From: **Journal of Forecasting** onbehalfof@manuscriptcentral.com
Subject: Journal of Forecasting - Decision on Manuscript ID FOR-17-0244
Date: April 6, 2018 at 8:11 AM
To: pfrank830@earthlink.net
Cc: p.young@lancaster.ac.uk

JF

06-Apr-2018

Dear Dr Frank,

I write regarding Manuscript ID FOR-17-0244 entitled "Propagation of Error and the Reliability of Global Air Temperature Projections" which you submitted to Journal of Forecasting.

In view of the comments of the referee(s) found at the bottom of this letter, your manuscript has been declined for publication in Journal of Forecasting. On the basis of his review, I decided that, since I am myself interested in climate forecasting, I should look at the paper myself. I am afraid thatI have to agree with him that your paper is more suitable for a climate or meteorological journal and would need to be shortened and modified considerably for the JOF audience. I hasten to add that this is a question of suitability and not necessarily any reflection on the technical merit of the paper.

PLEASE NOTE: The reviewer may have returned their comments to me as a separate file. We do not attach these files to the decision letter as they can interfere with successful delivery of emails. You can view the reviewers attachment by visiting your Author Centre in Manuscript Central. Click on "Manuscripts with Decisions" queue. In the list appearing at the bottom of the screen click on "View Decision Letter".
In the new window the files can be seen in the section "Files Attached"
at the bottom.

If you feel that your paper could benefit from English language polishing, you may wish to consider having your paper professionally edited for English language by a service such as Wiley's at http://wileyeditingservices.com. Please note that while this service will greatly improve the readability of your paper, it does not guarantee acceptance of your paper by the journal.

Thank you for considering Journal of Forecasting for the publication of your article. I hope the outcome of this specific submission will not discourage you from submitting future manuscripts.

Yours sincerely

Professor Peter Young
Journal of Forecasting
p.young@lancaster.ac.uk

Referee(s)' Comments to Author:

Reviewing: 1

Comments to the Author
General comments

The question asked by the author, which we might rephrase in terms more familiar to the audience is: what prediction intervals should be applied to century-scale forecasts of global surface temperature from general circulation models? This is a question of interest to the journal audience. However, the technical detail will be lost on most if not all of the readership. The style of the paper is more suitable for a climate or meteorological journal and needs to be shortened and redirected to an audience with limited knowledge of the physics and chemistry of the earth's climate.

GCMs take the approximately 340 Wm-2sec-1 of solar energy directed on average to the earth, the natural and anthopogenic emissions, and with constant forcings (usually allowing for the variation in the solar cycle) can simulate centuries of weather-like outputs. In control runs they must therefore, on average, maintain incoming and outgoing energy in balance. Without knowing any details, when modelers engage in "tuning" their models, proportion of cloud cover is likely to need to be parameterized. If the proportion of cloud cover is biased and all other fluxes are correctly (i.e., unbiasedly) estimated, then a long-term simulation would eventually go off track as, for example, surface temperature would gradually diverge from a constant climate. For this not to happen, then something else in their system must be biased in the opposite direction. Therefore, showing that one variable out of the hundreds generated in a GCM simulation has a sizeable error does not in itself provide a measure of forecast error.

Since surface temperature is a major focus of the CMIP5 project, and has the largest number of output files, a strong case would be to compare decadal hindcasts with actual surface temperatures. In fact, if this exercise is done with enough hindcasts then the actual forecast errors at different horizons can be used to estimate prediction intervals of any magnitude. For example, the 68% prediction interval (one standard deviation) would be drawn where 32 out of 100 errors fall outside the bounds. The shape of the intervals can be controlled to have the same shape as shown in Figures 8(b) or 9(b). If this is done the intervals will be much narrower than shown in the above figures. These are conditional forecasts since the historic forcings are known. Where they have to be forecast then that error must be added to the conditional forecast error, which is doubtless the reason for showing projections for different RCPs.

Section 2.1 is an example of excessive technical detail. Its purpose is to divide the total global warming into a component that is

Section 2.1 is an example of excessive technical detail. Its purpose is to divide the total global warming into a component that is water vapor alone and a component that is water-vapor-enhanced CO_2 warming. The three observations from Manabe and Wetherald (1967) are fitted to the function
T(K) = a ln[CO_2]ppmv+b as shown in Figure 1b. Assuming cloud cover stays at 66.7% then the contribution of water vapor alone is the sum of global surface temperature with 1 ppm CO_2 (taken from the fitted curve) weighted by the proportion of cloud cover minus TOA temperature all divided by 33. This gives rise to a water vapor contribution of 58% and a CO_2 enhanced component of 42%. These are approximately the widely quoted 60:40 forcing contributions.

Section 2.2 uses the 42% calculation in an equation that links temperature anomaly to the ratio of cumulative forcing to initial forcing. That is, the parameters are not estimated by fitting to a data series, they are derived from first principles. The section title is a misnomer: these are not projections of climate models, they are emulations. In Figure 2 it is noticeable that the residuals of ensemble mean minus fitted value are not random so that a different parameterization of the equation would give smaller and more nearly random residuals. Then in Figures 3 and 4 the parameters of the equation are changed for each model emulation though how the curves are fitted to the model output is not stated.

Section 2.3 compares hindcasts from GCMs of total cloud fraction with actual observations over a 25-year period. The residuals display high first-order autocorrelation, indicating bias. Second, cross-correlations of errors between pairs of GCMs show many values above 0.5 while for randomly constructed models the expected correlation is zero. This quantifies the well-known observation that GCMs have many common features so that simulations from different GCMs do not represent independent outputs.

In section 2.4 the root mean square error of aggregate GCM hindcasts of global total cloud fraction of 12% or .12 is converted to equivalent radiative forcing of 4 Wm-2year-1 which is compared with all greenhouse gas forcing (anomaly since 1750) of approximately 2.6 Wm-2year-1. This flux error is then applied to conditional surface temperature forecasts from the RCP forcings to show that the prediction intervals are dramatically wider that the variability implied by model ensembles.

Specific comments

p2 lines 20-22 I do not see what the big deal is about annual average global surface temperature. Of course it must be calculated from a large number of observations both temporal and spatial and so is a statistic.

p2 line 50 "meteorology"

p7 line 31 "33 K greenhouse temperature" The audience will not know what this means. Need to explain that the top of atmosphere temperature is taken to be -18 deg C, which would be the earth's surface temperature in the absence of greenhouse gases (mainly water vapor). Since the earth's average surface temperature is around 15 deg C, the difference between the two 33 deg C or deg Kelvin is the greenhouse temperature.

p20 line 19 What is the meaning of "intensive to all climate models"?

p29 line 46 Say "In Lauer and Hamilton (2013) . . ." Also p30 line 42.

p30 lines 10-31. I'm having some trouble following this. You have a three-dimensional data set: years x models x gridpoints and annual values are actually averages of a large number of observations or simulations. Your interest does not seem to in bias, though from figure S10 it is clear that some GCMs display bias since their curves never or rarely cross the curve of observed values. If you are looking for a measure of forecast accuracy then a mean square error criterion will cover both variance and bias. If you are planning to use an ensemble or composite forecast from all available GCMs then individual gridpoints represent errors and aggregating over all gridpoints gives an overall measure of forecast accuracy for this particular variable.

p31 line 52. It is 50% larger or 150% of anthropogenic GHG forcing

p31 line 55. That would be 2.6 Wm-2year-1