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 Anonymous Referee #2

We thank the reviewer for providing a thorough and generally positive 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 paper proposes a diagnosis method for precipitation of general circulation models

(GCMs) by using a native temporal and spatial grids and discusses dependency of temporal and spatial averages of precipitation. Precipitation behaviors of GCMs have been usually evaluated by climatological mean states. However, this study clearly shows that even if the climatological mean (or 3-hr average) precipitation is almost the same, its temporal and spatial behaviors are very different if analyzed by the native grids and original time step. This aspect of precipitation might affect large scale behaviors and hence must be more focused for the analysis, evaluations, or improvements of GCMs. The methodology is clear, and it implication is sound. Thus, I suggest publication of this study after minor revisions described below. Although the proposed diagnosis will be useful, the authors can go further more. In the text, the authors mention "persistency" or "intermittency" of precipitation. We need to compare many figures (e.g. Fig. 4 vs Fig. 7) to evaluate "persistency" or "intermittency". The authors should consider some quantifications of "persistency" and "intermittency", and show summary of these quantities of the models with difference samplings.

We agree with the reviewer and thank the reviewer for this suggestion. Reviewer #1 (Christian Jakob) 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 #1.

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 $5.6° \times 5.6°$ 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.

Specific comments

p. 3, L18, "Such diagnostics": It is not clear which "diagnostics" is referred to in this paragraph. Please clarify.

We were referring to diagnostics that estimate the coherence of precipitation in space and time, which we mentioned in the previous sentence. In the revised version of the manuscript, we have replaced "such diagnostics" with "diagnostics of precipitation coherence".

p. 4, L5-6, "Both products are derived from a combination of infrared and microwave sounders and calibrated against gauge data.": The authors should add more information on the difference between TRMM and CMORPH for readers who are not familiar to the details of the products of precipitation. In addition, since TRMM 3B42 is not solely

<div align="center">C3</div>

based on the TRMM data, it is not appropriate to call it "TRMM". The authors should make a remark on it if the abbreviation of "TRMM" is to be used.

We disagree with the reviewer that it is inappropriate to call the 3B42 product "TRMM". In fact, the creators of the dataset, the U.S. National Aeronautics and Space Administration (NASA), refer to the product as "TRMM 3B42". Please see their webpage here:

http://disc.sci.gsfc.nasa.gov/precipitation/documentation/TRMM_README/TRMM_3B42_re

The Tropical Rainfall Measuring Mission (TRMM) is a project, which includes a satellite that is—confusingly, we agree—also called TRMM. The microwave instrument on the satellite is often called the TRMM Microwave Imager (TMI). Some of the products that TRMM (the project) produces use data exclusively from TRMM (the satellite), while other TRMM (the project) products use data from multiple instruments, including TRMM (the satellite). The 3B42 product falls into the latter category. Many studies, and NASA itself, refer to the 3B42 product as "TRMM 3B42." It is not our place to correct this confusion; rather, we use the commonly accepted nomenclature.

We have added details to Section 2 about the TRMM 3B42 and CMORPH algorithms, including differences in how they produce their merged microwave–infrared precipitation estimates. We have also added a note that states that the TRMM 3B42 product is produced from multiple microwave sounders, not solely TMI. We have also added "3B42" to all figure captions and tables where we use TRMM 3B42 data.

p. 5, L23, "We find the central point in each region and extract the timeseries of precipitation.": I suggest that "find" should be replaced by "define" or an appropriate word.

We have replaced "find" with "select". We hope that this is acceptable.

p. 5, L26, "in Figs. 2b and 2c for CMORPH": These should be "Figs. 2c and 2d".

We agree with the reviewer. We have corrected this error by replacing "in Figs. 2b

<div align="center">C4</div>

and 2c" with "in Figs. 2c and 2d".

p. 6, L10, "these computations result in a matrix of correlations with distance and time, as shown in Fig. 2c.": I guess that this is for Figs. 2e and 2f.

We agree with the reviewer. We have corrected this error by replacing "in Fig. 2c" with "in Figs. 2e and 2f".

p. 6, L14, "Fig. 2b": This should be replaced by "Figs. 2c and 2d".

We agree with the reviewer. We have corrected this error by replacing "in Fig. 2b" with "in Figs. 2c and 2d".

p. 6, L16, "For the ranges shown here, the CMORPH 0.25° correlations decline more quickly with time than with space.": It is ambiguous to say which is "more quick" between time and space. Add more explanations.

We thank the reviewer for pointing out this ambiguity. On reflection, this sentence does not contribute anything meaningful to our discussion and is a likely source of confusion. We have removed it from the revised manuscript.

p. 8, L20, "The 1D histogram suggests that MetUM-GA3 oscillates between lighter (< 9 mm day^{-1}) and heavier (> 30 mm day^{-1}) rain rates, with almost no instances of moderate rates (9–30 mm day^{-1}).": This is an interesting behavior of precipitation of the MetUM-GA3. Please consider adding sample figures of time sequence and snapshot distribution of precipitation of MetUM-GA3.

We agree with the reviewer. We have added a figure to our revised manuscript (Fig. 7) that compares MetUM-GA3 to GEOS5. GEOS5 produces persistent rainfall and has a timestep and horizontal resolution similar to MetUM-GA3. In Fig. 7, we compare timeseries of precipitation at an example gridpoint in the Indian Ocean (0°, 90°E) for an example 48-hour forecast from our datasets (4 November 2010); we also compare snapshots of instantaneous precipitation rates for an example timestep (20:00 UTC

C5

on 4 November 2010). These figures confirm the results of our other diagnostics: precipitation in MetUM-GA3 is temporally and spatially intermittent, while precipitation in GEOS5 is temporally and spatially persistent. We have added a paragraph to Section 3.1 of our revised manuscript that discusses this figure and its implications.

We thank the reviewer for this suggestion, which has helped to provide further visual evidence of our main conclusions.

p. 8, L25, "The bi-modal 1D histogram suggests that most deep convection in MetUM-GA3 is strong.": It is not clear how "strong" or stronger than what? Please add more explanations.

We agree with the reviewer that this statement is not clear. We mean that most deep convection in MetUM-GA3 is as intense as it possibly can be, at that horizontal resolution and scientific configuration of the model. We have modified this sentence to state that "most deep convection in MetUM-GA3 is as strong as possible, given the horizontal resolution and scientific configuration of the model."

p. 9, L10: "Despite having the finest horizontal resolution" should be replaced by "Because having the finest horizontal resolution"?

We disagree with the reviewer's suggestion. As our explanation in the sentence following the one that the reviewer quoted states, we would expect (naïvely, perhaps) that finer horizontal resolution would increase spatial correlations, particularly when those correlations are measured as a function of the native gridpoint. A finer-resolution model has a smaller physical distance between gridpoints, which means that a precipitation feature of the same physical dimension would span more gridpoints in a finer-resolution model than in a coarser-resolution model. Thus, one would expect the finer-resolution model to have a higher spatial correlation, measured in native gridpoints. MetUM-GA3 has the lowest spatial correlation and the finest spatial resolution, which is at odds with this expectation. Therefore, we have maintained our use of "despite".

C6

We agree with the reviewer. We have added "for MetUM-GA3" to the end of this sentence.

We agree with the reviewer. We have corrected the formatting of the Δ symbol.

We agree with the reviewer. We were referring specifically to the lack of verifying observations for similar spatial domains and time periods to those used in our analysis of the 2-day hindcast data from the MJO inter-comparison project. We acknowledge that there are high-resolution observations, such as those the reviewer cites, that could be compared against model simulations, assuming care was taken to perform those comparisons at similar spatial and temporal resolutions, based on the results from our study. In fact, the comparison of high-resolution, convection-permitting model simulations against radar data is an area of ongoing research.

We have modified the sentence in question to read: "While there are no observation-based constraints on timestep rainfall for similar spatial domains and temporal periods as the model data analysed here . . . ". We believe this statement is accurate.

Further, we have added a short paragraph to the Discussion section that discusses the use of high-resolution radar data to validate simulated timestep precipitation, as the

reviewer suggests. This is the second paragraph of the Discussion section (Section 4) in our revised manuscript.

We agree with the reviewer. We have added a sentence to Section 2 to note this bias: "Both products have been shown to under-detect light rainfall rates (e.g., Tian et al., 2010)." Further, we note that we commented on the under-detection of light rainfall on page 11, line 34 of the original manuscript. We have added the Tian et al. (2010) reference above to that sentence as well.

We agree with the reviewer. We have removed the erroneous "s" after "Fig.".

We agree with the reviewer's first point. We have added the statement "(compare Fig. 9 to Fig. 4)" to the end of the sentence in question.

We also agree with the reviewer's second point. We have added metrics for spatial coherence and temporal persistence to our revised manuscript (see response to "General Comments" above). These metrics clearly demonstrate that averaging from timestep to 3-hr means increases the intermittency in models with more-persistent timestep precipitation (see Table 3 in revised manuscript). We have added the following sentence to the end of the paragraph in question: "Table 3 confirms that temporal averaging on the

native grid reduces inter-model variations in the temporal persistence summary metric, by increasing values for models with relatively low scores (e.g., MetUM-GA3, CanCM4, CNRM-AM) and reducing the values for models with relatively high scores (e.g., CAM5, MIROC5, MRI-AGCM3)."

Please note that we revised the sentence in question to read "Conversely, models with more persistent timstep precipitation (e.g., GEOS5, MRI-AGCM, CAM5 and MIROC5) display reduced persistence when data are averaged to 3-hr means" to prevent the mis-understanding that these models were the most intermittent at the 3-hr among all the models considered. In fact, these models are still the most persistent models, as our new summary metrics show, but they show reduced persistence (and lower values of our metric) for 3-hr data than for timestep data.

p. 11, L28-29, "SPCAM3, ECEarth3 and CanCM4 are perhaps closest to TRMM and CMORPH, but are still more persistent.": Again, it is not clear how these models are close to the observations. Please consider quantification of the persistency.

We agree with the reviewer. As above, we have created new metrics to quantify temporal persistence and spatial coherence, which can be found in Table 3 of the revised manuscript. These metrics demonstrate that SPCAM3, ECEarth3, CanCM4 and CNRM-AM show temporal persistence in 3-hr precipitation that is most similar to TRMM and CMORPH, but that all models are too persistent with respect to both datasets. We have added references to our summary metrics to the paragraph in question, noting that these four models show values closest to the satellite-derived observations.

p. 11, L34, "We note that there are also differences between TRMM and CMORPH over this short period: CMORPH displays more frequent light precipitation than TRMM, which has been shown to under-detect light rainfall (Huffman et al., 2007, e.g.,). TRMM is more intermittent than CMORPH.": The authors should note why these differences come from between the two observations. "(Huffman et al., 2007, e.g.,)" should be "(e.g., Huffman et al., 2007)".

We disagree with the reviewer's first point. The purpose of our manuscript is not to understand the difference between the TRMM and CMORPH datasets. Because we do not have direct access to the algorithms used to produce the TRMM 3B42 and CMORPH products, we cannot investigate the reasons for the differences in temporal coherence between the two datasets. Any hypotheses would be mere conjecture that would not be worthy of publication. Instead, we compare these datasets only to provide a measure of observational uncertainty in our diagnostics. We have added a sentence to Section 2 to clarify this: "We employ two observation-based datasets to provide a measure of observational uncertainty in our diagnostics."

We agree with the reviewer's second point. We have corrected this errors by moving the "e.g.," to the beginning of the parenthetical citation.

p. 12, L3, "All models display higher correlations": Add "at 3h interval (Fig. 10b)" at the end of this sentence.

We agree with the reviewer. We have added "When using 3-hr data" to the start of the sentence in question and a reference to Fig. 10b at the end of the sentence.

p. 12, L11, "Spatial averaging reduces timestep intermittency in all models (Fig. 11).": The observations of TRMM and CMORPH should also be added to Fig. 11.

We disagree with the reviewer. It is not possible to show data for TRMM and CMORPH in Fig. 11, because Fig. 11 shows timestep data from models, not 3-hr data. Comparing timestep data (12–60 minutes) from models to 3-hr data from TRMM and CMORPH is not valid. Fig. 14 in our manuscript compares 3-hr data from TRMM and CMORPH to 3-hr data from models, when all datasets have also been interpolated to a $5.6° \times 5.6°$ horizontal grid. This is the equivalent of Fig. 11 for 3-hr data; it provides a fair comparison between models and satellite-derived observations.

p. 12, L18-20, "This suggests that using a common horizontal grid or a common timescale does not necessarily create a fair comparison between models, due to differ-

We agree with the reviewer. We were referring specifically to the diagnostics we produced in this study. Of course, interpolating to a common grid or timescale is appropriate for some purposes, outside the remit of this study. We have amended this sentence to read "For the purposes of these diagnostics, using a common horizontal grid or a common timescale does not necessarily create a fair comparison between models . . . ".

We do not know exactly which aspect of this sentence the reviewer finds unclear. Our point is that when one averages over a broad region, such as our $5.6° \times 5.6°$ boxes, two models can produce the same precipitation rate (or spectrum of rates) from different combinations of precipitation frequency and intensity on the native gridscale. For example, Model A might produce very infrequent, but very heavy precipitation at its native resolution. Model B might produce very frequent, but light precipitation. When data from these models are averaged across a $5.6° \times 5.6°$ box, the average precipitation rates might be the same. In our study, MetUM-GA3 is like Model A. Within each $5.6° \times 5.6°$ box at any given time, there are very few precipitating gridpoints, but those gridpoints show heavy precipitation. Our new Fig. 7—which the reviewer suggested—demonstrates this well.

We have revised this sentence to read "For instance, the comparison of Fig. 4a and Fig. 12a suggests that MetUM-GA3 likely has only a few precipitating gridpoints in each $5.6° \times 5.6°$ region, but that those points show very heavy precipitation (e.g., 90–130 mm day^{-1}), as indicated in Fig. 8a."

C11

We agree with the reviewer. We have corrected this error by deleting the extraneous right-hand parenthesis.

We agree with the reviewer. Again, we were referring to the lack of verifying observations for similar spatial domains and time periods as the ones covered by the 2-day hindcast dataset we analysed. We have modified the sentence in question to read: "Although there are no verifying observations for the model timestep data that cover comparable spatial and temporal domains . . . ."

We agree with the reviewer. We have corrected this error by replacing "Fig. 4a" and "Fig. 4b" with "Fig. 7a" and "Fig. 7b", respectively.

We agree with the reviewer. We have added a short paragraph to the Discussion section of our revised manuscript (second paragraph of Section 4), in which we discuss the potential use of single-column model simulations to investigate the causes of undesirable intermittency in simulated precipitation, including the references that the reviewer pointed out. However, SCM simulations can only address errors in the sub-

C12

gridscale physics, not errors in the coupling between that physics and the resolved dynamics. Hence, we have added the following sentence to the revised manuscript: "Model development efforts to reduce or remove undesirable intermittency may involve single-column model experiments, in which the effects of changes in sub-gridscale physics can be isolated from feedbacks through the resolved dynamics (e.g., Satoh and Hayashi, 1992; Takata and Noda, 1997; Woolnough et al., 2010), although we stress that physics–dynamics coupling may have a substantial effect on the model behaviors and diagnostics presented here."