

1 Reviewer 1

1.1 Reviewer's comments

General Comments

This paper presents use of Multiple Linear Regression (MLR) and a Machine Learning (ML) algorithm as post-processing methods to improve MJO forecast of the European Centre for ECMWF model. It is generally well written and showcases successful results to improve MJO forecasts. Still, manuscript needs improvement in describing technical implementation that is somewhat short and confusing in places, such as relation between input and output neurons and lead time, as well as on MLR implementation.

Specific Comments

Line 105: "After selecting the number of output neurons (which is even and in fact defines our lead time, $\tau = Nh/2$)" – shouldn't be N_{out} instead of N_h ?
Line 110: It appears to me that for each lead time L ($1 \leq L \leq 46$), ML takes as input the predicted ECMF trajectory RMM1,2 up to day L , and as output RMM1,2 in ERA5 observations up to day $L+3$ – please elaborate and clarify by confirming or correcting as necessary. Also, what is done when $L=44,45$ and 46 ? Section 2.5: implementation of MLR is barely described at all, please expand, such as do you use regularization to avoid overfitting, etc... Line 115: Please explain what a "walk-forward validation" is.

1.2 Authors' response

We thank the reviewer for a careful revision of our manuscript that has allowed us to improve our work. With respect to the comments of the reviewer:

1. Line 105: "After selecting the number of output neurons (which is even and in fact defines our lead time, $\tau = Nh/2$)" – shouldn't be N_{out} instead of N_h ?

Authors' response: We thank the reviewer for noticing this typo, Yes indeed, $\tau = N_{out}/2$

2. Line 110: It appears to me that for each lead time L ($1 \leq L \leq 46$), ML takes as input the predicted ECMF trajectory RMM1,2 up to day L , and as output RMM1,2 in ERA5 observations up to day $L+3$ – please elaborate and clarify by confirming or correcting as necessary. Also, what is done when $L=44,45$ and 46 ?

Authors' response: We have revised the manuscript to clarify this point. The number of inputs is generally larger than the number of outputs, since we use the information we have of future ECMWF-predicted RMMs, as input. So, for each lead time L ($1 \leq L \leq 46$), we will have $(L+3)*2$ inputs, or, $N_{in} = N_{out} + 6$ (line 110). if L is 44-46, N_{in} will be equal to N_{out} , since we don't have access to the future values (line 110: $N_{in} = N_{out} + 6$ with an upper limit of 92 inputs). For lead times longer than 30 - 35 days, the prediction skill becomes poor (COR and RMSE already crossed the 0.5 and 1.4 thresholds), and thus, the last lead times (44-46 days) are not crucial.

3. Section 2.5: implementation of MLR is barely described at all, please expand, such as do you use regularization to avoid overfitting, etc...

Authors' response: We have revised the manuscript to clarify this point. We do not include a regularization term like in Ridge or Lasso. MLR is the ordinary least squares (OLS) linear regression. This choice is also due to consistency with Kim et al. 2021.

4. Line 115: Please explain what a “walk-forward validation” is.

Authors' response: We have revised the manuscript to clarify what “walk-forward validation” is. The procedure is as follows. First, we train the network on an expanding train set, and then test its performance on a validation set that contains the N samples that follow the train set. In our case, we found the best minimum number of samples for the train set, out of 2200 available, to be 1700. Then, the train set is extended by 100 samples (~ 1 year) for each run, and validated on the subsequent 200 samples (~ 2 years). This method of walk-forward validation ensures that no information coming from the future of the test set is used to train the model.

2 Reviewer 2

2.1 Reviewer's comments

Report on "Improving the prediction of the Madden-Julian Oscillation of the ECMWF model by post-processing" by Riccardo Silini et al Silini et al show that post-processing of forecasts of an ECMWF MJO prediction can improve its prediction, both in amplitude and phase. Besides this practical application their work addresses the interesting question if it is advisable to use a simple forecast model in conjunction with a deep neural network (which is expensive to train) as done in Kim et al 2021, or the most advanced forecast model (i.e. ECMWF model in the context of MJO) in conjunction with a shallow single layer neural network. The authors suggest that their approach of combining the advanced forecast model with a single single layer network is preferable in the context of MJO prediction.

The authors train on (past and future) observations, in particular on two leading EOFs. The authors also compare simple multiple linear regression and a single neural network as post-processing methodologies.

The questions and results the authors consider are definitely of a wide interest and the comparison with KIM et al (2021) is illuminating. I would, however, like to see a few clarifications:

the authors find that their ML post-processing improves MJO propagation mostly across the MC. Given the different performance across the different phases/sectors, would a stratification of the training data with respect to the initial conditions and their respective sectors be a good idea? the supplementary material Silini 2021b shows Wheeler-Hendon phase diagrams for all the data. I am not an expert on MJO but I was surprised how bad their forecasts perform. What is missing to get a better prediction? I imagine that RMM1 and RMM2 are not sufficient to predict the error between the forecast and the observations. Does one get better results if more EOFs are taken as input training data? Or can one include other variables as input training data? a minor comment: it may help the reader to begin the Discussion section with an introductory sentence which relates to their initial question of forecast model/postprocessing method complexity rather than starting without motivation "It is interesting to compare ...".

2.2 Authors' response

We thank the reviewer for his / hers comments which allow us to improve our work, in particular, the discussion. The corresponding author (RS) has recently defended his doctoral thesis, and the reviewer's pertinent and relevant comments overlap in part with those of the expert PhD committee.

1. The reviewer wonders "would a stratification of the training data with respect to the initial conditions and their respective sectors be a good idea?" Intuitively this is a great idea, but technically it is not possible because of the lack of data to train the networks. We work with the ECMWF predictions,

which are available every two weeks, for 20 years. If we split the dataset into 8 (phases), we wouldn't have enough data to perform a reliable training of the network. In fact, we can see already from our training that we found the minimum number of samples needed for the training to be 1700 out of the 2200 available. We will include a comment about this point in the revised manuscript.

2. The forecast on the Wheeler-Hendon phase diagram using machine learning techniques as standalone are not as good as those obtained with state-of-the-art dynamical models. The reviewer wonders "What is missing to get a better prediction? "

We believe that it might be possible to obtain better results with more EOFs, it probably depends on the importance of the associated variance. Since we were proposing a different approach to predict the MJO, we had to be able to fairly compare our results with other models, and the best models use RMM1 and RMM2 only. For what concerns including other variables as input data, it is a very promising idea, in fact RS and CM developed in 2021 (Silini, & Masoller, *Sci. Rep.* 11, 8423) a fast and effective metric to compute information transfer among variables, which would allow to identify variables that can be used as informative inputs. This is the object of future work and in the revised manuscript we will include a comment about this point.

3. We agree with the reviewer's suggestion, and we will modify the revised manuscript accordingly.