



EGUsphere, author comment AC1  
<https://doi.org/10.5194/egusphere-2022-2-AC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC1

Riccardo Silini et al.

---

Author comment on "Improving the prediction of the Madden–Julian Oscillation of the ECMWF model by post-processing" by Riccardo Silini et al., EGU sphere,  
<https://doi.org/10.5194/egusphere-2022-2-AC1>, 2022

---

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:

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

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

- Line 110: It appears to me that for each lead time  $L$  ( $1 < L < 46$ ), ML takes as input the predicted ECMF trajectory RMM<sub>1,2</sub> up to day  $L$ , and as output RMM<sub>1,2</sub> 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  $L$  ( $1 < L < 46$ ), 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.

- 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.

- 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.