Patrick Frank 25 January 2015
Propagation of Error and the Reliability of Global Air Temperature Projections
Manuscript #JOC-14-0623

Author Response to Reviewer 1.

Summary

The reviewer:

- betrays no acquaintance with the distinction between precision and accuracy fundamental to science (items 2, 8 and 9).
- has made object-free criticisms (items 3, 5 and 6).
- has misconstrued the PWM emulator as a fit (item 4). It is no such thing.
- has inadvertently acknowledged the validity of the error analysis (item 4).
- betrays no understanding of the significance or impact of mean model error (items 6, 7 and 9).
- apparently does not realize that projection similarity among models with tuned anti-correlated parameters provides no surety of accuracy (item 10).

The reviewer is quoted in full in italics, followed by the author response.

Detailed Response:

1. *I strongly recommend rejection of this paper, which presents an ill-designed analysis based on invalid assumptions and misconceived ideas about climate and climate models. Sorry. There are many aspects which I think are wrong and I think it's impossible to recover the analysis.*

   1. The author will return to this statement at the end of the response.

2. *For starters, the author thinks that a probability distribution function (pdf) only provides information about precision and it cannot give any information about accuracy. This is wrong, and if this were true, the statisticians could resign.*

   2. The reviewer has ignored the distinction between accuracy and precision drawn in the Introduction Section (p. 4 line 15ff). The author's point is that pdfs of an ensemble of simulations about a model mean alone reveal nothing about physical accuracy. An accuracy metric requires reference to an observable. If the reviewer does not understand that, s/he might consult the International Joint Committee for Guides in Metrology.[ISO/IEC, 2008]

   Alternatively, the reviewer might consult [Bevington, PR and Robinson, DK, 2003], "*Data Reduction and Error Analysis for the Physical Sciences*," where the distinction is made on page 2:

   "It is important to distinguish between the terms accuracy and precision. The accuracy of an experiment is a measure of how close the experiment is to the true value; the precision is a measure of how well the result has been

determined, <u>without reference to its agreement with the true value</u>. (underline added)"

This distinction applies just as forcefully to climate model simulations. Comparison among simulations alone reflects precision, not accuracy. There is no accuracy metric without reference to observables.

Either the reviewer has completely missed the point raised in the manuscript, or else the reviewer does not understand a distinction absolutely fundamental to science.

3. *He presents a statement about hindcasts in the paper which makes me think that he does not know what he is talking about.*

3. This criticism has no specification to anything in the manuscript, leaving it empty of content. What "*statement about hindcasts*"? The reviewer does not say. Nevertheless, the author searched the manuscript for "hindcast" to examine possible misstatements. Examples after the Abstract include this first instance:

"*The magnitude and structure of the error in total cloud fraction (TCF) hindcasts produced by CMIP5-level GCMs is then described.*"

This first appearance is followed by:

"*CMIP5 TCF hindcasts comprising 25-year (1980-2004) annual projection means were compared to the A-train observations [Jiang, JH et al., 2012].*"

The remaining "hindcast" usages are semantically identical, and are completely consistent with those in IPCC AR5 Chapter 9 or as recently used by, e.g., Meehl and Teng. [IPCC, 2013; Meehl, GA and Teng, H, 2012]

Meehl and Teng define "hindcast" as, "*an initialized "prediction" for a time period in the past,*" which is exactly the manuscript usage throughout. One is hard-pressed to find any justification for the reviewer's criticism.

4. *Also, a naive and simple linear framework for emulating global climate models (GCMs) is presented, that looks more like a fit to data. It is argued that it is skillfull, but given the initial fit and forcing data as input, this is hardly a tough test.*

4.1 The reviewer is mistaken: there is no "*initial fit*." Manuscript Section 2 derives the linear passive warming model (PWM) without a fit of, or reference to, any air temperature projection.

Manuscript Figure 1a and Figure 1b, and Auxiliary Material Figure S1, all show credible emulations of climate model air temperature projections without any fitting whatever. The reviewer has completely misstated the manuscript result.

As exemplified in Figure S2, the later fits merely yield a climate sensitivity coefficient for a specific individual GCM simulation. With the single derived coefficient, the PWM accurately emulates every single CMIP3 climate model projection.

These results rigorously demonstrate that the air temperature projections of advanced GCMs are merely linear extrapolations of greenhouse gas forcing. Linear error propagation necessarily follows.

4.2 With the comment concluding, "*this is hardly a tough test,*" the reviewer has in fact acknowledged that GCMS are mere linear extrapolation machines. Once this admission is made, the validity of the linear propagation of error is an unavoidable conclusion.

This reviewer, therefore, has inadvertently acknowledged the validity of the manuscript analysis.

5. *Furthermore, this emulation framework muddles external forcing with feedbacks, ...*

5. The reviewer again did not specify any muddled manuscript usage. Manuscript page 9, line 18 distinguished water vapor feedback from CO_2 forcing as, "*the sum of intrinsic CO_2 forcing with the calculated positive feedback following from the condition of atmospheric constant relative humidity*." with reference to [Held, IM and Soden, BJ, 2000].

That is, the follow-on effect of water vapor from CO_2 forcing is described as a feedback. The reviewer apparently missed this definition.

Throughout the manuscript, cloud forcing is used exactly in the manner of [Lauer, A and Hamilton, K, 2013].

These usages leave the reviewer criticism without any foundation.

6. *...and the treatment of errors assumes that each increment is independent of each other.*

6. Only two sorts of errors are discussed in the manuscript. The first is multi-model average CMIP5 cloud forcing error (±4 Wm-2), which is the 25-year mean annual error. There is no incremental character to a multi-year multi-model average. Mean error necessarily propagates through the serial calculations of any futures prediction.

The other sort of error is shown in Figure 3, namely the 25-year average cloud fraction (CF) error made by CMIP5 models. Clearly, no assumption of incremental independence is made here, first because average CF error itself is nowhere incremented, and second because the manuscript specifically shows CF errors are systematic.

The reviewer again has made a criticism absent any specification. Reference to the actual usage of error in the manuscript shows that the reviewer's criticism is, in any case, groundless.

7. *The author looks at zonal means of cloud biases, and does not realise that the latitudinal structure is due to well-known phenomena and circulation patterns - we should not expect a white (or red) noise type stochastic structure of the residuals, because the cloude climate varies with latitude. There is also varying degrees of freedom, as the space 'converges' in the polar regions due to the geometry of a sphere. Also, the effect of clouds vary with latitude both due to the solar inclination and cooler poles.*

7. The reviewer dismisses the cloud fraction error analysis by referencing cloud circulation patterns, by noting that cloud climate varies with latitude and has certain degrees of freedom, and by observing that spherical space converges in the polar regions (by reason of equatorial rotation, not spherical symmetry as such).

The reviewer does not notice, however, that these attributes of clouds are exactly what should be reproduced in a GCM simulation. Were the simulations true to observations, to physical reality, the difference residuals would be free of all these deterministic phenomena, leaving only noise.

That is, in a perfect simulation the error residual would contain no trace of circulation patterns, cloud climate, or polar convergence. However, the difference residuals do include a deterministic signature.

The reviewer acknowledges this, but does not understand the significance. As perfect simulations would leave no unaccounted residual cloud climatology, there is no, '*Of course the residuals should include cloud climatology.*' They should not do.

In the event there is residual climatology, the simulations are not accurate. This is the point of the manuscript analysis, and the point the reviewer has completely missed.

The cloud simulations of Figure 3 are 25-year GCM averages. All random error should be 5× attenuated. The fact that latitudinal autocorrelation remains in the CF error residuals shows evidence of deterministic error. That is, the simulations did not successfully reproduce the physically real circulation patterns, cloud climate, or polar convergence. The autocorrelation of residuals demonstrates systematic theory bias.

That is a major point of the analysis; a point that has entirely escaped the reviewer and a point that empties the reviewer criticism of any relevant content.

8. *The best way to test the errors of the GCMs is to run numerical experiments to sample the predicted effects of different parameters, which indeed has been done and presentated in the IPCC reports - eg natural versus total forcings.*

8. One might ask the reviewer how the simulation errors would be quantified without reference to physically real observations. This critically central standard of science appears nowhere in the reviewer's description of model error analysis.

   To illustrate: the reviewer suggests using parameter sensitivity test-simulations carried out using GCMs, to be compared with natural forcing simulations also carried out using GCMs.

   The reviewer's view is that physical error = GCM simulation differences.

   But GCM errors will appear in both sets of simulations, though different in degree and distribution, and in each case of unknown extent and magnitude. How, then, does the reviewer expect that taking a difference between cryptically erroneous simulations will expose GCM physical error?

   The answer is, it will not.

   It exposes only the difference between simulations, which yields no indication of the structure of physical error.

   The reviewer's view merely exemplifies the unappreciated distinction between precision and accuracy that apparently universally plagues the thinking of climate modelers.

9. *Any analytical or simplified emulation must reproduce these kind of the results of such experiments - also the error bars.*

9. Differences between simulations and their means produce self-referential model variance. They do not produce physical error bars. The reviewer's approach to physical error is wrong. Only differences between observations and simulations produce physical error bars. The latter are the sorts of error bars applied in the manuscript, and are the sorts of error bars apparently unknown to the reviewer.

   The experiments described in the manuscript prove the PWM emulation of GCM air temperature projections. The emulations demonstrate that climate models merely linearly extrapolate greenhouse gas forcing. This simple output allows linear extrapolation of error. This is standard error analysis in the physical sciences, as provided in Chapter 3 of [Bevington, PR and Robinson, DK, 2003]. It is a standard never met in climate modeling.

10. *The most obvious indication that the error framework and the emulation framework presented in this manuscript is wrong is that the different GCMs with well-known*

different cloudiness biases (IPCC) produce quite similar results, albeit a spread in the climate sensitivities.

10. The reviewer is apparently unfamiliar with the work of [Kiehl, JT, 2007], who showed that the similarity among climate model projections relies entirely on tuned anti-correlated parameters. See also [Huybers, P, 2010; Knutti, R, 2008]. Anti-correlated parameters, adjusted to reproduce the identical target, enforces the similarity of projections. This circularity falsifies any recourse to proof by similarity.

    The reviewer's criticism 10 also implies a lack of understanding of the distinction between growth of error and growth of uncertainty. Independent growth of error can produce a dissimilarity of projections among climate models. Growth of uncertainty due to physical error negatively impacts the predictive reliability of climate models, even given identical simulations across all models.

    This distinction of model error and predictive uncertainty is also fundamental to the physical sciences, and again is one that seems ubiquitously missing from the professional perceptions of climate modelers.

The author now returns to the reviewer's opening statement that the manuscript describes, "*an ill-designed analysis based on invalid assumptions and misconceived ideas about climate and climate models.*"

Following assessment, the reviewer evidences no apparent understanding of the distinction between accuracy and precision absolutely critical to science, has included object-free criticisms, has wrongly described the PWM as a fit, does not understand the logic that linear variable extrapolation is vulnerable to linear error propagation, and evidently understands neither the meaning of physical error nor the significance and impact of model average physical error.

The reviewer is apparently also unaware of tuned anti-correlated parameters within climate models, their tendentious impact on climate simulations, and the vitiation thereby of any comfort to be taken from similarity among multi-model climate projections.

The author ends by noting that reviewer's conclusion, stated at the outset, is completely unsupported by anything the reviewer has presented. Instead, the review illuminated a distinct lack of understanding of physical error and its analysis; all too common among climate modelers.

References:

Bevington PR, and Robinson, DK. 2003. *Data Reduction and Error Analysis for the Physical Sciences*, 3rd edn. McGraw-Hill: Boston.

Held IM, and Soden, BJ. 2000. Water Vapor Feedback and Global Warming. *Ann. Rev. Energy Environ.* **25**(1): 441-475, doi:10.1146/annurev.energy.25.1.441.

Huybers P. 2010. Compensation between Model Feedbacks and Curtailment of Climate Sensitivity. *J. Climate* **23**(11): 3009-3018, doi:10.1175/2010jcli3380.1.

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

ISO/IEC. 2008. Guide 99-12:2007 International Vocabulary of Metrology - Basic and general concepts and associated terms (VIM)Rep., International Organization for Standardization, Geneva.

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

Knutti R. 2008. Why are climate models reproducing the observed global surface warming so well? *Geophys. Res. Lett.* **35**(18): L18704, 18701-18705, doi:10.1029/2008gl034932.

Lauer A, and Hamilton, K. 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.

Meehl GA, and Teng, H. 2012. Case studies for initialized decadal hindcasts and predictions for the Pacific region. *Geophys. Res. Lett.* **39**(22): L22705, doi:10.1029/2012GL053423.