Reviewer #1: The submitted paper attempts to document the impact of structural errors in cloud radiative forcing within the current generation of of GCMs on the long term climate projections made by those models. The analysis is conducted by proposing a linear model relating global mean forcing of the climate system to temperature response. The author proposes that the systematic error in radiative balance the CMIP5 generation GCMs introduces a potential error orders of magnitude greater than the range of temperature projections seen in the 5th assessment of the IPCC, and thus concludes that the GCMs are not a reliable source of information to inform future climate projections.

The author motivates his study by considering the lack of end-to-end uncertainty analyses in GCM projections, and to some degree, he is correct that a comprehensive study documenting the propagated uncertainties of the assumptions made in GCM parameterizations and systematic choices would be desirable - and although various studies have made significant inroads into making such an assessment, the field is far from complete.

However, this author's proposed error propagation study is certainly not an accurate assessment of model error, and confuses a number of fundamental properties of the climate system and how it might behave in a zero-order model.

Firstly, the linear emulator proposed by the author represents the global mean temperature as a linear function of the forcing. This effectively ignores any thermal inertia in the system, and assumes that the system will instantly equilibrate to a new forcing level. Ignoring the thermal mass of the ocean is not an appropriate assumption, even for a simple emulator.

However, even allowing for this oversight, the critical flaw in the author's logic is the treatment of errors in global mean cloud forcing, and how this is then related to each year's incremental forcing change. The author proposes that there is a +/- 4Wm^-2 error in global mean longwave cloud forcing in the CMIP5 generation GCMs. He then proposes, as a result, that this error accumulates over time, i.e. that year 1 is subject to a 4Wm^-2 uncertainty, year 2 would be almost 8Wm^-2 etc.

In making this assumption, the author is confusing forcing and feedback. It is true that the models in AR5 exhibit systematic errors in cloud forcing. This is well documented in the very papers the author uses to obtain his cloud forcing data. No author has suggested that these errors are random, or should be treated as such. But the biases are just that, they are constant, and each model has already equilibrated to whatever mean state cloud bias that it exhibits. In the author's simple linear model without an ocean (eq. 6), the +/-4Wm^-2 should be appended to the F_0 term, not to the \Delta F_i term.

What the author should be considering in this study is the GCM uncertainty in cloud feedback. For sure, the mean state bias can inform the degree to which we trust each GCM. Various studies (Sherwood et al 2014, Fasullo and Trentberth 2013 amongst others) have documented promising methodologies for relating the mean state biases to feedbacks, but this remains an active field of research. But the author's approach, assuming that the bias will accumulate with each year of the simulation is simply incorrect.

On the basis of these fundamental errors in the logic of the study, I find the paper unacceptable for publication. In addition, I note the following additional minor points.

Page 3, 1st para: The likelihood of the future climate warming significantly is not conditional only on the use of GCMs. Considerations of the paleoclimate record, the observed warming during the satellite era alone or simply the radiative impact of increasing CO2 alone (as the author presents) makes it hard to make the argument that the Earth will not warm in a high CO2 future.

Page 3, 2nd para: A skillful representation of atmospheric processes is exactly what gives us confidence that GCMs are more meaningful than a simple linear extrapolation of global mean surface temperatures. Their ability to represent complex coupled processes: ENSO, the Madden Julian Oscillation, the climate of the deep past - is exactly what gives us confidence. These processes are emergent properties of the simulated system, conditional on the representation of dynamics, radiation, clouds, ocean currents, sea ice, biological feedbacks. Thus the ability of the model to represent these coupled processes as a sum of simpler parts gives us confidence that the system we have built is able to represent the emergent behavior of the climate.

Page 2, 1st para: Parameter sensitivity tests are not tests of precision. They are tests of propagated error in the purest sense. Our assumptions are the parameter values, and by varying these parameters within a range of plausibility and running climate simulations we can assess how these assumptions are impacting our projections of long term climate change.

Page 2, 1st para: Taylor diagrams, however are not a measure of accuracy because they are not predicting an out-of-sample metric. Taylor diagrams are used to tune climate models to the observed climate. Therefore, they are not measures of predictive skill, but the degree to which the model has been tuned to replicate the observed climate. Accuracy in future projections is not guaranteed by the models' ability to match the observed climate (although the latter is a necessary condition for us to consider the model a plausible candidate).

Page 6, 2nd para: "error bars" are regularly published in studies of propagated

errors - see Sexton et al (2013), Rowlands et al (2012), Collins (2012), Yamakazi (2013) amongst many others.

Page 6, 2nd para: systematic energy flux errors are not inputs to the system, they are resolved outputs (which might exhibit errors). This is the origin of the author's primary logical miscalculation.

Page 7, 2nd para: evaluating total cloud fraction is a complex process, and the different GCMs report it in different ways. It is an entire field of study to assess how best to compare observed cloud properties to their representation in GCMs, requiring satellite simulators to be built into the GCMs themselves to replicate the inverse process which is used to detect cloud properties from satellites. This field cannot be realistically summarized with general statements about "Global" cloud fraction without strictly defining how cloud fraction is to be defined.

Page 8, 1st para: what is the justification for ignoring all other feedbacks rather than the water vapor feedback? Longwave and (primarily) shortwave cloud feedback is our primary uncertainty in future climate change, but any comprehensive feedback model also needs to consider land ice, sea ice and land surface feedbacks, not to mention carbon feedbacks and ocean circulation feedbacks.

Page 8-9: this section is entirely irrelevant, given that 1ppm CO2 is entirely outside of any Earth-like state.

Yamazaki, Kuniko, et al. "Obtaining diverse behaviors in a climate model without the use of flux adjustments." Journal of Geophysical Research: Atmospheres 118.7 (2013): 2781-2793.

Sexton, David MH, et al. "Multivariate probabilistic projections using imperfect climate models part I: outline of methodology." Climate dynamics 38.11-12 (2012): 2513-2542.

Rowlands, Daniel J., et al. "Broad range of 2050 warming from an observationally constrained large climate model ensemble." Nature Geoscience5.4 (2012): 256-260.

Collins, Matthew, et al. "Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles." Climate Dynamics 36.9-10 (2011): 1737-1766.