Anonymous Referee #1

The authors used the spatial plotting method to identify significant patterns of fully coupled GCMs from ten sets of NMME. Overall, the manuscript is well crafted with clear structures.

We appreciate the positive comments.

Some grammatical errors exist and need careful proofreading. I have some minor comments.

Thank you for the constructive comments. We have improved the paper accordingly and proofread the whole paper.

1. Do you use some bias correction or downscale methods for the NMME forecasts? If yes, suggest to give some details.

Thank you.

"In the analysis, the attention is paid to the retrospective forecasts

$$F_{GCM} = \left[f_{s,l,n,y,x} \right]_{GCM} \quad (1)$$

In Eq. (1), f represents forecast values that are specified by the 5 dimensions; F, which is the set of forecasts, is marked by the GCM that generates the forecasts. It is noted that in NMME, F_{GCM} are raw forecasts generated by GCMs and are not bias-corrected or downscaled." (Page 4, Lines 103 to 107)

2. Suggest to give the difference between the anomaly correlation and the spatial plotting method. Or the advantage of the spatial plotting method against the simple anomaly correlation.

Thank you for the constructive comment. The advantage of spatial clustering is highlighted in the introduction:

"Spatial plotting with latitude and longitude has been extensively used to handle the dimensionality for the verification of GCM forecasts [Kirtman et al., 2014; Hudson et al., 2017; Slater et al., 2017]. The fact that forecasts are commonly generated by GCMs

as grid-based data makes spatial plotting a particular tool of choice for verification [Merryfield et al., 2013; Saha et al., 2014; Jia et al., 2015]. As to anomaly correlation, spatial plotting overcomes tedious eyeball search by grid cell and is effective in locating where there is a good correspondence between forecasts and observations and where the correspondence is not satisfactory [Luo et al., 2013; Saha et al., 2014; Crochemore et al., 2016; Zhao et al., 2018, 2019b]. Similarly, spatial plotting applies to other verification metrics, such as bias and CRPS, and facilitates the examination of forecast attributes [Hersbach, 2000; Gneiting et al., 2007; Kirtman et al., 2014].

The extensive use of spatial plotting underlines the importance of testing the significance of spatial patterns. In spatial statistics, one of the fundamental issues is "are the spatial patterns displayed by the spatial plots significant in some sense and therefore worth interpreting?" [Cliff and Ord, 1981; Anselin, 1995; Getis, 2007]. However, the test of significance is commonly missing in the spatial plotting of GCM forecasts. In other words, verification metrics, such as anomaly correlation, are calculated for each grid cell and then shown as they are. To some extent, the interpretation of predictive performance depends on the color schemes, which are selected subjectively to represent the scale of verification metrics. There is the first law of geography – "everything is related to everything else, but near things are more related than distant things" [Tobler, 1970]. As to spatial plotting, the indication is that when verifying forecasts at one grid cell, attention also needs to be paid to forecasts at surrounding grid cells. For anomaly correlation, a grid cell with high correlation between forecasts and observations can be surrounded by grid cells with similarly high correlation, or by grid cells with low correlation. In the former case, the grid cell is located in a region where the GCM forecasts tend to perform well. But in the latter case, the high correlation can be a suspicious outlier. Moreover, previous studies observed grid cells with negative anomaly correlation, i.e., large (small) values of forecasts correspond to small (high) values of observations [Zhao et al., 2017b, 2018, 2019b]. In such a case, forecasts are cautiously wrong. Therefore, it is critical to characterize the different cases in spatial plotting and test whether the spatial patterns are significant and worth further attention.

In this paper, we are motivated to introduce spatial statistics [e.g., Di Luzio et al., 2008; Lu and Wong, 2008; Woldemeskel et al., 2013] to investigate the spatial plotting of anomaly correlation at the global scale…" (Pages 2 to 3, Lines 50 to 74)

3. Line 102: please check the start updated date of real-time forecasts.

Thank you for the insightful comment. More details on real-time forecasts are added:

"Ten sets of precipitation forecasts, as well as CMAP observations, in the NMME are downloaded from the International Research Institute at the Columbia University (https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/). Their retrospective forecasts are complete in the period from 1982 to 2010 [Merryfield et al., 2013; Saha et al., 2014; Jia et al., 2015]. In the meantime, their real-time forecasts are updated periodically in a slightly different setting; for example, CFSv2 forecasts are generated since January 2011 using initial conditions of the last 30 days, with 4 runs from each day (https://www.cpc.ncep.noaa.gov/products/CFSv2/CFSv2_body.html)." (Page 4, Lines 98 to 103)

4. Suggest to add the equation of the anomaly correlation calculation and give some details about the climatology for the anomaly.

Thank you for the suggestion. We have added the equation:

"The start time s in Eqs. (1) and (2) comprises year k, i.e., 1982, 1983, …, 2010, and month m, i.e., January, February, …, and December. The predictive performance of GCM forecasts exhibits seasonality [Yuan et al., 2011; Zhao et al., 2017a, 2017b]. Accordingly, in the analysis, forecasts are selected by fixing m while varying k, e.g., pooling forecasts initialised in June 1982, June 1983, …, June 2010. The anomaly correlation is calculated by relating forecasts to the corresponding observations

$$r = \frac{\sum_k \left(rf_k - \overline{rf}\right)\left(ro_k - \overline{ro}\right)}{\sqrt{\sum_k \left(rf_k - \overline{rf}\right)^2}\sqrt{\sum_k \left(ro_k - \overline{ro}\right)^2}} \quad (2)$$

The above formulation deals with k and omits other dimensions, including m, l, y and x, for the sake of simplicity. In Eq. (3), rf_k (ro_k) is the rank of year k's forecast ensemble mean (observation) in the 29 years' ensemble mean (observations); and $\overline{rf}$ ($\overline{ro}$) is the mean value of rf_k (ro_k). In general, the anomaly correlation characterises how well large (small) values of ensemble mean correspond to large (small) values of observations. Good (poor) correspondence makes r tend towards 1 (–1)." (Page 5, Lines 112 to 119)

5. Suggest to give some explanation for the forecasts of total precipitation in three months.

Thank you. More information is provided:

"The spatial clustering is performed for the anomaly correlation across the ten sets of forecasts in NMME. In the analysis, the attention is mainly paid to June, July, and August (JJA), which are generally boreal summer and Austral winter. Specifically, the start time of the forecasts is June, and the forecasts at the lead times of 0, 1, and 2 months are aggregated to form the seasonal forecasts. In the meantime, forecasts initialized in September of total precipitation in September, October, and November (SON), forecasts initialized in December of total precipitation in December, (the next) January, and (the next February) (DJF), and forecasts initialized in March of total precipitation in March, April, and May (MAM) are also investigated, with the results presented in the supplementary material." (Pages 7 to 8, Lines 168 to 174)

6. Are captions 4.1 and 4.2 the same? Please check.

We are very sorry for the typo. The captions have been modified in the revision:

"4.1 Anomaly correlation in JJA" (Page 8, Line 175)

"4.2 Anomaly correlation and its spatial lag in JJA" (Page 10, Line 209)

7. Line 320: The authors showed the spatial extents of clusters vary by season. Suggest to give more explanations/reasons for this.

Thank you very much for the insightful comment. In the revision, we have added a new section and illustrated the results of the other three seasons:

"4.5 Frequency of the case HH in SON, DJF, and MAM

Besides JJA, spatial clustering has been performed for the anomaly correlation of GCM seasonal forecasts of total precipitation in SON, DJF, and MAM. Similarly, it is observed that the anomaly correlation varies across the globe (Figures S1, S4, and S7 in the supplementary material), correlates with its spatial lag (Figures S2, S5, and S8), and exhibits significant spatial patterns (Figures S3, S6, and S9). In addition to Figures 4 and 5, the frequency of the case HH is counted for the other three seasons and shown in Figures 6 and 7.

ENSO is one of the most important drivers of global climate [Mason and Goddard, 2001; Saha et al., 2014; Bauer et al., 2015], and the CPC of NOAA has summarized the correlation between ENSO and global precipitation in different seasons (https://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/regressions/geplr.shtml). In this paper, the results in Figure 6 are associated with the global effects of ENSO.

In SON, the CPC shows that ENSO correlates negatively with precipitation in Eastern Australia and Southeast Asia, and positively with precipitation in part of Middle East and East Africa. From the upper part of Figure 6, it is observed that the frequency of the case HH is high in these regions. In DJF, ENSO is shown to correlate positively with precipitation in Southern North America and negatively with precipitation in Northern South America. In these two regions, the frequency of the case HH is high (middle part of Figure 6). In MAM, ENSO is illustrated to correlate negatively with precipitation in part of Southeast Asia, Eastern Brazil, and Eastern Australia. Therein, the frequency of the case HH seem to be high (lower part of Figure 6). Therefore, as previous studies found that GCMs in NMME generate skilful forecasts of ENSO [e.g., Kirtman et al., 2014; Saha et al., 2014; Zhang et al., 2017], Figure 6 suggests that the skill, as is indicated by anomaly correlation, of GCM forecasts in NMME can also be related to ENSO. In Figure 7, the percentage and cumulative percentage of the frequency of the case HH are illustrated for SON, DJF, and MAM. Similar to Figure 5, the results show the complementarity among the ten sets of forecasts.

Besides ENSO, there are other drivers of global climate. For example, North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) extensively affect the climate in Europe, Asia, and North America [Hurrell et al., 2001; Ambaum et al., 2002]. Several sea surface temperature indices of the Atlantic and Indian Oceans and ENSO jointly impact the climate in Africa [Rowell, 2013]. As can be observed from Figures 4, 5, 6, and 7, there is still substantial room for improvement of seasonal precipitation forecasts for large parts of Europe, Asia, and Africa. The overall neutrally skilful precipitation forecasts in these regions can possibly be due to that GCM formulations of other climate drivers are not as effective as the formulations of ENSO. In the meantime, the difficulty of global climate forecasting due to spatially-temporally varying teleconnections between regional precipitation and global climate drivers is noted [Merryfield et al., 2013; Saha et al., 2014; Jia et al., 2015; Hudson et al., 2017; Kushnir et al., 2019].

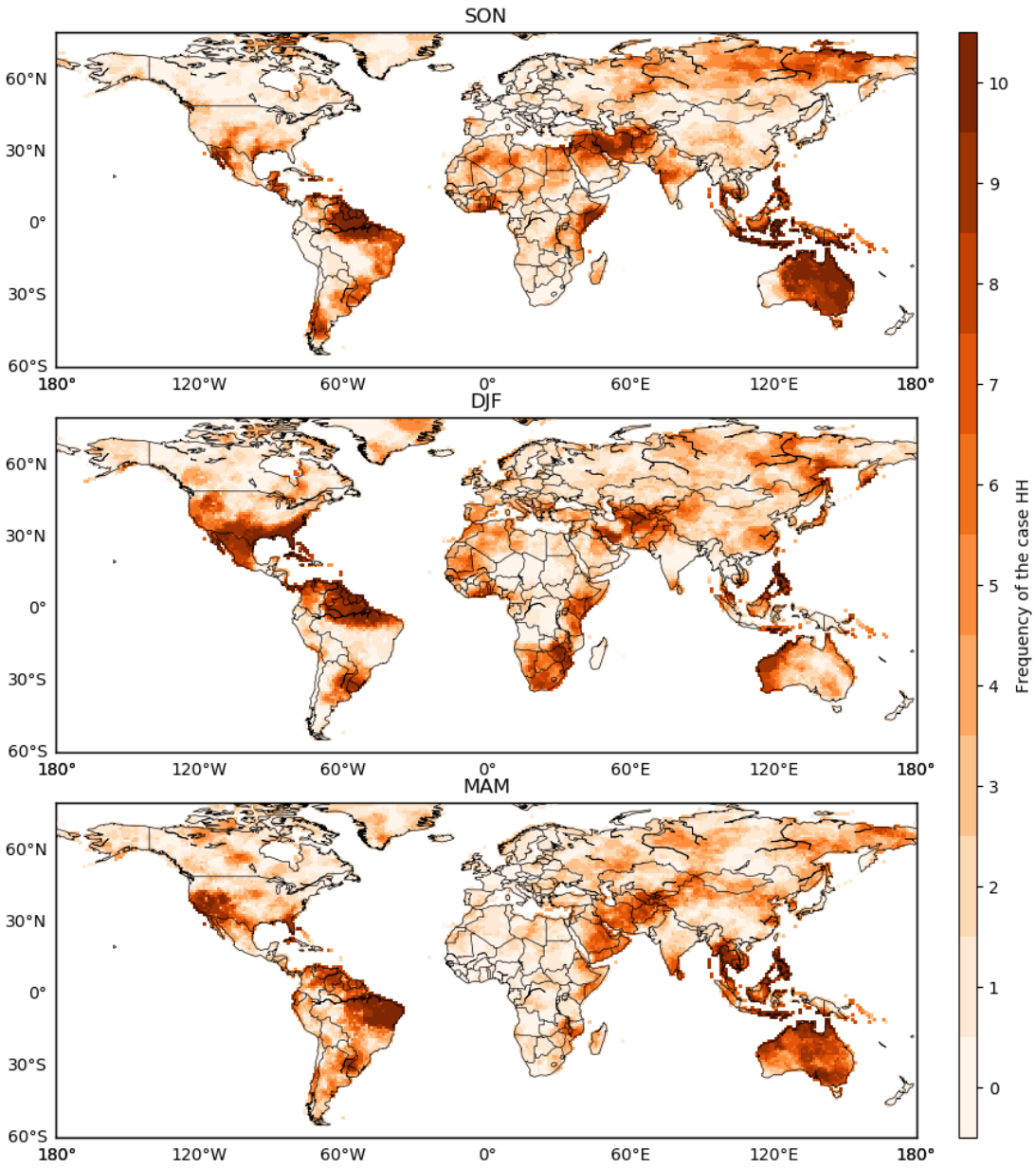

Figure 6: As for Figure 4, but for SON, DJF, and MAM

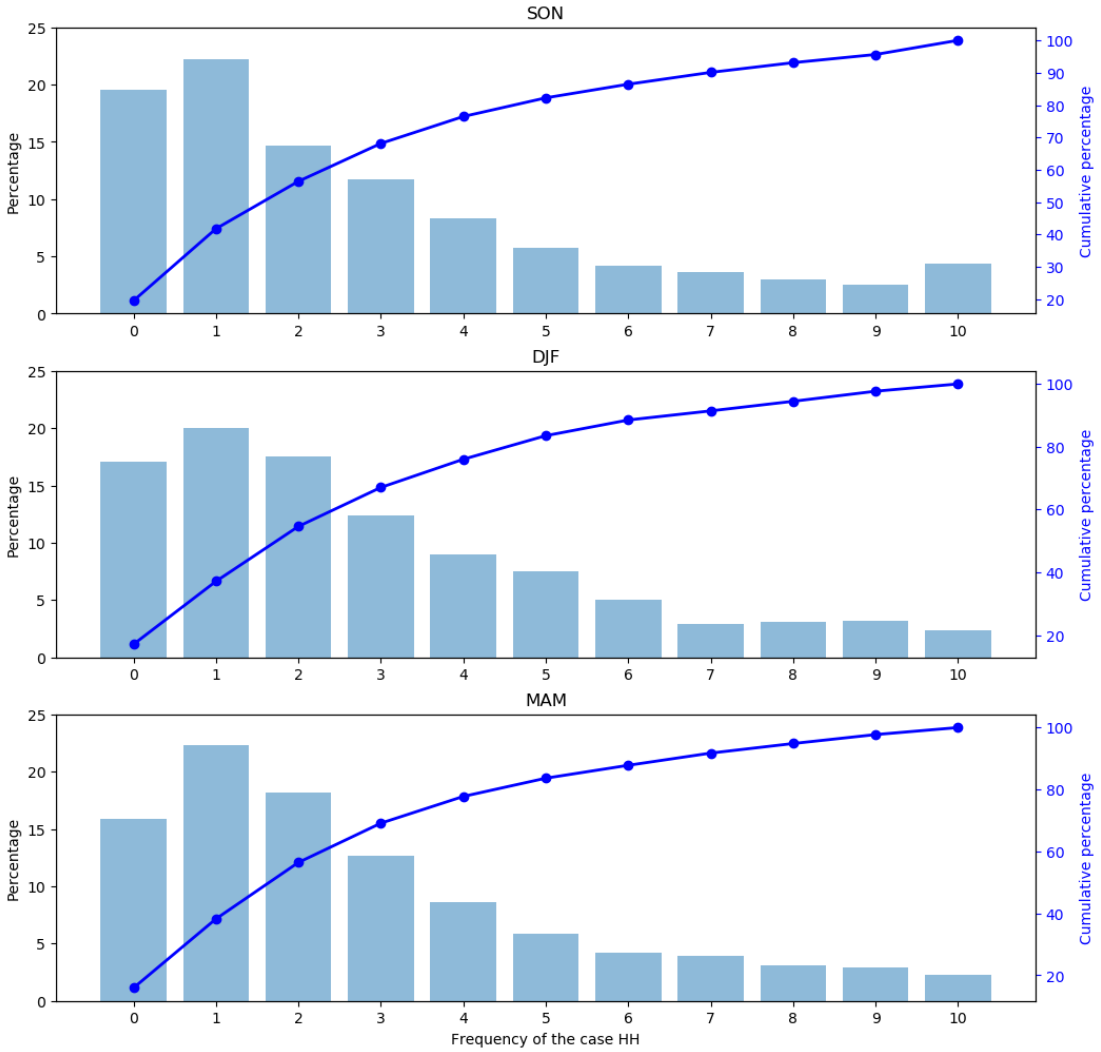

Figure 7: As for Figure 5, but for SON, DJF, and MAM”

(Pages 16 to 19, Lines 311 to 345)