Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-161-AC1, 2016

Geoscientific Model Development Discussions

Interactive comment on "ASoP (v1.0): A set of methods for analyzing scales of precipitation in general circulation models" *by* Nicholas P. Klingaman et al.

Nicholas P. Klingaman et al.

n.p.klingaman@reading.ac.uk

Received and published: 23 November 2016

GMD-2016-161
Response to comments from Christian Jakob

We thank the reviewer for providing a positive and encouraging review of our manuscript. In our response below, the reviewer's comments appear in red text; our response to the reviewer's comments appears in black text.

General comments

This is an excellent paper that introduces a number of interesting and useful diagnos-

tics to study the behaviour of precipitation in weather and climate models at various space and time-scales all the way to single grid points and time steps. The paper is very well written and the analyses and arguments are sound. It adds several new ideas for model analysis and evaluation, all of which will add to the arsenal available to the community. I have a few minor comments, which I list below.

We thank the reviewer for these positive comments.

Specific comments

1) The new diagnostics introduce are very nice and worth looking at in the present form. However, as there are many of them, it would be nice if the authors could consider producing a few summary measures that could be presented more easily when comparing many models and/or evaluating them against observations. This would contribute to the growing interest in having "performance metrics", so that changes in models can be more easily assessed and their effects quantified. I am aware that there is no single metric that identifies good or bad models, but by having a collection of them—well beyond this study—I believe the community will be able to better communicate model improvement in the future. For one-dimensional histograms, there are simple statistical techniques one could use, such as the Kolmogorov-Smirnov two sample test, which allows an assessment of the likelihood that two samples are drawn from the same population. This would be especially useful where models are compared to observations. It would be nice to also have a summary measure of the two-dimensional histograms presented here, but that might be more difficult.

We agree with the reviewer and thank the reviewer for this suggestion. Reviewer #2 also proposed that we quantify the persistence and intermittency of precipitation in models and observations. We have combined the two reviewers' suggestions and provide a unified response below, a copy of which appears in our response to Reviewer #2.

We have created summary metrics of spatial and temporal coherence in precipitation, which allow the reader to more easily evaluate models, either against each other or against observations. These metrics are based on the persistence of upper- and lower-quartile precipitation in time (measured from one timestep to the next) or space (measured at neighboring gridpoints). Considered together, the metrics summarise aspects of our two-dimensional histograms, as well as our correlations of precipitation as functions of distance and time, but without relying on the choice of a threshold correlation value or spatial or temporal scale. The metrics are scaled to range from -1 to +1: positive values indicate that persistence is more common than intermittency; negative values indicate that intermittency is more common than persistence. Table 3 of the revised manuscript shows these metrics for all models, as well as for TRMM and CMORPH satellite-derived observations, and for all horizontal grids and temporal frequencies considered in our study (i.e., timestep and 3-hr, native-grid and $5.6^{\circ} \times 5.6^{\circ}$ averages). The metrics confirm many of the conclusions of our study concerning the effects of averaging in space and time on the spatial and temporal coherence of simulated precipitation features. These quantitative metrics also confirm our qualitative conclusions about the relative levels of spatial and temporal coherence in the models we analysed.

We have added a description of these metrics to Section 2.1.4. The metrics are shown in Table 3. We discuss the metrics throughout the Results section (Section 3) and refer to them in the Discussion (Section 4) and Conclusions (Section 5) sections.

These metrics have helped to demonstrate quantitatively the qualitative conclusions we had drawn from our analysis, so we are very grateful to the reviewers for suggesting them.

2) Page 12, Line 6-7: This is a very strange argument. It's likely you did look at the lagged time correlations for 3-hourly rainfall and found something. Why not just state what you found—no need to show a figure if you can say it in words. Saying that there

could be a problem, rather than that there is one, sounds like you have something to hide.

We agree with the reviewer. We have changed the paragraph in question to read: "For model data, lagged correlations of 3-hr precipitation (i.e., as in Fig. 6b but for 3-hr data) over a 36-hr window were dominated by the overly strong and regular diurnal cycle of precipitation in the models, which manifested itself in our diagnostics as a pronounced peak in the correlations at a 24-hr lag (not shown). TRMM and CMORPH displayed a much weaker and broader peak across lags of 18–30 hr, suggesting greater day-to-day variability in the timing of the diurnal maximum in tropical maximum in the satellite-observations than in the models."

We believed that because the lag-correlation diagrams for 3-hr data were dominated by the diurnal cycle in the models, showing the diagrams would not address our original objective of analysing the temporal coherence in the 3-hr precipitation data on sub-daily scales (i.e., lag-1 or lag-2 correlations in the 3-hr data). Instead, it would unduly focus the reader's attention on the errors in the diurnal cycle, which are outside the scope of our study.

3) Page 13, Line 28-29: The atmosphere is not in radiative convective equilibrium at the scale of 600 km over 3 hours. If the models were, that would be completely wrong. To provide evidence, I attach an unpublished figure from my own work, which plots daily averages of atmospheric cooling derived from CERES observations against daily averages of rainfall from GPCP for increasing areas centered on a 1x1 degree gridpoint in the Tropical Western Pacific. If we consider the 90x90 degree area as close to RCE, it is evident that the averaging scale one could argue starts to approach it is about 30-40 degrees (i.e., $(3000\text{-}4000\ \text{km})^2$), far from the 600 km scale speculated about here. In fact, not surprisingly, at those scales, the atmosphere is a far from RCE as it can be, as the heavy rainfall is associated with large cloud fields that reduce the atmospheric radiative cooling substantially. More likely, at the 600 km scale the convection is in

balance with dynamical systems at the synoptic scale, which do exist in the tropics, and are perhaps relatively well captured by the models.

We agree with the reviewer and thank the reviewer for this suggestion. We have replaced the sentence in question with: “We hypothesize that these broader scales represent those at which simulated convection is in balance with the synoptic-scale, dynamical systems that produce precipitation, predictions of which should be highly similar among the models in the short, 2-day hindcasts we analysed.”

We replaced a similar sentence in the Discussion section with: “This convergence of model behavior may be enhanced by the fact that these data are from short (2-day) forecasts initialized from the same ECMWF analyses, which means that the representation of these dynamical systems are much more similar among models than if the data came from free-running climate simulations.”

4) Page 6, Line 7-8 and Figures 2 and 5: The bin below 0.5 * Delta x appears pointless and makes for an awkward plot. I suggest changing the txt to acknowledge the theoretical possibility of the existence of such a bin but then states that in practice it does not exist for this study. Then you can remove the unnecessary and distracting XXX columns from the figures.

We agree with the reviewer. In the revised manuscript, we have removed the bin below $0.5\Delta x$ from our analysis and the column from our revised versions of Figs. 2 and 5.