not fully appropriate for non-linear models, because of the non-linear transformations. It is for this reason that the linearity of manuscript equation 6 is emphasized and is so central to the uncertainty calculation. That is, linear propagation of systematic error is exactly appropriate to the linear emulation equation.

[*Vasquez and Whiting*, 2006] also mention that the systematic errors of thermodynamic data are not constant and tend to be a function of the magnitude of the variables measured. In the case of non-linear models, the magnitude and perhaps the sign of systematic error may be a function of the expectation value. All such errors propagate into an extrapolation as the ±rms uncertainty.

Interestingly, [*Vasquez and Whiting*, 2006] also point out that, "*random errors produce skewed distributions of estimated quantities in nonlinear models. Only for linear transformation of the data will the random errors cancel out*." Given the almost universally deployed Gaussian pdf error models, it seems unlikely this point is widely recognized.

7.6.1 *More generally, the author's argument seems to be that because climate model projections are step-wise in nature (lines 101-105), he believes that the errors associated with the GMST projections for each year should be cumulatively summed together with the errors from* *previous years using his equation 2 (lines 89-94). This is incorrect for several reasons.*

7.6.1 The reviewer's logic is not correct. The summation involves uncertainties, not errors. This difference is critical. Root-sum-square cumulation of uncertainty through a calculation is standard.

Next, the uncertainty arises from the simulated atmospheric thermal flux, **not** from the reviewer's GMST. The uncertainty in temperature derives from the cumulated uncertainty in thermal flux. The uncertainty in temperature is not independently summed as the reviewer's language implies.

This distinction is fundamental. Unfortunately the reviewer has apparently failed to grasp it.

The logic concerning uncertainty is first given in manuscript lines 542-552 and elaborated in manuscript Section 2.4.2, and again in Section 3, lines 736-763.

To summarize, the manuscript argument is that the uncertainty in simulated atmospheric thermal flux is a property of the model and enters every single projection step. The uncertainty of simulation step *n* must propagate into simulation step *n+1* because the simulation of step *n+1* initializes from the unknown errors introduced from step n. In step *n+1*, the model again makes some error in LWCF, of unknown magnitude but within the average range of ±4 Wm^{-2}.

This further error in step *n+1* is again of unknown magnitude and direction. It increases the uncertainty -- the level of ignorance -- in the phase-space position of the simulated climate relative to the unknown physically correct position.

Again, the physical error magnitudes are unknown in a projection of future climate, but the errors themselves are certainly present. The known stepwise introduction and propagation of error produce stepwise uncertainties that must then themselves propagate forward.

The physical errors are known present and introduced into each projection step by way of model calibration experiments, such as that presented in [*Lauer and Hamilton*, 2013].

7.6.2 *Yes, climate models projections are step-wise in nature. However, while the global surface mean temperatures in a given step do have an influence on the projections of the next steps, this is generally an indirect influence (through feedback loops, etc) and is definitely not the only (or even a major) factor. This can be demonstrated using a simple thought experiment: if the GMST of one year was solely a function of the GMST of the previous year, then GMST would be effectively constant over time. Instead, the global warming projected by the CMIP5 models is mostly a consequence of rising GHG concentrations.*

7.6.2 The manuscript uncertainties do not derive from erroneous air temperatures. They arise from model errors of unknown magnitude in projected atmospheric thermal energy flux.

Further, in any year, global average air temperature is derivative of global atmospheric thermal energy flux. That is, air temperature is a dependent variable. It is not independent and cannot influence its future magnitude.

Manuscript lines 532ff and equation 7 describe how uncertainty in global annual average air temperature derives from uncertainty in the simulated atmospheric thermal energy flux. Manuscript line 561: "*This uncertainty in simulated total thermal energy flux produces an uncertainty in simulated air temperature.*"

Neither air temperature itself, nor uncertainty in air temperature, plays any rôle whatever in the year-by-year propagation of uncertainty.

The reviewer's analysis is again fundamentally misguided.

7.7 *I agree that GMST of the previous years have SOME influence on the next in the models through feedbacks (changes in sea ice, atmospheric circulation, etc). This can be seen from the difference between "Equilibrium Climate Sensitivity" estimates of the long-term expected GMST increase from a step doubling in CO2 and "Transient Climate Response" estimates derived using simulations involving a gradual increase in CO2. However, it is only an indirect influence, and it is not a simple one either.*

7.7 Response items 7.6.1 and 7.6.2 resolve this issue. Again, the reviewer's GMST plays no part at all in the production or propagation of simulation uncertainty.

The reviewer's argument further reveals a continuing misperception that manuscript eqn. 6 is about climate physics. It is not. Eqn. 6 is about the behavior of climate models.

The reviewer seems to think that the uncertainty in air temperature is air temperature itself. This is to mistake a statistic for a thermodynamic magnitude. Such a mistake could not be more basic.

7.8 *So, the author's use of equation 2 for estimating the uncertainty of GMST projections is invalid.*

7.8 It should be clear at this point that the reviewer's conclusion has no analytical weight. No part of review section 7 survived critical scrutiny.

8. *Comments on Section 3*

8.1 *In the Summary and Discussion section, the author focuses almost entirely on his analysis in Section 2.4. However, in my opinion, this is probably the weakest part of the paper.*

8.1 Noted, but see response items 7.1 through 7.7.

8.2 *There are many reasons for questioning the reliability of the current GCMs' GMST projections including the uncertainties over their ability to model clouds, as well as the references I* *mentioned in the Introduction. However, the author's application of equation 2 is inappropriate and leads to an overly dramatic and unrealistic dismissal of their reliability. In my opinion, the author should drop his focus on his equation 2 (which he has been doing since Frank, 2008). Instead, he should focus on the more justifiable concerns over the current GCMs, such as the ones I mentioned in the introduction.*

8.2 In view of the above critical responses and the evidence presented, the author respectfully suggests that the reviewer's opinion is entirely without force and that manuscript eqn. 2 is entirely appropriate.

8.3 *I also think he should try to be more constructive in his criticism of the current GCMs. Even climate modellers agree the current models have many flaws (e.g., Fyfe et al., 2013; Palmer, 2016). However, this doesn't mean that they are useless or can't be improved. The author should try to provide practical and realistic recommendations for how the modelling community can improve the GCMs rather than implying that they are useless.*

8.3 The author nowhere implies that climate models are useless. The author merely concludes that climate models are unable to resolve an annual perturbation of ~0.035 Wm^{-2} in atmospheric thermal energy flux (average since 1979). This is not a harsh judgment; nor is it particularly surprising given the acknowledged very large model errors in the published literature.

The shock in this conclusion comes entirely from the public investment and high visibility the field has made in promoting the ability to resolve such a tiny perturbation. That this ability was impossible was very clear on publication of [*Soon et al.*, 2001], but that cautionary analysis was roundly ignored. The systematic blindness of the field to a recognition of obvious physical error is unprecedented in the history of science, in the author's view.

8.4 *If the author could generate more realistic lower bound estimates of the uncertainty in the accuracy of the projections, this could be of help to the modelling community, as well as the rest of the climate science community. However, the estimates he has calculated in this draft are completely unrealistic, mostly because of his inappropriate use of his equation 2.*

8.4 Again, given the response items above and the evidence presented, the author respectfully suggests that the reviewer's opinion is mistaken.

The uncertainty limits are not at all unrealistic in view of the fact that the annual average uncertainty of ±4 Wm^{-2} in atmospheric thermal energy flux is ±114 times larger than the annual average thermal perturbation of ~0.035 Wm^{-2} (since 1979) to be resolved. Further, the projections extend out a full century.

This combination of a relatively huge error and a long extrapolation make inevitable the large uncertainties.

The reviewer might like to know that the author extensively discussed the results of [*Anagnostopoulos et al.*, 2010] in [*Patrick Frank*, 2015].

Anagnostopoulos, G. G., D. Koutsoyiannis, A. Christofides, A. Efstratiadis, and N. Mamassis (2010), A comparison of local and aggregated climate model outputs with observed data, Hydrolog. Sci. J., 55(7), 1094–1110.
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.
Bevington, P. R., and D. K. Robinson (2003), Data Reduction and Error Analysis for the Physical Sciences, 3rd ed., McGraw-Hill, Boston.
Essex, C., R. McKitrick, and B. Andresen (2007), Does a Global Temperature Exist?, J. Non-Equilib. Thermo., 1–27.
Frank, P. (2008), A Climate of Belief, Skeptic, 14(1), 22-30.
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.
Jiang, J. H., et al. (2012), Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations, J. Geophys. Res., 117(D14), D14105, doi: 10.1029/2011jd017237.
Kiehl, J. T. (2007), Twentieth century climate model response and climate sensitivity, Geophys. Res. Lett., 34(22), L22710, doi: 10.1029/2007gl031383.
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.
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.
Pyle, J., et al. (2016), Chapter 1. Ozone and Climate: A Review of Interconnections, in Safeguarding the Ozone Layer and the Global Climate System: Issues Related to Hydrofluorocarbons and Perfluorocarbons, edited by T. G. Shepherd, S. Sicars, S. Solomon, G. J. M. Velders, D. P. Verdonik, R. T. Wickham, A. Woodcock, P. Wright and M. Yamabe, IPCC/TEAP Geneva.
Soon, W., S. Baliunas, S. B. Idso, K. Y. Kondratyev, and E. S. Posmentier (2001), Modeling climatic effects of anthropogenic carbon dioxide emissions: unknowns and uncertainties, Climate Res., 18, 259-275.
Taylor, B. N., and C. E. Kuyatt. (1994), Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, *NIST Technical Note 1297 Rep.*, 20 pp, National Institute of Standards and Technology, Washington, DC.
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.