Patrick Frank 12 August 2018

PLoS Manuscript PONE-D-18-14400

Response to the Editorial Review

Summary Response

Scientific points:

1. Major editorial lacunae

1.1 The Editor nowhere addressed the core of the study: GCM global air temperature projections are linear extrapolations of forcing. Therefore, they are subject to linear propagation of error.

This point is repeated throughout the manuscript (Abstract lines 22,23; 213-215; 507-511; 745-750; 950-952), but the Editor completely ignored it.

The first (linearity) is fully demonstrated. The second (error propagation) necessarily follows. This is QED for the study.

1.2 The Editor ignored the central fact that the annual average increase in CO_2 forcing (0.035 Wm^{-2}) is 114 times smaller than the average lower limit of model resolution ($\pm$4 Wm^{-2}), although this was pointed out repeatedly: lines 26 (Abstract), 691-693, 828-833, and 1047-1049.

It is impossible that GCMs can resolve a perturbation $\pm$114-fold below their lower limit of resolution.

1.3 The Editor ignored the presentations of model calibration error (lines 74-86, 102, 222-228, and all of Section 2.3 (515-632)), of error propagation (lines 60, 156-185, 731-733, and all of Section 2.4.2 (735-841)) and the problem of step-wise repetitive initial value error (lines 177-180, 735-738, 853-906 including Scheme 1, and 985-1002).

1.4 In short, the Editorial review ignored the substantive content of the analysis and provided no analytical content of its own. It is dismissive rather than critical.

2. Major critical mistakes (with reference to Full Response items below):

The Editor:

1. mistakenly claimed that the manuscript analysis does not disprove the IPCC consensus, items 3.1.1 through 3.1.3.
2. confused precision with accuracy, items 3.2.1, 3.2.1.3, and 6.1.2;
3. mistakenly supposed that climate models “work,” when in fact they do not, items 3.2.1, 3.2.1.1, 3.1.1.2, 3.2.2.1 through 3.3.2.2, and 6.1.1 through 6.2.2;
4. uncritically dismissed clear evidence of model theory error, item 3.3.2.1;
5. apparently does not know that ocean models do not converge, item 3.3.2.3;
6. does not realize that measurement errors necessarily propagate into model simulations and produce projection uncertainty, item 3.3.4;
7. made unfounded claims of "excessive language," items 3.4.1 through 3.4.3;
8. does not distinguish between inductive engineering models and deductive scientific models, item 4;

9. ignored the demonstration that GCM air temperature projections are linear extrapolations of forcing, item 5.2. In this, he ignored the very demonstration that forms the core of the analysis;
10. showed no understanding of error propagation, item 5.1, 5.3, 5.4.1 and 5.4.2;
11. asserted a "lack of evidence" after completely ignoring the abundance of evidence, item, 7.4.

3. Other critical lacunae:

1. clearly did not consult the provided and overwhelming evidence that climate modelers display no expertise in error analysis, items 1.1, 1.2.1, 1.4.1.1, 1.4.1.2;
2. displayed an unprofessional personal umbrage, items 1.4.2, 1.4.3, and 1.4.4;
3. has a professional conflict of interest with the manuscript, item 1.4.3.2;
4. published climate modeling papers that, like all others, neglect physically valid uncertainty estimates, item 1.4.3.2;
5. dismissed the critically central discussion of neglected physical accuracy and the analytical representation of climate models as a "lack of focus," items 2.1.1.1 through 2.1.1.3;
6. dismissed the logical progression in physical reasoning as "very bad balance," item 2.1.3;
7. inverted the support given by references 56 and 57, item 2.1.5.

Full Response: The Review Editor is quoted in full (*italics*) below, followed by the indented author response.

The Editorial review began with informal comments. Many of these comments required a response

1.1 *The work that you submitted for consideration is challenging for several reasons. First you acknowledge that your work has been rejected for publication in other journals with up to 27 reviewers having checked previous versions. As an Editor this is something that I appreciate, being honest and transparent about the study.*

1.1 The author is grateful for the Editor's appreciation. However, the cover letter also provided evidence that the prior reviewers made mistakes of such a fundamental nature as to obviate their critical force.

A link was provided to the prior reviews documenting their extraordinary errors (https://uploadfiles.io/f5luc). However, there is no indication anywhere in his commentary that the Editor ever examined the documentation.

1.2 *Secondly you request that we ban from the review process some of the more respected experts in the field of your work and what is more, you suggest that only experts in one field have got the understanding to judge your work. This complicates the assessment of your work.*

1.2.1 The field of work is physical error analysis. Climate modelers are not expert in error analysis (*cf.* 1.1).

The cover letter transmitted this point, with evidence that included more than two dozen examples showing that climate modelers are apparently untrained in physical error analysis.

Also mentioned was the very serious professional and economic conflicts of interest climate modelers have with a manuscript demonstrating that climate models have no predictive value.

The Editor has not acknowledged either of these grounds.

1.2.2 The author suggested physical error experts from two fields: condensed matter physics and analytical chemistry; not the Editor's one.

1.3 *Finally your manuscript questions very well established and broadly accepted research results. Going against the consensus is not easy and therefore your previous requests are partially acceptable.*

1.3 The author acknowledges "*broadly accepted,*" however "*well established*" requires an extensive prior body of physically valid demonstrations.

No such body of demonstrations exists. Climate modeling studies rely on precision with neglect of physical accuracy. They are not physically valid demonstrations no matter how many nor how broadly accepted.

The Editor apparently does not recognize this critically central distinction.

Nevertheless, neglect of accuracy leaves climate modeling results physically unestablished.

The manuscript Introduction extensively documented the problem of neglected accuracy. In referenced previous work, the author showed that consensus climatology neglects accuracy as a general rule (1, 2, 3).

1.4.1 *I have to say that despite all of this your letter sounds slightly arrogant.*

1.4.1.1 It is not arrogant to notice that prior climate modeler reviewers had repeatedly mistaken uncertainty statistics to be physically real temperatures or energetic perturbations.

Further they,

- failed to distinguish between accuracy and precision.
- confused variation about an ensemble mean with propagated error.
- supposed offsetting model calibration errors ensure accurate predictions of evolving physical states.

These mistakes reflect an extraordinary lack of understanding.

1.4.1.2 The files at https://uploadfiles.io/f5luc show previous reviewers, almost certainly climate modelers all, uniformly evidenced complete ignorance of physical error analysis, of the meaning of uncertainty, and of the notion of calibration. This is not an opinion but a fact in evidence.

These are seriously fundamental mistakes. They indicate naïveté in the student and incompetence in the trained. It is not arrogant to notice this.

1.4.2 *Being very open about this, I do not share your view and I can not accept your statement saying that for example, I do not have the capacity to assess your work.*

1.4.2 My cover letter did not name Dr. Añel. Nor was there any direct or implied criticism of him or of his ability to review the work.

The author regrets the Editor's aggrievance. However, the editor has chosen to be personally aggrieved over an objectively true observation concerning a group of which he is not a member.

Nevertheless, given the author's uniform experience that climate modelers lack expertise in physical error analysis, documented above, it seems very justifiable to request they not be recruited to review a manuscript that deals exclusively with physical error analysis.

1.4.3 *I have got a BSc, an MSc and a PhD in Physics, with specific courses on the study of error measurement. On one of the shelfs in my office I have a book dealing specifically with accuracy and reliability and I have cited it in some of my works. The fact that people like me has specialized on Earth Physics does not mean that we do not have extensive training on radiative balance, chemistry, etc.*

1.4.3.1 The above is an argument from authority. In any case, the author's observation concerned climate modelers, not Earth Physicists. Much of Dr. Añel's work seems to concern analysis of measurement fields. His self-identification with climate modelers seems therefore thin.

1.4.3.2 However, given the Editor's argument, the author has here taken the opportunity to examine recent of the Editor's papers that included climate simulations. These papers indicate the Editor has a professional conflict with the author's analysis.

Figure 1 in the Editor's 2016 paper with Chiodo, et al., shows projections of air temperature derived from climate simulations using the "community Earth system model (CESM) version 1.0.5" (4). The temperature projection displayed in Figure 1 has no physically valid uncertainty bars; such as those that might derive from physical simulation errors such as described in the 2013 paper of Lauer and Hamilton (5), among others (6, 7, 8, 9).

There is no indication in Chiodo, et al., 2016 of the uncertainty in projected air temperature that arises from the known problem that even tuned climate models nevertheless differ by factors of 2 or 3 in climate sensitivity (10, 11).

Similarly, no physical uncertainty indicators appear in Chiodo, et al., 2016 remaining Figures 2-6.

The term "error" itself does not appear in Chiodo, et al., 2016. Projection uncertainty is judged in terms of consistency among model simulations, which is a measure of precision, not accuracy. Physically valid uncertainty limits do not appear in the climate model expectation values reported in the Editor's co-authored work in Toohey, et al., 2013, either (12).

1.4.4 *You should have it into account, because when submitting a work, to begin a letter saying that your are the smartest person in the room could condition an editor or reviewer.*

1.4.4.1 The author made no such saying.

The author merely observed that climate modelers appear untrained in physical error analysis. Their mistakes were summarized in the cover letter. A link was provided to extensive objective evidence sustaining that conclusion (see also 1.4.1.1 and 1.4.1.2 above).

The author then requested a reasonable accommodation, namely that climate modelers be not recruited to review a manuscript dealing centrally with physical error analysis.

The Editor has met a reasonable request with a statement of personal resentment. It is difficult to believe the Editor could remain critically dispassionate in assessing the manuscript. One notes from 1.4.3.2 above, that the manuscript study removes significance from the Editor's own work.

1.4.5 *Now I would like to make clear that this has not impacted my evaluation of your manuscript.*

1.4.5 The author's email response of 23 July pointed out certain review content that renders the Editor's statement in item 1.4.5 unfortunately very moot (see also item 2.1.5 below).

2 *That said, there are several reasons why I consider that your work is far of being publishable in its current form. Some of them are more 'formal' and others more 'scientific'.*

Beginning with the formal points:

2. The Editor's criticisms are addressed point-by-point below.

2.1.1- *lack of focus on the topic that you address: from the title and abstract your manuscript is about error propagation in climate models, however in a 50 pages main manuscript, the first 25 have nothing to do with it.*

2.1.1 The first 25 pages include the Introduction (through page 11), the derivation and validation of the emulation equation in terms of climate models (through page 18), and the verification of the emulation equation against climate model air temperature projections (through page 25).

Concisely, the Editor has ignored the absolutely critical necessity of demonstrating the analytical relevance of an emulation model before beginning the object analysis. Results Sections 2.1-2.3 (pp. 12-25) establish this analytical relevance.

The error analysis and the rest of the study would be left bereft of any direct relevance to climate models without this opening demonstration.

Items 2.1.1.1 through 2.1.1.3 below describe these analytical connections more fully. In sum, they show the Editor's 2.1.1 to be without analytical merit.

2.1.1.1 The Introduction:

- establishes the necessity of a physical measure of model reliability, i.e., physical accuracy.
- documents that the climate modeling community focuses instead on statistical precision.
- introduces the neglected idea of model calibration.
- outlines the meaning and impact of calibration error.

- explains why tuning to obtain offsetting errors does not remove projection uncertainty (13).
- describes the standard propagation of calibration error.

All of these concepts are critically central to the subsequent analysis.

Lines 201-220 (pp. 10-11) introduce the necessary analytical connection between the emulation equation and the internal dynamics of climate models in their production of global air temperature projections.

However, the Editor has dismissed this necessary connection as irrelevant.

2.1.1.2 Section 2.1 (pp. 12-18) uses a mean-free-path argument to derive the fraction of GHG warming due to CO_2 as relevant to climate models. This model-relevant fraction establishes the analytical connection between the emulation equation and advanced climate models.

The model-specific variation of this fraction refers each individual emulation to the internal state of a foundational climate model.

Manuscript Figure 1a and associated text shows that the logarithmic relation of CO_2 forcing begins at about 1 ppm concentration. This result justifies the asymptotic extrapolation of Figure 1b.

All this is made eminently clear throughout Section 2.1. For example,

- line 252 (p. 12): "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 only to general circulation climate models*."
- line 344: "Carbon dioxide forcing ... departs from linearity with band saturation."
- line 365: Section "*2.1.3 The fractional GH Temperature due to wve CO2 forcing, as relevant to climate models."*

It is very difficult to understand how the Editor can label all of this as irrelevant.

2.1.1.3 Page 18, Section 2.2: "*A General Emulation of the GASAT Projections of Climate Models*"

The section heading itself makes its analytical relevance eminently clear.

Section 2.2 introduced the full emulation equation. Figure 2, Figure 3, and Figure 4 (pp. 20-24) demonstrate that the equation successfully emulates the internal dynamics of CMIP2+, CMIP3, and CMIP5 climate model air temperature projections.

How can it possibly be thought that Sections establishing the relevance to advanced climate models and demonstrating the thoroughly successful emulation of their outputs, "*have nothing to do*" with an analysis of climate models?

2.1.2 *This 25 initial pages should be a different manuscript itself. They are about a completely different topic, computation of the radiative balance in the troposphere. There is no point on presenting all this material together.*

2.1.2 Following from response items 2.1.1*ff* above, it should be clear that manuscript Sections 1 through 2.3 are the critical foundation of the study. The Editor's dismissal as irrelevant ignores clear analytical logic.

It is very difficult to understand how someone trained to a Ph.D. in Physics can dismiss such obvious physical reasoning.

2.1.3 - *bad organization: your manuscript has a very bad balance, only 3 sections in the main part and then you go up to 10 subsections that you actually need to make the point sometimes*

2.1.3 It is again difficult to follow the Editor's complaint. The manuscript organization is:

- 1 Introduction
- 2. Results
  - 2.1. Mean free path, water-vapor-enhanced (wve) CO_2 forcing, and a fractional greenhouse temperature relevant to GCMs
    - 2.1.1 The mean free path of 15 μm radiation.
    - 2.1.2 The onset of CO_2 forcing.
    - 2.1.3 The fractional GH Temperature due to wve CO_2 forcing, as relevant to climate models.
  - 2.2. A General Emulation of the GASAT Projections of Climate Models.
  - 2.3 CMIP5 model calibration error in global average annual total cloud fraction (TCF).
    - 2.3.1 The structure of CMIP5 TCF error.
  - 2.4 A lower limit of uncertainty in the modeled global average annual thermal energy flux
    - 2.4.1 The magnitude of CMIP5 TCF global average atmospheric thermal energy flux error.
    - 2.4.2 Error propagation and the uncertainty in projected GASATs
    - 2.4.3. Differencing from a base-state climate does not remove systematic error.
    - 2.4.4 Predictive Reliability.
- 3. Summary and Discussion

Each section or subsection informs the reader of the topic there at hand. Even a casual inspection of the Section list shows that the topical organization is analytically sequential, logically connected, and smoothly continuous.

How is this a bad organization?

Should the Editor prefer, perhaps the subsections can be promoted to full sections. But then the organization could conceivably be judged bad because, e.g., 12 sections are required to make the author's point.

2.1.4 - *lines 112-137 are completely irrelevant here*

These lines can be removed; although GCMs are engineering models.

2.1.5.- *as an example, in line 273 you cite 56 and 57 and in my view these works do not support your claim, indeed they could be interpreted as serious criticisms to what you expose.*

2.1.5 The following response shows the Editor is mistaken, and is a modified version of that given in the authors email of 23 July 2018.

Manuscript lines 270-273 say this: “The extent of radiative forcing by various concentrations of atmospheric CO_2 is primarily determined by the absorption of radiant energy in the extended 15 µm IR radiation band originating from the warm terrestrial surface (1, 56, 57).”

Reference 56, 57 are to Houghton 1995, 2005, respectively.

The following quotes show these references exactly support manuscript lines 270-273, rather than criticize them.

Houghton 1995 (reference 56) agrees that:
“*the absorption and emission of radiation in the atmosphere by the “greenhouse” gases, especially by carbon dioxide … lead to greenhouse warming of the lower atmosphere and the surface*” and “*most of the absorption by carbon dioxide of radiation from the surface occurs within 30 m of the surface…*”

Houghton 2005 (reference 57) says (abstract):
“*… ‘greenhouse gases’, of which the most important is carbon dioxide. Such gases absorb infrared radiation emitted by the Earth’s surface and act as blankets over the surface keeping it warmer than it would otherwise be.*”

And within the text of Houghton 2005,
“*The greenhouse effect arises because of the presence of greenhouse gases in the atmosphere that absorb thermal radiation emitted by the Earth’s surface ...*”

The plain message of references 56 and 57 supports manuscript lines 270-273, and the subsequent analysis. They are not "*serious criticisms*."

The Editor then wrote, “*Now I would like to make clear that this has not impacted my evaluation of your manuscript*.”

Nevertheless, the Editor's evaluation inverted the clear meaning of references 56 and 57. They support the author's position.

3. *Scientific points:*

3.1 - *great claims need great support/evidence: basically you state that the IPCC is wrong. The point is that if you want to prove it (and to go against the broadly accepted consensus) you need to refute at least some of their claims, and in your work you do not do it. I can accept that it could not be fair from the formal point of view to request you to demonstrate that others are wrong, but at least you should strut your results with demonstrations of the style 'and this probes that the result obtained in previous works [X] is wrong because of this and this and this';*

3.1.1 In contrast, the author's work has in fact shown that the consensus is wrong and has refuted IPCC claims. The entire position of the IPCC depends upon the causal diagnosis rendered from climate models. The manuscript study completely refutes the causal reliability of these selfsame climate models.

In sequence, the manuscript shows that:

- GCM air temperature projections are merely linear extrapolations of GHG forcing (pp. 20-24, 40, and 46, as well as SI pages 5-11).
- CMIP5 long wave cloud forcing (LWCF) error is systematic and correlated among climate models (pp. 27-31)
- GCM annual LWCF error is ±114 times larger than the annual increase in CO_2 forcing (pp. 34. 41 and 51).
- Propagation of GCM LWCF error yields ±15 C of uncertainty in projected global air temperatures (pp. 36-40, 46).
- GCM air temperature projections are without any physical meaning and are unreliable (p. 41, lines 839*ff*, p. 45, lines 917*ff*, p. 46, lines 941*ff*, and SI p. 30, Section 10.2).

These demonstrations fully refute the consensus position of the IPCC.

3.1.2 It is impossible to resolve the effect a perturbation that is ±114 times smaller than the lower limit of model resolution.

Therefore, the study shows that GCMs have no predictive capacity. They can assign no causal meaning to air temperature changes. The IPCC position is refuted.

There is no discernable human fingerprint in global climate change.

The refutation is based upon the demonstration that the climate models cannot resolve the effect of CO_2 on the climate, and in fact cannot resolve this fingerprint even if it does exist.

These demonstrations are definitive.

3.1.3 The Editor asked that the author show, "*previous works [X] is wrong because of this and this and this*"? The Editor's *"this and this and this*, are the listed demonstrations of GCM error and lack of resolution. How did the Editor miss this?

Propagated longwave cloud forcing error shows that air GCM temperature projections are physically meaningless. Climate models are demonstrated to have no predictive value.

This demonstration refutes the IPCC claim of evidence that CO_2 has changed global air temperature. Absent climate models, there is no basis for that claim.

That is, climate model attribution or validation studies are without meaning.

The conclusion of the study, "*an anthropogenic temperature signal cannot have been, nor presently can be, evidenced in climate observables*," is entirely justified by the critical content.

The conclusion is demonstrated beyond any reasonable doubt. The IPCC position is explicitly refuted.

3.2.1 - *you do not address a big point in all the debate. For example, empirical observations all over the world evidence that despite any issue with error propagation the method of using ensemble mean works.*

3.2.1 By "*empirical observations*" the Editor is apparently implying that the statistical skill of the ensemble mean hindcasts of tuned models proves the models "work" (are accurate).

The Editor's implication confuses precision with accuracy.

Statistical skill provides no assurance that the underlying physical climate state is simulated accurately. Therefore, statistical skill does not mitigate predictive uncertainty.

The mistake of conflating precision with accuracy is widespread in the climate modeling community. This problem is discussed in detail in the Introduction, but is ignored here by the Editor.

Indeed, his equation of statistical skill to physical accuracy indicates that, Ph.D. notwithstanding, the Editor is evidently confused about the source of meaning in science.

The response items below address the three central issues: the simulated climate energy state, forecast reliability, and physical meaning. These will show that the Editor has not thought deeply enough about the problem of projection uncertainty. And that the Editor's confidence is badly misplaced.

3.2.1.1 <u>The problem of simulated climate states versus physically real climate states</u>.

Figure 3.2.1.1 below is adapted from Figure 1 of Rowlands, et al., 2012 (14).

The Figure shows the spread of HadCM3L air temperature projections when certain parameters are varied across their range of uncertainty. The vertical red line and horizontal black lines were added by the author.

The added horizontal black lines show that the HadCM3L will project the same temperature change for multiple climate energy states. For example, the bottom lines shows that with different sets of parameters the HadCM3L projects a constant 1 C temperature increase for every single annual climate energy state between 1995-2050, relative to 1985.

The scientific question is, which of the thousands of 1 C simulations is the physically correct one?

Likewise, a constant 3 C increase in temperature is projected to result from every single annual climate energy state between 2030-2080. Which one of those is correct?

Which of these is correct? No one knows.

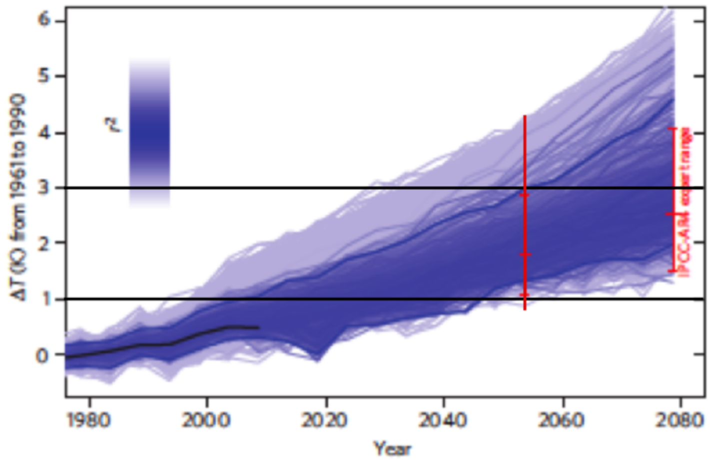

Figure 3.2.1.1. From the original Legend: "Evolution of uncertainties in reconstructed global-mean temperature projections under SRES A1B in the HadCM3L ensemble. Blue colouring indicates goodness-of-fit between observations and ensemble members, plotted in order of increasing agreement (light to dark blue). Black line, the evolution of observations, ..."

The identical logic applies to the vertical red line. This shows that the HadCM3L will project a multitude of temperature changes for a single climate energy state (the 2055 state).

Likewise, every single annual climate energy state between 1976-2080 has dozens of simulated air temperatures associated with it.

Again, none of the different parameter sets producing these simulated temperatures is known to be any more physically correct than any other set. None of the simulated climate states is known to be physically correct. There is again no way to decide which, among all the different choices of projected annual air temperature, is physically correct.

This set of examples is typical of advanced GCMs. None of them can produce a unique solution to the problem of the climate energy state.

No set of model parameters is known to be any more valid than any other set of model parameters. No projection is known to be any more physically correct (or incorrect) than any other projection.

This means, for any given projection, the internal state of the model is not known to reveal anything about the underlying physical state of the true terrestrial climate. More simply, climate models cannot tell us anything at all about the physically real climate, *at the level of resolution of forcing from CO_2 emissions.*

The same is necessarily true for any modeled climate energy state, including the modeled energy states of the past climate.

It does not matter that any given GCM hindcast may reproduce the known past air temperature trend since, e.g., 1880. The underlying climate energy states remain unknown. Therefore, even a statistically skilled hindcast must have a large physical uncertainty associated with it.

The uncertainty reflects the fact that the underlying simulated physical energy state of the climate is not known to be correct.

Every individual projection therefore has a large uncertainty associated with it from parameter uncertainties alone, i.e., even without regard for physical error.

In Figure 3.2.1.1, every single projection should properly include an uncertainty envelope from propagated error. The uncertainty in the ensemble mean is then the root-mean-square of the uncertainties of the individual projections.

None of these accuracy measures was included in the published Figure.

3.2.1.2 <u>The problem of forecast skill</u>.

Figure 3.2.1.2 is from Fyfe, et al., 2013 (15). This Figure compares projected temperatures with observed temperatures. The correspondence is poor. The Figure makes it quite clear that even tuned models do not work in the sense meant by the Editor.

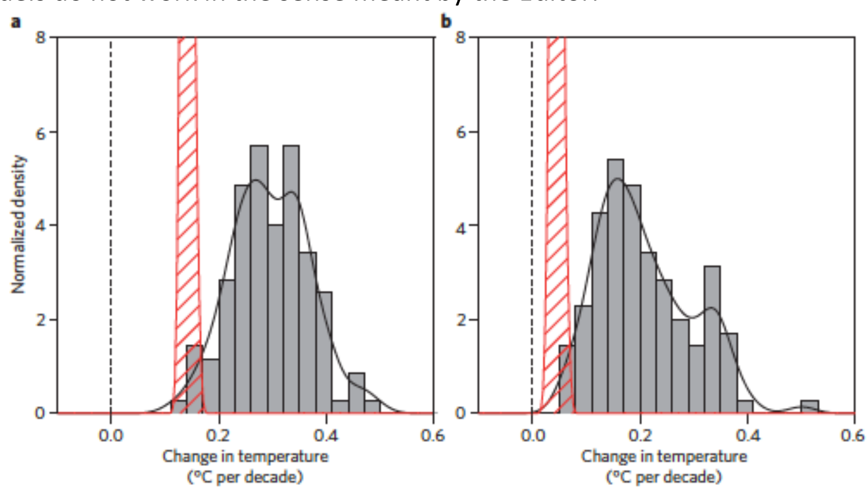

Figure 3.2.1.2. From the Legend: "Trends in global mean surface temperature. **a**, 1993–2012. **b**, 1998–2012. Histograms of observed trends (red hatching) are from 100 reconstructions of the HadCRUT4 dataset1. Histograms of model trends (grey bars) are based on 117 simulations of the models, and black curves are smoothed versions of the model trends."

Next, Anagnostopoulos, et al., 2010, tested the simulated temperature and precipitation fields of multiple climate models against observations at regional and continental scales (16, 17). The models performed poorly. They found that projected air temperatures,

> "*differ substantially from the observed time series. ... The results are worse for precipitation. ... The results for AR4 are no better than those for TAR. In some, the annual mean temperature of the USA is overestimated by about 4–5 ℃ and the annual precipitation by about 300–400 mm.*"

They concluded,

> *"It is claimed that GCMs provide credible quantitative estimates of future climate change, particularly at continental scales and above. Examining the local performance of the models at 55 points, we found that local projections do not correlate well with observed measurements. Furthermore, we found that the correlation at a large spatial scale, i.e. the contiguous USA, is worse than at the local scale."*

These tests against observations do not support the Editor's suggestion that, "*empirical observations all over the world evidence that despite any issue with error propagation the method of using ensemble mean work."*

3.2.1.3 The problem of physical meaning.

Finally, Figure 3.2.1.3 is from the author's own work. It also appears as Figure S12 and Figure S13 in the Supporting Information.

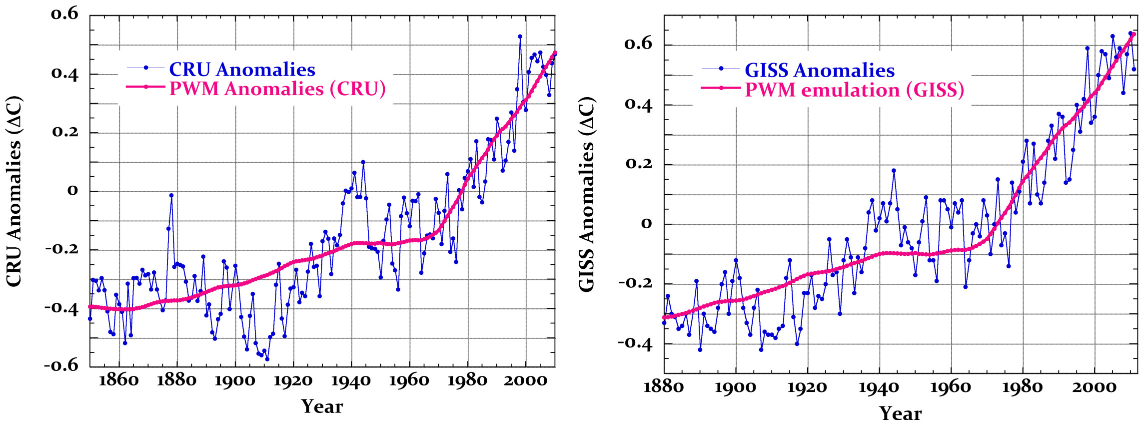

Figure 3.2.1.3. Points and thin blue lines are the global air temperature record as compiled by, left, the UK Met Office, and; right, by the NASA Goddard Institute of Space Studies. The red lines are projections made using the manuscript emulation model (ms. eqn. 6) and the requisite IPCC forcings (see text).

The forcing used in the Figure above were the combined annual “CO_2” plus “GHG Other” plus “Aerosol (Total)” forcings of AR5 WG1 Annex II Table AII.1.2, over the full historical period (18).

That is, using the accepted forcings, emulation equation 6 can reproduce the full centennial air temperature record. The emulations display high statistical skill.

Emulation equation 6 therefore meets the Editor's definition of works in his comment 3.2.1. If we are to believe the Editor in his comment, the successful emulations above prove that eqn. 6 provides a physically believable reproduction of past air temperature, and is therefore a useful source of reliably projected air temperatures.

Clearly, however, equation 6 includes no physical theory of the terrestrial climate. A conclusion of accuracy is physically untenable.

Instead, this demonstration shows that statistical skill in fact provides no confidence of physical accuracy at all.

Nevertheless, the very same refuted supposition that statistical skill equates to physical accuracy forms the basis of the Editor's confidence in the scientific truth content of general circulation climate model air temperature projections.

3.2.2 *If we assume that your work is ok, at lest as a summary in your conclusions for a well balanced discussion it is necessary to state something like 'despite all these problems and limitations over the last 30 years the models are proved to be useful and to be able of forecasting with a very high skill in 1980s the mean climate for today' (empirical evidence);*

3.2.2.1 Climate models do not have forecast skill. Published work shows this.

For example, Kravtsov, 2017 observes that (19), "*Here we show that state-of-the-art global models used to predict climate fail to adequately reproduce such multidecadal climate variations. ... Therefore, our ability to interpret the observed climate change using these models is limited*."

Kravtsov goes on to note that the problem is with, "*biases in the models' forced response or models' lack of requisite internal dynamics, or a combination of both.*"

That is, there are inadequacies in the physical theory deployed within the models. The present manuscript makes the same diagnosis from model cloud errors.

3.2.2.2 Prof. John R. Christy presented Figure 3.2.2.2 during testimony before the U.S. House Committee on Science, Space, and Technology (20).

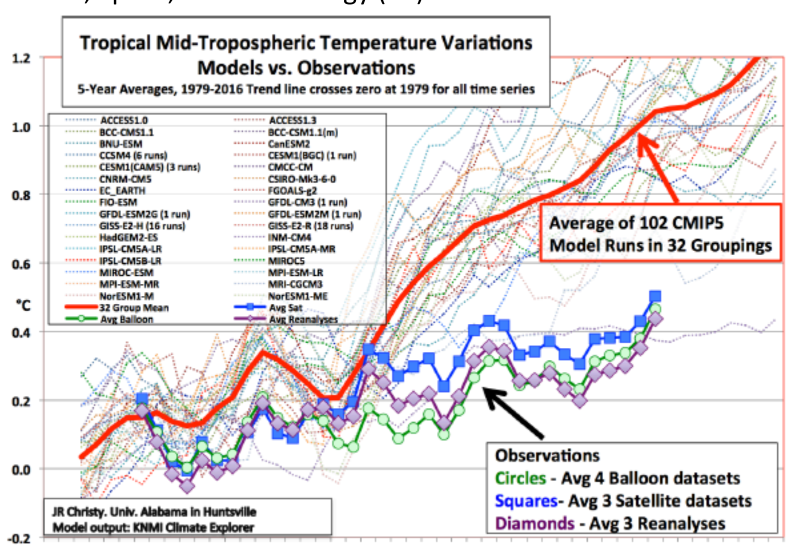

Figure 3.2.2.2, from the original Figure Legend: "*Five-year averaged values of annual mean (1979-2016) tropical bulk T_{MT} as depicted by the average of 102 IPCC CMIP5 climate models (red) in 32 institutional groups (dotted lines). ... Observations are displayed with symbols.*"

"T_{MT}" is the mid-tropospheric air temperature. Figure 3.2.2.2 shows that climate model global air temperature simulations began departing from observations immediately following the anomaly zero point.

Prof. Christy concluded that, "The scientific conclusion here, if one follows the scientific method, is that the average model trend fails to represent the actual trend of the past 38 years by a highly significant amount."

Prof. Christy also noted that the IPCC buried this CMIP5 failure within the Supplemental Material of Chapter 10 of the AR5, where it would not be noticed. He also observed that the data were plotted in a way to make their meaning "*difficult to understand and interpret."*

Prof. Christy concluded, *"I demonstrate that the consensus of the models fails the test to match the real-world observations by a significant margin. As such, the average of the models is considered to be untruthful in representing the recent decades of climate variation and change, and thus would be inappropriate for use in predicting future changes in the climate or for related policy decisions.*"

John Christy is the Distinguished Professor of Atmospheric Science and Director of the Earth System Science Center at the University of Alabama in Huntsville.

3.2.2.3 Following from response items 3.1 and 3.2.1.1 through 3.2.2.2, it should be clear that the Editor's confidence is very poorly-placed.

The same should be clear following any knowledgeable reading of the error analysis presented in the manuscript.

Climate models are not able to forecast the climate state at all. Projection ensemble averages from tuned models are not at all representations of accuracy. They are merely statistically skilled, as the model tuning guarantees (21).

3.3.1 - *I do not agree that we are facing 'physical' errors as you claim.*

3.3.1.1 The author does not claim physical errors. Lauer and Hamilton [(5) manuscript reference 80] demonstrate GCM physical errors. As noted above, others also show GCM physical errors (15, 16, 19, 20).

Figure 3.3.1.1 below is the Taylor diagram from Lauer and Hamilton Figure 3, showing the GCM physical errors in longwave cloud forcing. They compared 20 years of simulated clouds relative to observations, from twenty-seven CMIP5 climate models. All of these GCMs had previously been tuned to reproduce hindcast target observables.

All twenty-seven CMIP5 GCMs made significant errors compared to observations. These are model calibration errors, relative to the observational standard.

The inter-model error correlations of manuscript Table 1 indicate all twenty-seven GCMs make the same error. The cloud forcing errors are systematic and represent mistaken expressions inherent in the models themselves.

These systematic model errors will necessarily enter into every single step of any climate simulation. They must be propagated step-by-step through model simulations as an estimate of the reliability of any climate projection.

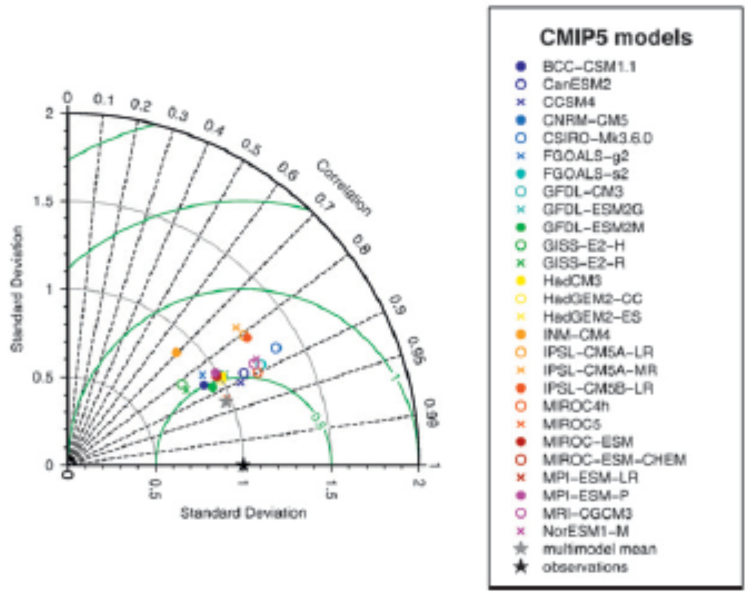

Figure 3.3.1.1, from the original Legend: "Taylor [diagram] showing the 20-yr annual average performance of the ... CMIP5 models for [Top of the Atmosphere longwave cloud forcing] as compared to satellite observations."

Figure 3.3.1 displays a 20-year annual mean of physical error for each model. The black star at point 1,1 represents observations. Every model simulation exhibits a significant departure. The manuscript analysis utilizes the reported ± 4 Wm^{-2} root-mean-square error (rmse) of the 27 model annual error means. This annual error mean-of-means is uncertainty in the simulated tropospheric longwave thermal energy flux. It is present in every single step of a simulation.

Any random cloud-forcing error component will have been reduced by $\sqrt{(20\times 27)}$ = 23-fold. Thus, the ± 4 Wm^{-2} consists almost entirely of systematic model error.

This error does not average away in an ensemble mean projection. It cannot be ignored. It must condition any climate projection simulated using CMIP5 models.

3.3.1.2 Likewise, model physical errors are reported in Dolinar, 2015 (clouds), Chase, et al., 2004 (surface air temperature), Collins, et al., 2011, 2012 (feedbacks and forcings), and Gleckler, 2005 (surface energy balance) as arbitrarily chosen examples (6, 8, 22, 23, 24).

Soon, et al., assayed climate model physical errors, and found they summed up to nearly 150 Wm^{-2} (25). That is, climate models misallocated about half the solar energetic flux energy entering the climate.

All of these physical errors have been ignored in the published literature. None has ever been used to assess the physical uncertainty attending the projections of climate models.

3.3.2 *Any deviation from reality is not because of the physical system here, it is because of the tool and measurements and completely independent of the level of understanding of the completeness of the theory. They are mathematical/statistical deviations. In a text dealing so much with accuracy making clear this difference is important.*

3.3.2.1 The Editor is not correct. The error correlation matrix in manuscript Table 1 (p. 31, line 627) shows beyond any doubt that GCM cloud error is systematic, is shared among models, and must reflect an error in the deployed physical theory.

See also the discussion under Figure 3.3.1.1 above. Any random component is minimal in the 20-year, 27-model average of CMIP5 ± 4 Wm^{-2} tropospheric longwave thermal energy flux error. T

This error is inherent in the CMIP5 models and clearly arises from an inadequate physical theory.

The possibility that the error is random is also eliminated by its extraordinarily high inter-model correlation (correlation $R > 0.95$).

The analysis showing that model cloud error is indeed systematic is found in manuscript lines 563-617; Section 2.3.1 *The Structure of CMIP5 TCF Error* and p. 31, Table 1.

The Editor has merely dismissed this analysis without ever addressing it.

3.3.2.2 Further, Martin, et al., 2010 explicitly acknowledge that GCMs make serious theory-based systematic errors (26).

In view of the Editor's confidence, it is worth quoting Martin, et al., 2010. From the abstract:

> *"The reduction of systematic errors is a continuing challenge for [climate] model development. Feedbacks and compensating errors in climate models often make finding the source of a systematic error difficult."*

Martin, et al., explicitly include cloud error, which is pertinent to the manuscript analysis. From the paper:

> *"The growth of systematic errors in general circulation models (GCMs) remains one of the central problems in producing accurate predictions of climate change for the next 50 to 100 years (Randall et al. 2007). Although great advances in global climate modeling have been made in recent decades (Solomon et al. 2007), **there are still large uncertainties in many processes such as clouds**, convection, and coupling to the oceans and the land surface (e.g., Cubasch et al. 2001; Koster et al. 2004). (author bold)*

"One issue impeding progress is that attributing the growth of systematic errors to the modeling of particular physical processes is notoriously difficult in climate GCMs. This is due to both a lack of observational data and to the nonlinear interactions amongst various physical processes and the errors in modeling them."

Martin, et al., go on to candidly discuss the many systematic errors inherent in global climate models. The phrase "*errors in modeling them*" is an explicit admission of deficits in theory. They are not merely random excursions.

The itemized uncertainties in cloud processes, convection, and coupling to and between oceans and land surfaces exemplify deficiencies in the physics. These are real deficiencies in theory, which the Editor claims are not present.

3.3.2.3 Further, Wunsch, 2002 has pointed out that GCM ocean models do not converge (27). This non-convergence is still true today (28).

What is the physical meaning of a non-converged physical model? Does the Editor recognize that a non-converged ocean simulation includes unaccounted physical errors in a global climate simulation?

3.3.3 *Other fields of study are also dominated by formalism without analytical solutions and not because of this those using results projected consider they wrong.*

3.3.3 Such formalisms are found to be wrong when their predictions are inconsistent with observations. These errors are typically traced to deficiencies of theory (13, 29, 30).

One of the author's recent co-authored papers was the first to severely test solution-phase molecular dynamics theory, and illustrates this test of theory against experiment (31).

3.3.4 *Also the error of a measurement can be known partially a priori because of the instrumental error and this does not mean that the result is considered 'erroneous' in any discipline or field.*

3.3.4 The editor's deficit is quite extraordinary here. Errors in measurements render their magnitude uncertain in every single field of physical science (32, 33, 34).

When measurements are known to include systematic errors, the mean magnitude ceases to be the most probable value. The physically true value is then lost within the uncertainty envelope. Such measurements are known to be erroneous. Those errors propagate into model simulations.

For example, in the JCGM. Evaluation of measurement data, Section 3.3.1,

"*The uncertainty of the result of a measurement reflects the lack of exact knowledge of the value of the measurand (see 2.2). The result of a measurement after correction for recognized systematic effects <u>is still only an estimate of the value of the measurand</u> because of the uncertainty arising from random effects and from imperfect correction of the result for systematic effects (33)*." (author's underline)

This understanding is basic to empirical work.

Measurement errors induce model errors when they enter a simulation. Thus, JCGM note under 3.3.3, *"a "random" component of uncertainty in one measurement may become a "systematic" component of uncertainty in another measurement in which the result of the first measurement is used as an input datum*."

This caution is entirely applicable to climate model simulations. A random error within simulation step "*n*" become systematic when it is within the initial conditions entering into simulation step "*n+1*". This problem of random error propagating into systematic error is entirely neglected in the field.

When measurement errors are systematic, as in the global air temperature record (1, 3, 35), uncertainty is introduced into the simulation from any model using that record as a calibration target.

3.4 *Here you should be aware that in some cases your language use is excessive and this does not help to consider the presentation of your results fair: 'erroneous', 'mortal test', 'negligible CO2'...*

3.4 These terms are considered in turn. Their context of usage determines fairness.

3.4.1 "erroneous" appears in lines 96, 862, 879, and 883. The first usage conditions all the rest: "Models suffering from an incorrect and/or incomplete physical theory produce systematically erroneous results."

How is that statement incorrect or excessive? It is obviously true. The study goes on to show that climate models produce systematic errors in cloud cover.

Line 862: "The continuing theory-bias means that the erroneous flux magnitudes of the initial spin up state are again and further incorrectly partitioned within each subsequent climate state."

The remining two usages are derivative.

The Editor may dislike reading that climate models make theory-bias errors. However, noticing that serious errors exist and have consequences is not excessive.

3.4.2 "mortal tests" appears once, in line 141: "Scientific models are held to the standard of mortal tests and successful predictions outside any calibration bound. The represented systems so derived and tested must evolve congruently with the real-world system."

"Mortal test" here is a synonym for falsification. That word can be substituted, as the Editor likes.

Nevertheless, the idea of theory held to the mortal threat of falsification by experiment is entirely correct within science. Experiment and observation provide a mortal test of physical theory. That process is central to Physics and should not be defensively rejected in the way implied by the Editor.

The author has in mind Einstein's comment concerning his own theory, that, "If the redshift of spectra lines due to the gravitational potential should not exist, then the general theory of relativity will be untenable.", quoted in Popper's "The Unended Quest" [(36), p. 38].

Einstein held General Relativity to the mortal test of prediction and observation. Disproven, Einstein agreed that the theory is dead. Does the Editor disagree with this?

3.4.3 "negligible CO_2" occurs in lines 266, 284, 295, and 371. In every case, the term refers to the negligible forcing exerted by 1 ppm of CO_2.

Thus, line 266: "The condition of negligible CO_2 forcing reveals the fraction of the unperturbed greenhouse temperature produced by water vapor alone, f_{wv}."

The condition of negligible CO_2 forcing occurs when the concentration of atmospheric CO_2 is so low ($\leq$ 1 ppmv), that the mean free path of 15 μm radiation is greater than or equal to the average height of the troposphere.

"Negligible CO_2 forcing" is not a scientific misnomer in this case, nor is it a rhetorical extreme. It is a plain statement of physical fact.

This condition was derived in order to estimate the forcing due to water vapor alone. Thus, line 371: "The absence of CO_2 forcing means the remaining GH effect is due to water vapor alone."

How is it unfair to derive this necessary physical condition and discuss its meaning?

4- *your introduction to parametrization in lines 141-143 is 'obscure'. It is not clear to me what you want to say. What do you mean by not empirically tuned. Of course some parametrizations are empirically tuned, this is why they are parametrizations and not part of the free running of models;*

4. The Editor fails to distinguish between inductive engineering models and deductive scientific models. In the first, parameters are tuned to reproduce empirically known behavior. In the second, parameters are set to their best values. In scientific models, parameters are not tuned.

Engineering models are tuned to reproduce known observables. Climate models are tuned to reproduce known observables. They are inductive models that infer future outcomes by direct reference to past outcomes.

Scientific models make deductive predictions from physical theory that are then tested against observables.

The difference could not be greater, and is fundamental.

Engineering models infer observables only within their tuned bounds.

Scientific models predict observables over the entire physical range that the theory addresses.

In any case, the Editor wanted lines 112-137 to be removed (item 2.1.4 above). Removal of these lines will necessitate also removing lines 139-151, which require 112-137 as context.

5.1- *another issue comes out with your assumption about the cumulative nature of the error.*

5.1 Propagation of error (cumulative nature) is not an assumption. It is a standard method of error treatment in the physical sciences.

The manuscript provided establishing references to Bevington and Robinson (32), to the recommendations of the US National Institute of Standards and Technology (NIST) (34), and to the international standard detailed by the Bureau International des Poids et Mesures (33).

The approach is presented extensively in the manuscript in lines 60, 88-93 and 156-180 with references that fully establish the approach, including with complex models (37, 38).

Abstract sentence 4 provides the full justification: "The linearity of projections justifies linear propagation of error."

Line 214: "As linear extrapolations of forcing, GCM air temperature projections are subject to linear propagation of error, eqn. 2."

Likewise, lines 507-511, 731-733, all of Section 2.4.2 (lines 735*ff*) "*Error propagation and the uncertainty in projected GASATs*," and lines 748-751.

Line 951: "Linear propagation of model error follows directly from their linear extrapolation of GHG forcing."

Finally, lines 1018-1028 provided the analytical justification of the approach in the quote from the JCGM Guide to the Expression of Uncertainty in Measurement.

There is no reasonable doubt about the propagation of systematic model error through the sequential steps of a simulation.

5.2 *This could be a discussion/paper itself.*

5.2 Editor's comment 5.2 is quite extraordinary.

One would never guess from it that the linearity of GCM air temperature projections is established in manuscript Figure 2, Figure 3, Figure 4, and Figure 9, and in Supporting Information Figure S1 through Figure S8.

The evidence of linearity includes the air temperature projections from dozens of advanced climate models.

The Editor has apparently not understood this demonstration, or its significance.

5.3 *Part of this is to accept that the mean climate (or let's say the total cloud fraction) for the step 't+1' depends only on 't' and probably this is not true and it is completely model dependent. There are a lot of other factors that can contribute to TCF values in a model, beginning with a fully coupled land model with biogenic particles, statistical values of parametrized volcanic eruptions, number of chemical species interacting...*

5.3 Climate simulation step 't+1' does not depend only on step 't.'

The simulation step 't' provides the initial conditions of simulation step 't+1.'

Step 't+1' then introduces its own conditions, which may be different from the conditions of step 't,' such as having a greater CO_2 concentration.

For example, suppose step 't+1' includes the sudden aerosol forcing change due to a volcanic explosion. Nevertheless, simulation step 't+1' takes its initial conditions from step 't.' These conditions become modified by the introduced volcanic forcing change within step 't+1.'

Thus, regardless of any forcing changes in step 't+1,' all the simulation errors in step 't' propagate into step 't+1.'

The Editor has ignored the obvious truth of this sequence.

The injection of model error across simulation steps is discussed in detail, in manuscript Section 2.4.2 (line 735ff), lines 851-906, graphically illustrated in Scheme 1, and following line 867 and lines 985-1009.

However, the Editor has not addressed any of this.

5.4 *Why should we assume such additive nature in a model where boundary conditions at t are different than those in t+1?*

5.4.1 The additive nature is not assumed. It is fully demonstrated in Figures 2-4, 8, and 9, and throughout the Supporting Information, and as noted in items 2.1.1.3 and 5.2 above.

For example, manuscript Figure 9 shows the that emulation equation 6 fully reproduced the air temperature projection of the GISS Model II air temperature projections, which remain linear with forcing despite the non-linear excursions produced by volcanic aerosols.

5.4.2 The average CMIP5 cloud forcing error is constant and present in every single model time step, $t_0, t_1, ..., t_n$ during a simulation. This is the meaning of a mean model calibration error. As the simulation boundary conditions change, the model error remains.

An error in model theory that is present in every simulation step, propagates into and through every step, and produces an increasing uncertainty with every step (38, 39, 40).

6- *and just to add one more thing, for example, in lines 616-617 you use a reasoning pretty weak: if you argue that the fact that something works does not mean that it is ok, then you can not admit that statistical improbability is scientific proof of the opposite argument being true.*

6.1.1 The author's argument does **not** start from an agreement that GCMs work but are not OK. The author's argument in lines 616*ff* rest on the demonstrated grounds that GCMs do not work.

The Editor has here transposed his view onto the author, namely that the statistical skill of GCM hindcasts means that they are physically accurate, i.e., that they "work." The author does not hold that view, and nowhere expressed it.

Statistical skill does not provide causal information. Precision does not equal accuracy.

The Editor's argument that GCMs work was falsified in response item 3.2.1, above, and especially in 3.2.1.1 and 3.2.1.3.

6.1.2 In the manuscript Introduction, the conflation of precision with accuracy was thoroughly exposed as incorrect.

Manuscript lines 790-793 again exemplify this distinction in terms of Figure 7.

However, the Editor here revealed that he has apparently remained immune to these demonstrations and to this absolutely critical distinction. Physical science is impossible without it.

6.2.1 Lines 616-617 discuss the correlation of GCM systematic cloud simulation error. However, the Editor represents this correlation as though it were a statistical accident, or the outcome from a deliberate model tuning exercise.

Neither of those representations is correct. In the first place, a probability of 10^{-17} for any $R > 0.90$ correlation allows dismissal of accident. The fact that there are 12 such correlations removes the likelihood of this accident to one in $(10^{-17})^{12}$.

The Editor is apparently arguing an accident that would occur about once in the age of the universe expressed in femtoseconds.

6.2.2 Systematic model error indicates that observations are simulated incorrectly. The systematic error includes unsimulated data residuals.

The very high model error correlations are not merely a statistical improbability. They unambiguously show that <u>*all the GCMs make the same error.*</u> An error shared by all the models is intrinsic to the models.

7.1 *To summarize, obviously my assessment includes much more material that I have written here and these are just some picks representative of what your work includes and I have had into account.*

7.1 The Editor has completely neglected the core of the study. Namely: GCM air temperature projections are merely linear extrapolations of forcing. Therefore, they are subject to linear propagation of error.

There is no doubt that these points are correct. The first is unarguably demonstrated and the second necessarily follows from the first.

This is QED with respect to the conclusion that global air temperature projections are unreliable.

The study logic is so obvious that it is most difficult to understand how a PhD physicist has not grasped it.

The Editor's review included no analytical content at all. It is dismissive rather than critical.

7.2 *However because of all these concerns with the lack of focus on the presentation of your manuscript,*

7.2 Item 2.3.1 above shows the straightforward logical progression of the manuscript.

7.3 unbalanced discussion ...

7.3 The discussion follows the straightforward logic of the analysis in item 2.3.1.

7.4 and lack of evidence to support part of the results that you present,

7.4 A "lack of evidence" that included five manuscript Figures fully establishing the linearity of GCM projections.

A "lack of evidence" that demonstrated the physics of every analytical point.

A "lack of evidence" that provided literature support for every part of the analysis including the longwave cloud forcing error.

And a "lack of evidence" in the thirty-seven pages of the Supporting Information that provided a validating analysis and discussion that include fourteen more Figures.

This, the Editor calls a lack of evidence.

7.5 *your manuscript can not be accepted for publication. For future works I would suggest you splitting this work at least in three different parts that I think are pretty clear.*

7.5 This conclusion is unsurprising, but is not justified by the content of the review. This counter judgment is substantiated in the listed summary items 1 and 2 at the head of this response.

8. *I am sorry that we cannot be more positive on this occasion, but hope that you appreciate the reasons for this decision.*

8 It should be clear, following any fair reading of the above review and response, that the reasons for the decision are not grounded in fact, in critical analysis, or, indeed, in science itself.

References:

1. Frank P. Uncertainty in the Global Average Surface Air Temperature Index: A Representative Lower Limit. Energy & Environment. 2010;21(8):969-89.
2. Frank P. Imposed and Neglected Uncertainty in the Global Average Surface Air Temperature Index. Energy & Environment. 2011;22(4):407-24.
3. Frank P. Negligence, Non-Science, and Consensus Climatology. Energy & Environment. 2015;26(3):391-416.
4. Chiodo G, García-Herrera R, Calvo N, Vaquero JM, Añel JA, Barriopedro D, et al. The impact of a future solar minimum on climate change projections in the Northern Hemisphere. Environmental Research Letters. 2016;11(3):034015.
5. Lauer A, Hamilton K. Simulating Clouds with Global Climate Models: A Comparison of CMIP5 Results with CMIP3 and Satellite Data. J Climate. 2013;26(11):3823-45.
6. Collins M, Booth BB, Bhaskaran B, Harris G, Murphy J, Sexton DH, et al. Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. Climate Dynamics. 2011;36(9-10):1737-66.
7. Jiang JH, Su H, Zhai C, Perun VS, Del Genio A, Nazarenko LS, et al. Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations. J Geophys Res. 2012;117(D14):D14105.
8. Dolinar EK, Dong X, Xi B, Jiang JH, Su H. Evaluation of CMIP5 simulated clouds and TOA radiation budgets using NASA satellite observations. Climate Dynamics. 2015;44(7-8):2229-47.
9. John VO, Soden BJ. Temperature and humidity biases in global climate models and their impact on climate feedbacks. Geophys Res Lett. 2007;34(18):L18704.
10. Kiehl JT. Twentieth century climate model response and climate sensitivity. Geophys Res Lett. 2007;34(22):L22710.
11. Tebaldi C, Knutti R. The use of the multi-model ensemble in probabilistic climate projections. Phil Trans Roy Soc A. 2007;365(1857):2053-75.
12. Toohey M, Hegglin MI, Tegtmeier S, Anderson J, Añel JA, Bourassa A, et al. Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative. J Geophys Res: Atmospheres. 2013;118(20):11,847-11,62.
13. Beven K, Binley A. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes. 1992;6(3):279-98.
14. Rowlands DJ, Frame DJ, Ackerley D, Aina T, Booth BBB, Christensen C, et al. Broad range of 2050 warming from an observationally constrained large climate model ensemble. Nature Geosci. 2012;5(4):256-60.
15. Fyfe JC, Gillett NP, Zwiers FW. Overestimated global warming over the past 20 years. Nature Climate Change. 2013;3:767.
16. Anagnostopoulos GG, Koutsoyiannis D, Christofides A, Efstratiadis A, Mamassis N. A comparison of local and aggregated climate model outputs with observed data. Hydrolog Sci J. 2010;55(7):1094–110.
17. Koutsoyiannis D, Efstratiadis A, Mamassis N, Christofides A. On the credibility of climate predictions. Hydrolog Sci J. 2008;53(4):671-84.
18. IPCC. Climate Change 2013: The Physical Science Basis. Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2013. Report No.: 5.
19. Kravtsov S. Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century. Geophysical Research Letters. 2017;44(11):5749-57.

20. Christy JR. Testimony of John R. Christy to the U.S. House Committee on Science, Space & Technology Washington D.C.: U.S. House Committee on Science, Space & Technology; 2017 [Applying the scientific method to climate models from the IPCC AR5]. Available from: https://science.house.gov/sites/republicans.science.house.gov/files/documents/HHRG-115-SY-WState-JChristy-20170329.pdf.
21. Hourdin F, Mauritsen T, Gettelman A, Golaz J-C, Balaji V, Duan Q, et al. The Art and Science of Climate Model Tuning. Bulletin of the American Meteorological Society. 2017;98(3):589-602.
22. Chase TN, Pielke Sr RA, Herman B, Zeng X. Likelihood of rapidly increasing surface temperatures unaccompanied by strong warming in the free troposphere. Climate Research. 2004;25(3):185-90.
23. Collins M, Chandler RE, Cox PM, Huthnance JM, Rougier J, Stephenson DB. Quantifying future climate change. Nature Clim Change. 2012;2(6):403-9.
24. Gleckler PJ. Surface energy balance errors in AGCMs: Implications for ocean-atmosphere model coupling. Geophys Res Lett. 2005;32:L15708 1-4; doi:10.1029/2005GL023061.
25. Soon W, Baliunas S, Idso SB, Kondratyev KY, Posmentier ES. Modeling climatic effects of anthropogenic carbon dioxide emissions: unknowns and uncertainties. Climate Res. 2001;18(3):259-75.
26. Martin GM, Milton SF, Senior CA, Brooks ME, Ineson S, Reichler T, et al. Analysis and Reduction of Systematic Errors through a Seamless Approach to Modeling Weather and Climate. Journal of Climate. 2010;23(22):5933-57.
27. Wunsch C. Ocean Observations and the Climate Forecast Problem. International Geophysics. 2002;83:233-45.
28. Wunsch C. Do Ocean Models Converge? 2018. p. Personal Communication.
29. Mauritsen T, Stevens B. Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. Nature Geoscience. 2016;8(5):346-51.
30. Zhao M, Golaz J-C, Held IM, Ramaswamy V, Lin S-J, Ming Y, et al. Uncertainty in Model Climate Sensitivity Traced to Representations of Cumulus Precipitation Microphysics. J Climate. 2016;29(2):543-60.
31. Antalek M, Pace E, Hedman B, Hodgson KO, Chillemi G, Benfatto M, et al. Solvation structure of the halides from x-ray absorption spectroscopy. The Journal of Chemical Physics. 2016;145(4):044318.
32. Bevington PR, Robinson DK. Data Reduction and Error Analysis for the Physical Sciences. 3rd ed. Boston: McGraw-Hill; 2003.
33. JCGM. Evaluation of measurement data — Guide to the expression of uncertainty in measurement Sevres, France: Bureau International des Poids et Mesures; 100:2008. Report No.: Document produced by Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG 1).
34. Taylor BN, Kuyatt. CE. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. NIST Technical Note 1297. Washington, DC: National Institute of Standards and Technology; 1994 September.
35. Frank P. Systematic Error in Climate Measurements: the global air temperature record. In: Ragaini R, editor. The Role of Science in the Third Millennium. International Seminars on Nuclear War and Planetary Emergencies. 48. Singapore: World Scientific; 2016. p. 337-51.
36. Popper KR. The Unended Quest: an intellectual autobiography. La Salle, IL: Open Court; 1976.
37. Vasquez VR, Whiting WB. Uncertainty of predicted process performance due to variations in thermodynamics model parameter estimation from different experimental data sets. Fluid Phase Equilib. 1998;142(1-2):115-30.
38. Vasquez VR, Whiting WB. Accounting for Both Random Errors and Systematic Errors in Uncertainty Propagation Analysis of Computer Models Involving Experimental Measurements with Monte Carlo Methods. Risk Analysis. 2006;25(6):1669-81.

39. Allegrini F, Wentzell PD, Olivieri AC. Generalized error-dependent prediction uncertainty in multivariate calibration. Anal Chim Acta. 2016;903:51-60.
40. Draper D. Assessment and Propagation of Model Uncertainty. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):45-97.