Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-54-AC2, 2018

Geoscientific Model Development Discussions
Open Access

Interactive comment on "Baseline Evaluation of the Impact of Updates to the MIT Earth System Model on its Model Parameter Estimates" *by* Alex G. Libardoni et al.

Alex G. Libardoni et al.

alex.libardoni11@gmail.com

Received and published: 2 July 2018

1) I miss a description of the basic components and parameterizations of the model in the method section 3. I miss a section that describes model spin-up and the setup for the different model simulations, including external forcing factors. Further, it is not evident from the description why the model is called "Earth System Model". For example, are biogeochemical cycles included? Does dynamic vegetation affect albedo? Is it an ESM or rather an Earth System Model of Intermediate Complexity? I also miss a brief description of the metric used to compare model and data and how they are used to derive probability distribution. It is not sufficient to refer the reader to the literature

Printer-friendly version

Discussion paper

(Libardoni and Forest 2011).

Response: The MIT Earth System Model is an integrated model with sub-models for the atmosphere, ocean, land surface, atmospheric chemistry, ocean biogeochemistry, and the terrestrial ecosystem. When all of these sub-models are turned on, the model is set up as an Earth system model. However, under that set up, the model is too computationally expensive to be used for probabilistic studies of the model parameters like what is presented in the present study. Turning off all components of the model except the atmospheric, ocean, and land surface models simplifies the model to an EMIC that can be used for probabilistic estimates of the model parameters investigated in this work.

A more detailed presentation of the EMIC (climate component of MESM) has been added to Section 2. In that discussion, we describe the model components of the EMIC, the input forcings, and the model parameters. In the discussion of the model parameters, we describe how each of the three are adjusted and how the model is being modified to make the changes.

In Section 3, we have included a summary of the methods used to derive the probability distributions. We present the goodness-of-fit statistic used to evaluate the model. This statistic is the weighted sum-of-square residual between the model output and observed climate record for a given diagnostic. A reference to the likelihood function is provided and then we explain how the joint distribution is calculated from the goodness-of-fit statistic.

2) Section 3: The authors vary three parameters – ocean diffusivity, an aerosol forcing scaling, and the strength of the cloud feedback determining ECS and constrain the models with two parameters.

2a) There is little information in the method section what these parameters specifically influence. The aerosol forcing scaling is unclear. Does this mean that all aerosol forcings are lumped together and scaled with a constant time invariant factor? How are different uncertainties applying to different aerosol classes (e.g. sulfate versus soot) considered or not and what is the justification for this approach. Please discuss caveats related to your assumption of a scaling factor.

Response: In the description of the model parameters that was added to Section 2, we describe what each of the parameters influence. For completeness, we summarize them again here. ECS is modified by adjusting the strength of the net cloud feedback in the model. More specifically, a number of simulations where CO_2 concentrations have been doubled and the system brought to equilibrium have been run for different values of the cloud adjustment. These are used to provide a lookup table which gives the cloud adjustment needed for a specific ECS. Ocean diffusivity is defined by a latitude-dependent pattern based off of tritium mixing into the deep ocean. K_v represents the global mean value and specific diffusivity values are calculated by scaling the spatial pattern by the same value at all latitudes to achieve the desired global mean value.

The forcing due to all aerosols except sulfate are held constant during historical simulations and the sulfate aerosol is parameterized through adjustments to the surface albedo based on changes in the historical emissions of SO_2. The historical emissions have both spatial and temporal components, with the aerosol parameter setting the amplitude of the pattern in the 1980s. Adjusting the forcing in this manner is not without its drawbacks. As the only adjustable forcing component in the model, this forcing pattern also represents an estimate of all other forcings not included in the model. Thus, this is not a pure estimate of the aerosol forcing.

2b) Effective ocean diffusivity is a very loose term. Is this diapycnal, vertical or horizontal diffusivity or does the parameter refer to the diffusivity associated with Gent-McWilliams parameterization? The subscript v of K_v points to vertical diffusivity. I would hope that this parameter reflects diapycnal diffusivity as diapycnal diffusivity co-governs ocean overturning strength and thus surface-to-deep heat transport. In any case, I am puzzled about the range sampled. Diapycnal diffusivity in coarse resolution, dynamic ocean models is typically of order 0.1 10^{-4} m^2 s^{-1}. Here diffusivity is varied in steps of 1 10-4 m2 s-1 and a very wide range up to 64 10-4 m2 s-1 is used. The upper value is even much larger than applied in classical box-diffusion models (1-2 10-4 m2 s-1); in box-diffusion models the entire vertical transport (mixing, advection, convection) is parameterized by diffusion only. What is the justification for this large sampling range? As a minor point, please use SI units for diffusivity. Further, I though Gent-McWilliams parameterization is included in the MIT model. If yes, why is the Gent-McWilliams diffusivity not varied or is this parameter linked with the “effective diffusivity”?

Response: We have added text to the manuscript to address these concerns. We have clarified that a mixed-layer ocean model is used. In this model, horizontal heat transport is prescribed by the Q-flux calculation and the vertical mixing of heat into the deep ocean is prescribed by the spatial diffusivity pattern and scaled by Kv as discussed above. As Kv represents the mixing of heat into the deep ocean by all processes, it is greater than diapycnal diffusion values found in the sub-grid scale parameterizations of dynamic ocean models.

A wide range of Kv values was sampled to simulate many possible climate states, including those with very strong vertical ocean mixing. Similarly, wide ranges were also chosen for climate sensitivity and the aerosol forcing. For the most part, runs with extreme values of any parameter were rejected for being inconsistent with the model diagnostics. In the case of Kv, this supports the claim that such high values should not have been sampled to begin with. The penalty paid for this over sampling of the parameter ranges is a misallocation of computing resources.

2c) ECS is typically used to abbreviate Equilibrium Climate Sensitivity. Here, an effective climate sensitivity is introduced and termed ECS. What represents this effective climate sensitivity?

Response: We mistakenly expressed ECS as effective climate sensitivity, when it is, in fact, equilibrium climate sensitivity. The lookup table for ECS is derived from runs brought to equilibrium, so that any equilibrium climate sensitivity can be obtained through the proper adjustment of the cloud feedback. All references to effective climate sensitivity have been changed to equilibrium climate sensitivity.

3) Section 3: I question somewhat the application of only two observational metrics to constrain ECS, TCR, and sea level rise. Namely, pattern of surface air temperature change and "linear" ocean heat uptake are used as constraints by the authors. In my opinion, there is a lack of observational constraints to probe the timescales of deep ocean overturning (e.g. 14C). Thus it appears not surprising that the diffusivity parameter remains not well constrained. There is also a lack of metrics to probe the spatial pattern of heat uptake. This is particularly important as the thermal expansion coefficient varies by almost an order of magnitude in the ocean. Thus it matters, where the heat is taken up to estimate sea level rise. As another focus of the study is on TCR, it would also be nice to invoke additional metrics on thermocline ventilation as for example available by observation-derived fields of CFCs and bomb-produced 14C.

Response: Given the mixed-layer ocean model that is coupled to the atmosphere, we are somewhat limited to the diagnostics that can be used to evaluate the ocean system. As further explained above, the vertical mixing pattern is prescribed with latitudinal dependence, but also fixed throughout the run. The vertically-integrated horizontal heat transport is also prescribed based on offline Q-flux calculation. With these patterns fixed, incorporating ocean diagnostics with spatial dependence is not feasible at this time.

As an aside, developing additional model diagnostics to constrain estimates of the model parameters, TCR, and sea level rise is a task that should be undertaken and is of interest to the authors. Care should be taken to ensure that these metrics are independent of each other or that steps be taken to account for the correlation between metrics. However, developing such metrics is beyond the scope of this work.

4) Page 5 to page 7, results, The description of the difference in input forcing is useful,

Printer-friendly version

Discussion paper

but in my opinion misplaced. Solar and ozone forcings are model drivers (or forcings) and distinct from a particular model version. These forcings should be described in the method section where the simulations and the applied external forcings are to be described.

Response: While we recognize that the presentation of the model forcings may be better placed in the methods section, we believe that keeping them in the results sections is justifiable. The interpretation of the new forcings and their direct application to the model parameters are in themselves a finding in this study. Much of the reasoning for the shifts in the parameter estimates centers around these changes in the model forcings and are essential to the explanation of the results. In our opinion, keeping them together is appropriate.

5) P6, line 3ff; Q-flux adjustment: Does this mean that the authors apply temperature flux correction to their model? This should be explained in the method section.

Response: An explanation of the Q-flux adjustment has been added to the manuscript and discusses how it is related to horizontal heat transport in the ocean.

6) Section 4: I miss a figure comparing the modelled pattern of the median (or mean or best-guess version) with the observed pattern of surface air temperature change and similar for the global ocean heat uptake and its spatial pattern (and may be for upper air temperature) to illustrate how well the model is able to capture the observations.

Response: A figure comparing the model output to the observed surface pattern used in our diagnostic does not yield a clean comparison. As a result of weighting the model-to-observation residuals by the noise covariance matrix, the temperature patterns are rotated into a coordinate space defined by a set of orthogonal basis functions defined by the internal variability estimate. Thus, any attempt to compare the model output and observations in the unrotated space does not give a fair representation of an individual model run's fit to the observed record.

Printer-friendly version

Discussion paper

A fairer assessment of the model fit to the observations is obtained by comparing the global mean temperature time series. We have included a figure where the global mean surface temperature of each of the 1800 model runs is shown, along with the observed time series for each of the five datasets used in this study. We have also highlighted the model runs where the parameter settings most closely match the median values from the marginal distributions derived from each of the surface datasets. All anomalies are calculated based off of the 1906-1995 climatology used in the surface diagnostic.

Similar to the global mean surface temperature results, we also include a figure to show the spread in the ocean heat content linear trends calculated from our ensemble. We plot a histogram of the calculated trends from each individual run, while also showing the observed trend and highlighting the runs with parameter settings closest to the distribution medians. Given the fixed pattern used for ocean mixing in the ocean model, the spatial pattern of heat uptake does not vary between the model runs. Only the magnitude changes, making a comparison between the model and observations for individual runs redundant.

7) Page 12, line 7: How well does the polynomial fit represent the model results?

Response: In general, the polynomial fit represents the model results quite well, but is not without error. In our response to Reviewer #1, we discussed using first-, second-, and fourth-order fits, as well as some of the errors associated with the third-order fit.

8) Page 12, line 14: Why is the PDF for the TCR not directly estimated from the 372-member ensemble? Does the fitting add additional uncertainties to the procedure of estimating TCR?

Response: It is possible to directly estimate the PDF for TCR from the 372-member ensemble. Doing such would represent estimating TCR from a joint distribution where all values of ECS and Kv are equally likely to occur. In other terms, the ECS-Kv two-dimensional PDF would be uniform for all pairs within their respective domains. We have shown in this study that ECS and Kv are not uniformly distributed and that some pairs are more likely to occur than others. Drawing from this more realistic distribution yields a probability-weighted sampling of parameter pairs from which to estimate TCR.

Using the polynomial fit adds additional uncertainty to the procedure of estimating TCR by introducing interpolation error. As described in the response to Reviewer #1, the polynomial fit is not an exact match to the model results, and any error in the estimation propagates as an error in the TCR distribution. However, running the transient simulation for each ECS-Kv draw from the Latin Hypercube Sample is infeasible, so the fit is required to estimate TCR for the pairs where there is no corresponding run.

9) Discussion and conclusion: While the authors suggest that their approach should serve as a template for other groups, they fail to mention that similar, and sometime much more comprehensive approaches of parameter calibration, have been undertaken by other groups. They also fail to compare their estimate of TCR and ECS with published estimate and to put their findings in the context of the wider literature. See for example, Collins et al., IPCC, 2013 for the most recent assessment of TCR and ECS values by IPCC. Of course there are recent updates of these estimates and there are also many other studies that determine model parameters such as vertical ocean diffusivity. Examples that come immediately in my mind are Holden et al., Clim. Dyn., 2010, Richardson; Nat. Clim.Change, 2016, Schmittner et al., GBC, 2009, Steinacher et al., Science, 2013 or Steinacher and Joos, Biogeosciences 2016. It is the task of the authors to identify the recent literature to provide a relevant discussion.

Response: In both the abstract and the penultimate paragraph of the introduction, we state that the point of the study is to assess how the changes in the model can impact the distributions. The paper is not intended to discuss how the results compare with recent estimates of ECS or TCR distributions or specific methodologies for estimating probability distributions. We think the introduction's text reflects this and is included here.

"In this study, we provide a transparent method of testing and accounting for how the simulated behavior and probability distribution functions change in response to the recent model development. We derive a new joint probability distribution by closely following the methods of Libardoni and Forest (2011) to show the impact that the new version of the model has on the parameter estimates and find that the new version of the model leads to higher climate sensitivity estimates in addition to shifts in the distributions of the other model parameters. The effects on the parameter distributions due to changing observations and temperature metrics will be addressed in future papers to separate their impacts from those due to changes to the model framework alone."

The future work will provide the appropriate discussion of other studies as suggested by the reviewer while this work only documents the impact of changes in the model framework.

We are aware that other approaches exist and have avoided stating that our parameter estimation methodology is better. We do think this approach can serve as a template for testing how new versions of models can directly impact parameter estimates and that such tests should be documented in a similar fashion.

P1, Line 22: typo: sensitivity

Response: We have fixed this typo in the manuscript.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-54, 2018.