We thank the reviewer for taking the time to read our manuscript and for the comments. This document is a point-by-point reply to the comments.

General comments:

In this study, the authors presented the performance of ensemble MJO forecasts using a stochastic weather generator based on circulation analogs. As the MJO is an important source of predictability on the subseasonal time scale, a useful forecast of the MJO is of significant scientific and practical values. Although there have been quite a few studies on MJO forecasts, this study uses a unique approach which is novel in this area. The result is interesting. It shows that a useful skill of the MJO can be achieved at a lead-time of 40 days, which is considerably longer than most dynamical and statistical models. The paper is in general clearly written, although some clarifications and edits are needed. A little more reasoning for choice of variables and region for the analogs and explanation of the results would improve the paper.

Specific comments

The MJO is a planetary-scale tropical disturbance, but the tropical region for the analog calculation in the Indian Ocean (Fig. 2) is quite small. It is a little surprising that Z500 in such a small region can provide information for the MJO evolution. On lines 210-214, one reason for the choice is given which is based on the composition of the RMM index. This may explain why OLR is not used, but RMM does not include Z500 either. The MJO has a baroclinic structure, but 500 hPa is in the middle troposphere that cannot capture the vertical structure. In addition, geopotential height in the tropics does not represent well wind fields. Why not using zonal winds at upper or lower troposphere? Some more explanation on the choice of variable, region, and level would be very helpful.
⇒ **We computed analogs from other regions mentioned in figure B1. However, we obtain better results for the forecast scores by focusing on the small area we used. This is explained by the higher quality of analogs (lower distance).**
As mentioned in our manuscript, we computed analogs with OLR, which is a driver of the MJO. However, the forecast skill with OLR analogs is lower than for Z500 (with analogs over that region). Z300 analogs, which are close to where the MJO takes place, do not yield significant improvement over Z500 analog skill scores.
We also computed analogs from the zonal wind at the upper and lower troposphere (250 and 850 hPa). As shown in the figure below (Figure 1), the wind at 250 hPa, 850 hPa, and the ORL do not improve the bivariate correlation and RMSE forecast skill of the MJO index for a longer lead time (above 20 days) over Z500 or Z300, despite the fact they are the driver of the MJO.
Therefore, we take in our approach of forecasting RMM indices of the MJO the minimal physical assumptions on how they are constructed and use “all-purpose” predictor fields (Z500 or Z300). It turns out that geopotential height analogs lead to better scores (COR and RMSE) than other variables that are direct predictors of the MJO.
One reason for this apparently surprising behavior is that the composites of OLR and wind speed highly depend on the phase of the MJO (see Figure 1 of Kim et al., J. Clim., 2018). As our analog approach is constrained by the choice of a geographical

region, it misses the spatio-temporal variability of OLR and wind speed during the MJO. Geopotential heights, although less physically and dynamically relevant for the MJO, are more appropriate predictors from the statistical and mathematical constraint of the analog-based method. This will be discussed in the text, although we acknowledge that this is a conjecture of ours.

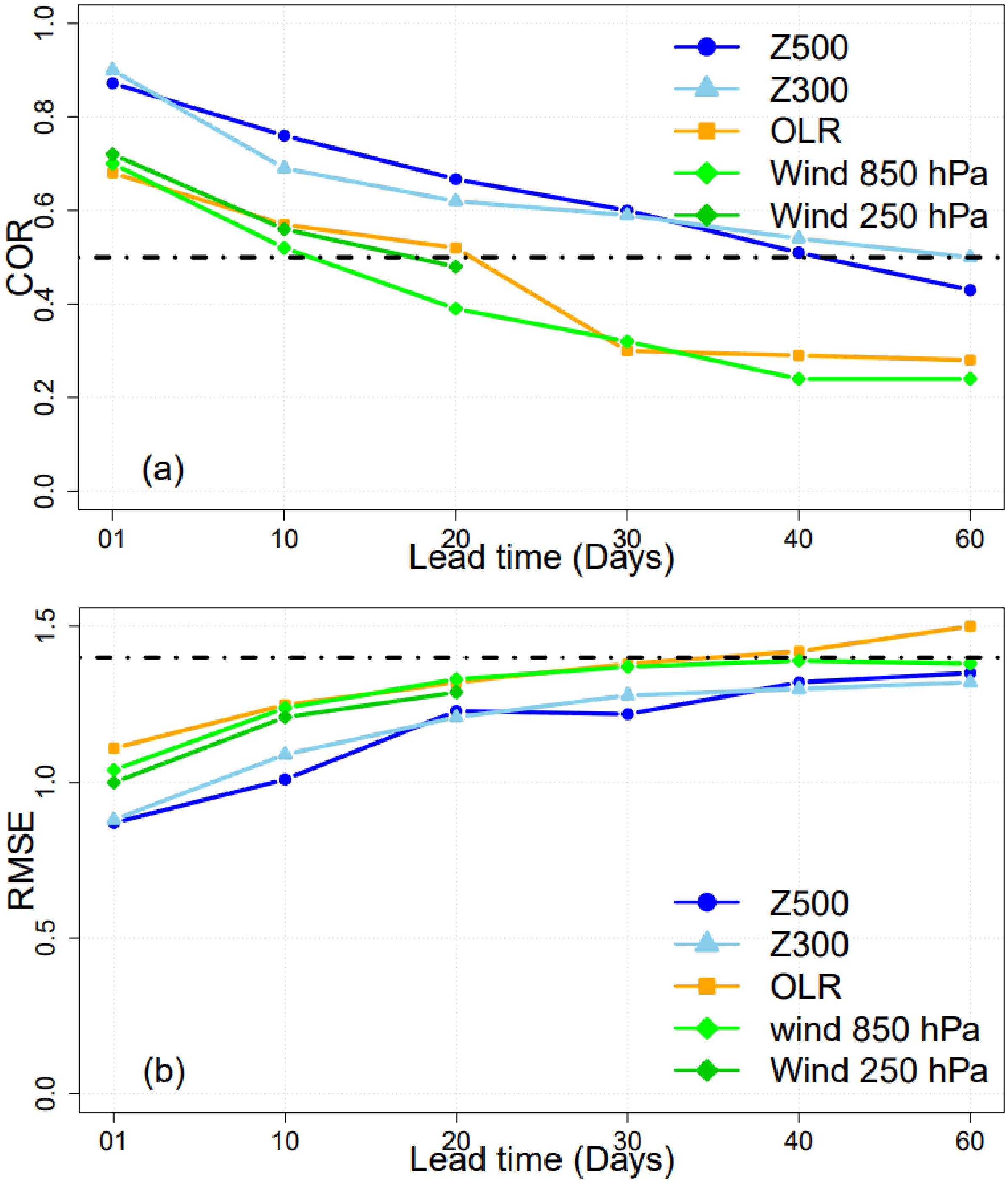

Figure 1. COR (a) and RMSE (b) values for lead times of the RMMs forecasts from 3 days to 90 days using analogs of OLR, wind at 250 hPa, wind at 850 hPa, Z300 and Z500.

Some justification for the choice of region is given on lines 217-219. The dependence of MJO forecast skill on initial phase is in fact not conclusive in previous studies. It would be interesting to see how this is the case in this study, i.e., the dependence of MJO skill on the initial phase. It would be interesting to see the skill dependence on MJO amplitude as well.

⇒ **The chosen region for the analogs is rather pragmatic, as the spatial correlation of Z500 shows meaningful structures with RMM indices (Figure 2 in the manuscript). We show below (Figure 2) the dependence of the CRPS values when the forecasts are initiated in each of the MJO phases (for this SWG configuration of analogs) and when the lead time is 10d. The differences are fairly small. The paragraph (ll. 217-219) will be rewritten to reflect that our choice is not based on a physical consensus on MJO forecasts, contrary to what we had suggested, but is just coherent with some aspects of the paper of Kim et al. (2018).**

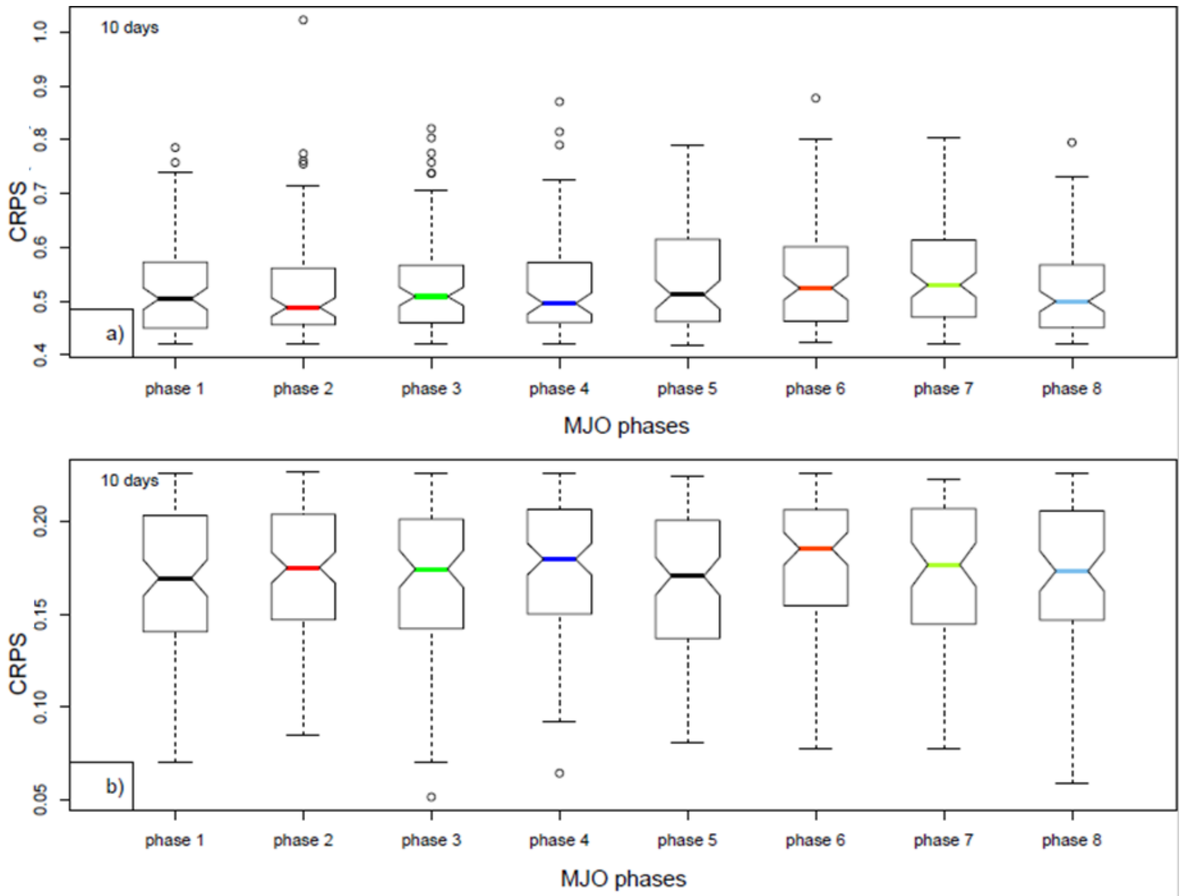

Figure 2. Relation between CRPS and MJO phases. (a) CRPS values above the 75th quantile and (b) CRPS values above the 25th quantile.

Section 6: Some introduction is needed for the two hindcasts of numerical models POAMA and ECMWF. More information on model resolution, version, ensemble size, hindcast period, etc., should be provided. A comparison as in Fig. 10 may not be very meaningful when these forecasts are for the different periods.

⇒ **We will provide more descriptions of the numerical models. In Figure 10, the forecasts that we used for the comparison from the ECMWF and machine learning are done over the same period.**

Minor comments:

Line 4: first two principal
⇒ **This will be corrected.**
Line 74: an MJO event
⇒ **This will be corrected.**
Lines 81-82: “over the region covering 15N-15S” is redundant.
⇒ **This will be corrected.**
Figure 2 caption last sentence: It seems the case for RMM2. How about RMM1?
⇒ **This will be corrected.**
Line 228: “other atmospheric circulations” à “other atmospheric variables”
⇒ **This will be rephrased.**
Line 278: the ensemble spread is increasing, instead of decreasing.
⇒ **This will be corrected.**
7b: How is the bias calculated? Is it the average bias of RMM1 and RMM2?
⇒ **The bias is the average bias of RMM1 and RMM2. It is calculated as follows:**
$bias\ =\ Ensemble\ average_{RMM1} - obs_{RMM1}$

Line 283: remove “the” in front of “a similar”
⇒ **This will be corrected.**
Line 284: A large RMSE does not necessarily mean a large spread.
⇒ **Indeed. This will be mentioned in the text.**

Line 309: Vitart (2017) also found that the MJO skill is higher in JJA for the ECMWF model
⇒ **Thank you for this reference. We will add it in our discussion.**

Line 351: machine learning
⇒ **This will be corrected.**

Reference:

Vitart, F., 2017: Madden-Julian Oscillation prediction and teleconnections in the S2S database. Quarterly Journal of the Royal Meteorological Society, 143, 2210-2220.