**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.**

Summary:

In this study a stochastic weather generator (SWG) based on the model analogs of the atmospheric circulation is formulated to predict the daily MJO index for a subseasonal lead time. The SWG method adopted by this study has been used to forecast climatic variables, precipitation and the North Atlantic Oscillation by the same authors, and this is the first time that they extend this method for the MJO prediction. The performance of the proposed method is compared against persistence, climatology, and state-of-the-art numerical models. In general the proposed method shows superior performance than both persistence and climatology forecast. Comparison against the full GCM model such as ECMWF forecast shows that the proposed model show larger RMSE (lower COR) than ECMWF model forecast for the 20-day forecast, but smaller RMSE (higher COR) than ECMWF model forecast for days 20-60.

In general, this paper is well written, and the method they proposed is interesting. One significant advantage of their method is the substantially low computational cost compared to a full GCM model (since they use the past model outputs to find the analog), but with superior performance than a full GCM model forecast for days 20-60.

It would be preferable if the authors can clarify why different variables are chosen to form the MJO index and the analog. Besides, I have some other comments shown below.

Recommendation: Major revision

Major Comments:

1. Why different variables and areas are used for RMM, and analog calculation?

Around Line 59: RMM1/2 are calculated from the satellite-derived OLR, and zonal wind at 250hPa, and 850hPa.

Around Line 84: analogs are calculated from the modeled geopotential at 500hPa, 300hPa, and OLR daily data.

Based on these, it seems you are using different variables to calculate RMM and analogs.

**⇒ The RMM1 and RMM2 are computed based on the wind at 250, 850 hPa and the ORL. In our study we have not computed the RMMs ourselves: we used RMMs from the IRI repository. However, we computed the analogs from other atmospheric variables (Z500, Z300 and OLR) as mentioned in the manuscript.**

**Although we do know how RMM indices are computed, this paper deliberately makes no (or very few) assumptions on their computation, and assesses how an "all purpose" predictor (e.g. Z500 that is not included in the definition of RMM) can predict RMM. In practice, analogs of OLR or zonal wind speed at 850 hPa were tested for the prediction, but the results are not optimal. One of the reasons stems from an examination of Figure 1 by Kim et al. (J. Clim., 31, 23; 10.1175/JCLI-D-18-0210.1, 2018), which shows the large longitudinal extent of OLR and wind speed anomalies. On such a large "window", the analogs yield rather large distances or low correlations. This implies that the analog SWG gets low skill scores, because the analogs are not very informative. The OLR or zonal wind speed analogs were computed on the optimal window obtained for Z500 or Z300 (figure below), and we find that the COR and RMSE scores are lower than for Z300 and Z500. This is because this "small" window is not appropriate for OLR or wind speed, as reflected in Figure 1 by Kim et al. (2018). This is a potential caveat (or feature) of analogs. The analog geometry needs to be imposed a priori in a rather simplistic way, which does not follow the spatio-temporal features of the MJO, which are known independently. This will be discussed in the text.**

My questions are:

(1) what is the motivation to use geopotential at 500hPa & 300hPa, instead of zonal wind at 250hPa and 850hPa, to calculate the analog?
⇒ **As stated above, the motivation is based on trial-and-error computation of forecast skill scores with analogs from various fields and geographical window sizes. The analog method does not seem optimal for forecasts with OLR and zonal wind due to the complexity of their spatio temporal structures. This will be mentioned in the discussion section.**

(2) Have you tried to use zonal wind at 250hPa and 850hPa to calculate the analog?
**We considered analogs from the zonal wind at 250 hPa and 850 hPa. The result is shown in the Figure below. The analog of the zonal wind does not help to get better forecast skill (bivariate correlation and the RMSE) than Z500 or Z300 analogs.**

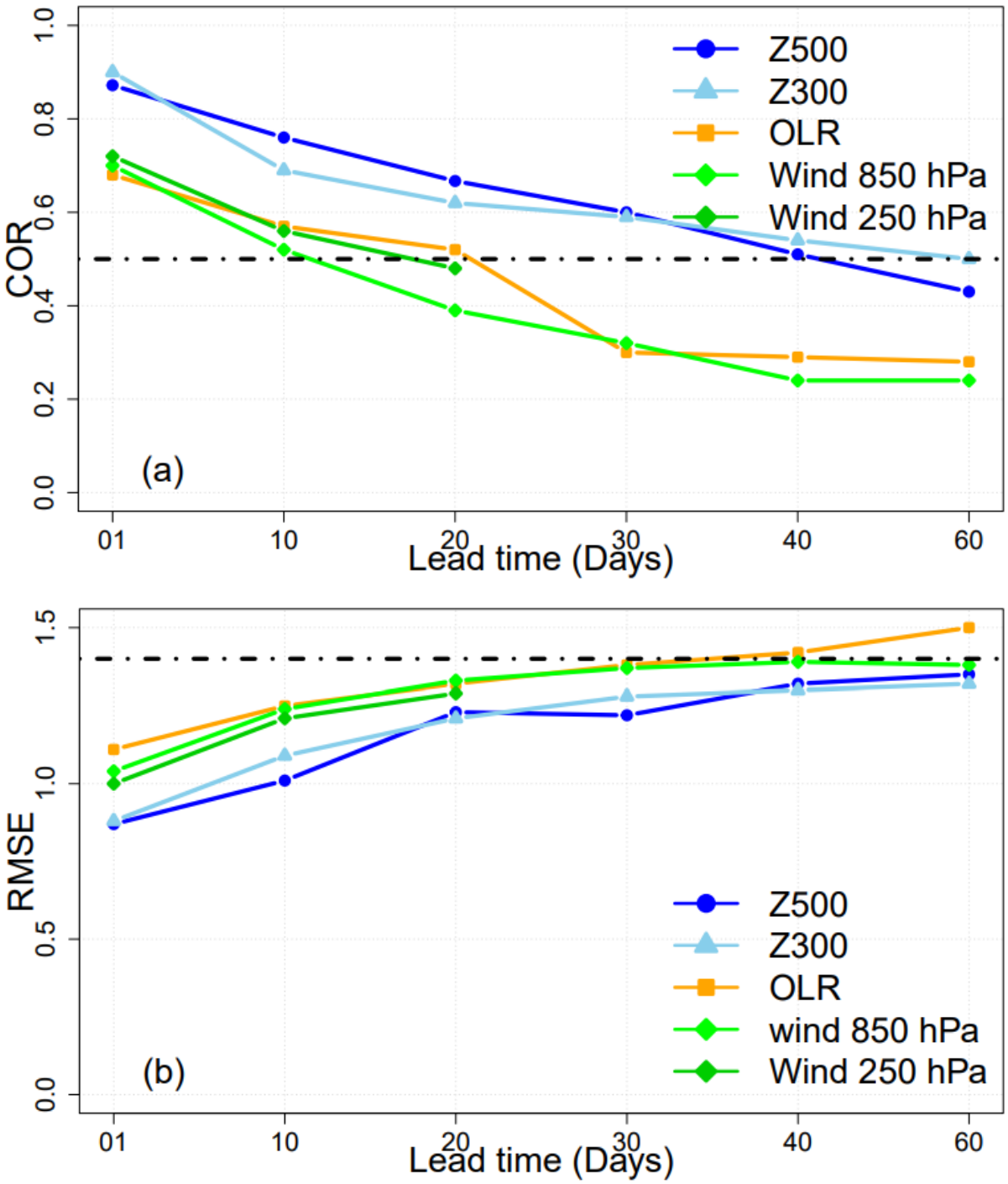


**Figure. COR (a) and RMSE (b) values for lead times of the RMMs forecasts from 3 days to 90 days using analogs of OLR, wind at 250 hPa, wind at 850 hPa, Z300 and Z500.**

(3) For the analog computation, you used the model data from NCEP. Can you indicate which reanalysis datasets you are using? Are you using CFS-R from NCEP?

⇒ **We used NCEP reanalysis data: NCEP-DOE AMIP-II Reanalysis (R-2): M. Kanamitsu, W. Ebisuzaki, J. Woollen, S-K Yang, J.J. Hnilo, M. Fiorino, and G. L. Potter. 1631-1643, Nov 2002, Bulletin of the American Meteorological Society.**

(4) Around line 59: RMM1&2 are calculated over the region between 15 deg N/S, while the analog is calculated based on the Indian ocean (around Line 100). I understand this is because Indian ocean is the onset place where the MJO occurs. My question is: is it necessary to only form the analog only based on the Indian Ocean? It seems that the analog formation process can be easily extended to later regions where MJO occurs. This might help the case where the initial signal is not well-captured by the initial analogs in the Indian Ocean.
⇒ **Indeed we computed analogs for other regions as shown in Figure B1. We tried to check the influence of the regions on the CRPS, but we did not notice any dependence.**

2.It might be better to reorganize the section 3.2 “Configuration of the stochastic weather generator”
⇒ **Ok, we will rephrase this section.**

(1) Around Lines 110-120: Is it possible to plot some schematics to illustrate the SWG process?
⇒ **We will add more explanations.**

(2) Line 110: “The random selection …that are computed are the products of two weights…rules”: rephrase the sentence. Also, it would be better to write an explicit equation combining \w_k^c and \w_k^\{Phi}. Though Lines 111-118 mention how the three w terms related, but it is better to reorder the sequence here and show the explicit equations for each w term.
⇒ **We will reformulate and explicit this part further.**

(3) Line 129 “The persistence and climatological forecasts are randomized by adding a small Gaussian noise”: Can you further clarify what kind of Gaussian noise did you add? How did you determine the magnitude of the variance of Gaussian noise (Any justification)?
⇒ **We use white Gaussian noise (mean=0, standard deviation = 1).**

Page 30, Figure B1: Why there are so many triangular white zones in the figure? What variable (indicated by color) is plotted in this figure?
⇒ **We will modify this figure. Figure B1 shows the different regions where we computed the analogs.**

Minor Comments:

Line 11: “We compare our SWG forecast with other forecasts of MJO”: It might be better if you can give a short summary (1-2 sentences) of the advantages of your methods over other MJO prediction methods in summary
⇒ **We will add a short summary.**

Line 65 “For this paper, …, the RMM1 and RMM2 allow to… (2004)”: rephrase the sentence.

⇒ **We will rephrase this sentence.**

Line 74 “For instance, we consider that there is a MJO event when A(t)>=1”. Why the value 1 is selected (any references)?

⇒ **Indeed the MJO is declared active when the amplitude A(t) is above or equal to one.**

**The reference is:**

**Lafleur, D. M., Barrett, B. S., & Henderson, G. R. (2015). Some Climatological Aspects of the Madden–Julian Oscillation (MJO), Journal of Climate, 28(15), 6039-6053. https://doi.org/10.1175/JCLI-D-14-00744.1.**

Figure 2: Is it possible to overlay the contour with p-values 0.05 so I know the correlated areas outside your boxed area.

⇒ **We will modify the plot.**