Patrick Frank 28 September 2014
Propagation of Error and the Reliability of Global Air Temperature Projections
Manuscript # JOC-14-0316

Summary

- The review is constructive and helpful.
- Water vapor enhanced CO_2 forcing has been defined at its first appearance.
- A new paragraph has been added to the renamed Section 3, Summary and Discussion, to indicate the $\pm 4 Wm^{-2}$ TCF error is a lower limit of model energy resolution.
- Model tuning and the similarity of climate projections is discussed (item 4).
- It is hoped that the remaining responses alleviate the reviewer's concerns.
- The manuscript has been modified throughout as the reviewer recommended.

Detailed response to Reviewer 2

Reviewer comments are presented in full, indented, numbered, and in italics. Author responses follow.

> *1. I think that it is very interesting paper. If the claim is proven to be true, this paper will largely resound in the science community which depends on the reliability of long-term climate prediction. The claim that errors due to cloud bias are propagating and thus should be quantified based on accuracy (not precision) sounds perfectly like what is required in normal science. Thus I think that this paper needs to be published in the end.*

1. The reviewer is thanked for this constructive overview and positive recommendation. It is fully agreed that an assessment of predictive capacity is completely standard in science. If the author may add, in recommending this standard, reviewer 2 is unique among the reviewers of this work.

> *2.1. There are some important questions to reassure that the claim of this paper is correct. The forcing error due to cloud bias may be damped by Stefan-Boltzmann feedback embedded in current climate models.*

2.1. The reviewer may well be correct about error-damping.

Please note however, that the manuscript is about predictive uncertainty rather than the evolution of model error. These are not identical.

Error is the difference between a predicted magnitude and an observed magnitude. Predictive uncertainty defines the *a priori* reliability of the prediction. Cancelation of error does not correct the underlying physics. Predictive uncertainty, *inter alia*, represents model bias itself. It is calculated before the observation is made and before

any error is known.

2.2. *The perturbed temperatures at time i due to cloud bias will be partly restored at time i+1 by the release or trap of energy by Stefan-Boltzmann feedback process.*

2.2. Please note that the ±4 Wm^{-2} total cloud fraction (TCF) error is a multi-year, multi-model average. Twenty-six CMIP5 models were assessed across 20 years of hindcast, making 520 model years of output. [Lauer, A and Hamilton, K, 2013]

Across 520 model-years, the $i \rightarrow i+1$ walk of error should have averaged out to a reasonable estimate of mean annual error. This mean error can then be propagated through a projection to yield an estimate of predictive reliability.

For a futures projection, where observations cannot exist, the shifts of error are not knowable. In this case, only a representative estimate of reliability is possible, which is attained by propagating known average error.

The contrast between unknowable error but knowable confidence intervals is discussed in manuscript Section 2.4.3.

3. In addition, cloud fraction bias is not all, leading to error of ±4 W/m^2 of cloud forcing. Models have many different substances such as sea ice/snow, vegetation, cloud properties, and precipitation, etc. all of which also act to add error, or compensate error.

3. The reviewer raises a good point here. Therefore a new second paragraph has been inserted in Section 3, Summary and Discussion, to make the point that the ±4 Wm^{-2} TCF error is only a lower limit of model energy resolution. Other sources of error do indeed exist. The confidence intervals thus constitute only a representative lower limit of uncertainty, and do not provide a full accounting of model error.

It is again important to understand that predictive uncertainty is not the same as model error.

Hindcasts allow climate model error to be evaluated against observations. However, there are no observations against which a climate futures prediction can be evaluated. Predictive uncertainty can be estimated, however, by propagation of known model errors through a futures projection.

4. Then even so, why are model-projected temperatures not too variant in the year of 2100 in Figure 5.

4. The reviewer's question is again very pertinent, and goes to the heart of the matter. Models give similar projections because they are tuned to the same climate

observables, especially the top-of-the-atmosphere (TOA) radiation balance. When models are tuned, they produce similar projections even when their parameterization schemes are different.

The reason is that tuning models to observables produces parameter sets with anti-correlated (off-setting) errors, so that the modeled projection always matches observations. For example, climate models all successfully hindcast the same 20th century air temperature record, even though their climate sensitivity varies by a factor of 2 to 3.

This specific problem has been discussed [Kiehl, JT, 2007], as has the problem of tuning more generally, e.g., [Bender, FA-M, 2008; Covey, C, 2000; Eisenman, I *et al.*, 2007; Kiehl, JT, 2007; Räisänen, J, 2007; Schwartz, SE *et al.*, 2007].

Model tuning is discussed in original manuscript Section 3, p. 50, line 10ff. Parameters with off-setting errors reduce the divergence from observations -- the visually apparent error -- but they do not reduce the uncertainty of the projection.

The reason confidence intervals are large even though visual error is small, is that the existent parameter uncertainty means that any projection is not a unique solution to the problem of the climate energy state. This uncertainty is propagated into a projection. Off-setting errors do not guarantee that the underlying climate state is correctly represented or that its evolution is projected correctly. Off-set errors do not produce correct physics. The large confidence intervals reflect this condition.

When models are tuned to have off-setting errors, the similarity of their projections is a kind of false precision. Projection similarity does not signify accuracy, or that models with similar projections have equally similar simulated climate states. This distinction is one of the major points of the manuscript, (p. 50, line 10ff in the prior manuscript, and similarly in the revised manuscript).

5. I hope that the author can properly reflect my concerns in the manuscript, so readers can be more confident with the unique claim of this study.

5. The reviewer is thanked for the constructive comments and questions. It is hoped that the responses alleviate the reviewer's concerns.

6. *Minor comments:*
6.1. P2, L13: Water-vapor-enhanced CO2 forcing is not familiar to the general readers. Please explain it.

6.1. Water-vapor-enhanced forcing is defined on manuscript p. 9, line 17. However, the manuscript has been revised so that the definition appears at the point of first appearance in the Introduction, in the 4th paragraph after eqn. 2.

6.2. P2, L15: Unclear terms: error averages? Its high correlation among GCMs?

6.2. The lines have been re-written as, 'CMIP5 GCMs produce an average 12.1% error in total cloud fraction (TCF). TCF error is strongly inter-correlated among GCMs, implying theory-bias, and produces an uncertainty in annual atmospheric thermal energy flux of ±4 Wm^{-2}.' It is hoped the meaning is now clear.

6.3 P3 to the end: Please follow the correct citation format for IJC throughout the manuscript. "Celsius. [IPCC, 2013; IPCC_2007 et al., 2007b]" should be corrected to "Celsius (IPCC, 2013; IPCC, 2007).".

6.3. The references have been changed to the recommended format.

6.4. P5, L47: Remove "(sure knowledge)".

6.4. Removed.

6.5. P7, L5: "error thought model..."?

6.5. Thank-you. Corrected to '... error through model...'

6.6. P7, L14: "water-vapor-enhanced (wve) greenhouse effect" was not explained, but explained later on P9, L17. Explain earlier when this term appears firstly.

6.6 The definition of water vapor enhanced CO2 forcing has been moved to its first appearance in the Introduction, 4th paragraph after eqn. 2.

6.7. P7, L32: "GHG" was not explained.

6.6. GHG is now defined where it first appears in the Introduction, 5th paragraph after eqn. 2.

6.7. P9, L36: Remove "or absolute" since it is confused with "relative humidity".

6.7. The phrasing has been re-written as, *'*... conditions of fixed relative humidity or absolute relative humidity...', i.e., the terminology used in [Manabe, S and Wetherald, RT, 1967].

6.8. P10 to the end: 15 μ should be replaced by 15 μm.

6.8. Replaced as recommended.

6.9. P10 to the end: The first sentence should be placed in the next line of the subtitle.

6.9. Line returns have been added after every subtitle.

6.10. P11, L56: Why "θ=45"?

6.10. Figure 1 represents an average condition. Relative to the horizontal tangent, 45° is an average emission angle of a surface uniformly radiating through π radians. The method was used in [Ramanathan, V and Coakley, JA, 1978].

6.11. P12, L57: "under ( ), clear sky, and; ( ), cloud cover" includes strange symbols.

6.11. Apologies. Somehow the original symbols were not translated properly by the pdf-builder. They have now been replaced with more conventional symbols.

6.12. P17, L56: "Model E"?

6.12. Model E is the climate model of NASA GISS. "GISS" has been inserted to make this clear.

6.13. P19, L13: It was not clearly explained how ΔFi is calculated.

6.13. An explanatory sentence has been added. Briefly, ΔF_i is the step-change in greenhouse gas forcing calculated using the equations of [Myhre, G *et al.*, 1998]. When the IPCC SRES projections are emulated, ΔF_i is the incremental change in forcing supplied by the IPCC.

6.14. P19, L20: "ΔΔTa"?

6.14. Thank-you. The notation has been rewritten and simplified.

6.15. P20, L45: "Figure 3:10 in [AchutaRao, K et al., 2004]"; please correct the format.

6.15. Changed to "Figure 3.10 in ..." I hope this is what the reviewer had in mind.

6.16. P25, L34: "The mean of the target MODIS and ISCCP2 A-train satellite observations"; how did you calculate the mean? For example, the data period and the method.

6.16. The mean is the sum divided by 2. Methodological details may be found in Auxiliary Material Section 6, "*MODIS and ISCCP2 Observed, and GCM Hindcast Total Average Global Cloud Fractions*." The text has been amended to include further information.

6.17. P25, L36: If possible, it is advisable not to use "Auxiliary Information", but to include all information within this manuscript.

6.17. Agreed. Methodological information has been added. However, the manuscript is already quite long. The author worries that inclusion of all the methodological details would very likely make it unmanageable. Reviewer indulgence is requested.

6.18. P25, L55: I cannot understand what is "lag-1 latitudinal autocorrelation".

6.18. Apologies. Lag-1 means every point *i+1* is plotted against point *i*. Lag-1 plots test whether data are auto-correlated, meaning that the magnitude of point *i+1* is dependent upon the magnitude of point *i*. Uncorrelated data show a random scatter. Auto-correlated data produce organized associations of points.

In a fit, auto-correlation of the unfit residual typically means that some deterministic information remains in the residual. The fit does not exhaustively explain the data.

Lag-1 latitudinal auto-correlation of the TCF residual means that there remains deterministic cloud structure that is not represented in the GCM simulation.

Figure 4 Legend has been amended to convey this.

6.19. P30, L30: "up-radiant"? Do you mean "upward"?

6.19. Yes. The term has been changed to 'upwardly radiant.'

6.20. P30, L49: I wonder why the authors did not consider SCF. SCF and LCF are compensating each other.

6.20. [Lauer, A and Hamilton, K, 2013] did report the SCF error of CMIP5 models. However, LCF is most pertinent because it directly represents the tropospheric thermal energy flux bath. Please see original manuscript p. 30, line 37-49.

Greenhouse gas forcing is part of the tropospheric thermal energy flux. An error in long-wave cloud forcing means the tropospheric thermal energy flux is simulated incorrectly. Therefore, the simulated air temperature is not represented by the correct physics. Uncertainty in projected air temperatures follows directly from this.

Although LCF and SCF are physically complementary, the manuscript analysis concerns simulation error and the projection uncertainty that is its consequence, rather than the actual physical behavior of SCF and LCF.

While the anti-correlation between SCF and LCF may remove visually apparent error, this does not decrease the uncertainty of a simulation. Indeed, inclusion of SCF simulation error will just increase the CIs because the scale of acknowledged physical error has increased.

6.21. P30, L56: "is not separate" should be "is not separated".

6.21. "Separate" has been replaced by 'separable.'

The idea is that once GHG forcing enters the tropospheric thermal flux bath, its identity is lost. The climate responds coherently to the total flux within the troposphere, not piecewise to individual flux contributors.

Annual LCF flux uncertainty is large compared to the annual increase in energy flux from GHGs. This means the effect of the GHGs on the troposphere cannot be resolved.

6.22. P33, L51: "root-sum-square"; I wonder why not "root-mean-square" that is more popularly used.

6.22. Root-mean-square error (rmse), $\pm\sigma = \sqrt{\sum_{i=1}^{n} \varepsilon_i^2 \Big/ N-1}$, is the mean of "$n$" individual errors, each of ε_i magnitude. This might represent the uncertainty in an ensemble mean, where each member of the ensemble is independent.

Root-sum-square error (rsse), $\pm\sigma = \sqrt{\sum_{i=1}^{n} \varepsilon_i^2}$, is the error that accumulates into a final magnitude after a sequence of "n" calculations, each of which contributes an error, ε_i. RSS error is propagated error. RSSE might represent the uncertainty gathered into a final state after a series of "n" calculated intermediate states, where each intermediate state projects its internal error forward into the next state.

6.23. P33: Please check equation 8; e_i(T) seems constant.

6.23. The reviewer is correct. The $e_i(T)$ term is constant in each emulation because the ±4 Wm^{-2} is a constant average error. From eqn. 6, $e_i(T)$ depends on (±4 Wm^{-2})/F_0, and so varies with F_0.

The meaning of the $e_i(T)$ is discussed more thoroughly on p. 38, line 27ff. A reference to this discussion is now entered into the text after the appearance of eqn. 8.

The reviewer is further thanked for bringing this point to attention. A small mistake in eqn. 7 was noticed. This has been corrected to show the growth of the TCF error statistic. Two explanatory sentences have been added to the following text.

6.24. P34: "descends directly from eqn. 2, eqns. 6&7, and Sections 2.2-2.4 above" is not correct format for IJC.

6.24. The text has been modified. I hope it is now correct.

6.25. P34, L41&L54: "4AR" should be "AR4".

6.25. This has been corrected throughout.

6.26. P44, L35: "Summary Conclusions" should be "Summary and Discussions".

6.26. Changed as recommended. Section 2 has also been changed to "Results."

6.27. P45, first paragraph: nobody knows previous version of this paper. I don't think this paragraph is necessary.

6.27. The reviewer's point is understood. Mention of previous versions has been removed.

It is expected that the audience will include climate modelers to whom the same objections will occur, because these same objection have already been offered by several prior reviewers who clearly were climate modelers. Indeed, Reviewer 1 and Reviewer 3 of this version offered similar or identical objections.

It is hoped to forestall these objections should they occur to future readers. The paragraph has been shortened to reflect the reviewer concern, and the author requests the reviewer's indulgence.

6.28. P46, L11: "evaluate" should be "to evaluate".

6.28. Thank-you. Fixed.

6.29. P47, L6: What is "Equation (10)"?

6.29. Regrets for the confusing presentation. The "Equation (10)" referred to the Guide to the Expression of Uncertainty in Measurement, [JCGM, 100:2008]. The quote has been interpolated to remove this ambiguity.

6.30. P49, L41: "poor resolution" sounds like spatial or time resolution. If it is not that meaning, please rephrase it.

6.30. The term is now qualified as 'energy resolution.'

6.31. P52, L44-48: 'degree' symbol is missing.

6.31. Thank-you; added in throughout.

6.32. P53, L53: "(CMIP5) an AGW signal" seems like typo.

6.32. Removed. The "(CMIP5)" was meant to represent the current level of theory, however it is true that the sentence is clearer without it.

Finally, the reviewer is thanked for the very helpful critical commentary. It is hoped that the changes and explanations resolve the reviewer's concerns.

References:

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

Covey, C. (2000), Beware the Elegance of the Number Zero, Climatic Change, 44(4), 409-411, doi:10.1023/a:1005572518719.

Eisenman, I., N. Untersteiner, and J. S. Wettlaufer (2007), On the reliability of simulated Arctic sea ice in global climate models, Geophys. Res. Lett., 34(10), L10501, doi:10.1029/2007gl029914.

JCGM (100:2008), Evaluation of measurement data — Guide to the expression of uncertainty in measurementRep. Document produced by Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG 1). Bureau International des Poids et Mesures, Sevres, France.

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.

Myhre, G., E. J. Highwood, K. P. Shine, and F. Stordal (1998), New estimates of radiative forcing due to well mixed greenhouse gases, Geophys. Res. Lett., 25(14), 2715-2718.

Räisänen, J. (2007), How reliable are climate models?, Tellus A, 59(1), 2-29.

Ramanathan, V., and J. A. Coakley (1978), Climate modeling through radiative-convective models, Rev. Geophys., 16(4), 465-489, doi:10.1029/RG016i004p00465.

Schwartz, S. E., R. J. Charlson, and H. Rodhe (2007), Quantifying climate change -- too rosy a picture?, Nature Reports Climate Change, 2(7), 23-24.