Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-152-AC2, 2020

Geoscientific Model Development Discussions

Interactive comment on "Testing the Reliability of Interpretable Neural Networks in Geoscience Using the Madden-Julian Oscillation" *by* Benjamin A. Toms et al.

Benjamin A. Toms et al.

ben.toms@colostate.edu

Received and published: 21 December 2020

Authors' response to anonymous referee #2; GMD 2020-152

We appreciate the thoughtful comments from the second anonymous referee. Our responses to each comment are provided below.

Comment 1: For a discussion on the current understanding of MJO theory, and the historical evolution of that understanding, the authors may wish to refer to the recently published manuscript of Jiang et al. (2020): http://dx.doi.org/full/10.1029/ 2019JD030911.

Response: We have added this additional reference to the introduction.

Comment 2: Many in the target audience for this manuscript (climate scientists studying the MJO) will not be familiar with neural network techniques. Adding some background information, or references for further information, to section 2.2 would help the community to understand and accept these techniques. In particular, it would help to understand what "hidden layers", the "ReLu activation function" and the "softmax operator" are. These are probably commonly used terms in computational science, but I believe the authors would agree that they want to avoid their audience treating this technique as a black box. A few sentences of explanation or a few references with further information would help guard against this.

Response: We have added a few citations to the end of Section 2.2 that can guide the reader to publications and books with more extensive details on the methodological details of neural networks. The focus of our paper is on a scientific application of neural networks, so we leave the reader to the extensive amount of free educational material available on the internet to learn more about neural networks.

Comment 3: In Figure 4, the authors show the probabilistic performance as a 2D histogram of predicted phase against target phase. A similar figure for the deterministic performance would be useful, to demonstrate whether the neural network technique performs similarly well for all target phases of the MJO. It is not easy to determine this from Fig. 4a, as the reader has to estimate the density of dots on the phase diagram.

Response: We now list this information in the text in the first paragraph of Section 3.1.

Comment 4: Further to the above, from Fig. 4a it seems that the neural network performs better for stronger MJO events, as there seem to be more red dots closer to the unit circle and more blue and grey dots further away from the unit circle. Did the authors examine performance as a function of target MJO amplitude?

Response: We did not explicitly evaluate the accuracy of the neural network as a function of MJO amplitude. The review is correct in that the neural network is more accurate for higher amplitude cases, which is likely related to the MJO signal being

more prominent compared to non-MJO signals in these cases.

Comment 5: In Figure 5, the authors show the seasonality of the deterministic performance of the neural network technique, but provide little interpretation of the seasonality of performance. Can we learn anything – either about the MJO or about the neural network technique – from the fact that the neural networks are less successful at predicting MJO phase in boreal summer than in boreal winter? Can these results help to support the authors' conclusions about the seasonality of the MJO itself?

Response: The reduced accuracy during the summertime months is possibly related to the MJO comprising a smaller percentage of the total OLR variability during these months (e.g. Kiladis et al., 2014). Because there is more non-MJO signal, the MJO signal is muddled and therefore more difficult to identify. It is also possible that the MJO exhibits even more spatial nonlinearity during boreal summer, and that our chosen neural network architecture would therefore need more nonlinearity in order to identify the summer and winter modes with similar accuracy. This is an interesting topic for future study, and we have added a statement about this possibilities to the text of the last paragraph in Section 3.1.

Comment 6: Related to the above, are there similar seasonalities in the probabilistic performance of the neural network technique? If so, is there any useful information we can gain from interpreting those seasonalities?

Response: Yes, this is another good point. There are indeed seasonalities in the probabilistic performance. The probability distribution is more tightly clustered about the correct phase for boreal winter and more disperse for boreal summer. This is also reflected in the accuracies for boreal summer being lower than boreal winter. We don't think there is much meaningful insight to be had here except for the fact that the neural network is more uncertain and thus has lower accuracy during boreal summer. There may be interesting physical explanations for the greater uncertainty/reduced accuracy during boreal summer, although such an analysis would extend the scope of the paper

beyond its current core focus of proving base-line applicability of interpretable neural networks to geoscientific studies.

Comment 7: In Figure 7, the authors compare classical composite diagrams of OLR anomalies by MJO phase (panels (a) and (b)) against the "interpreted" results from the neural network that highlight the most salient features for identifying the MJO phase. The authors' interpretation is that the neural network identifies a more focused area of active and suppressed convection as relevant for the MJO, versus the more widespread or diffuse anomalies in the classical composites. The common approach in composite analysis is to show only those anomalies that are statistically significant at some threshold (e.g., 5% significance) based on a t test or similar. Did the authors perform such a test on panels (a) and (b)? If not, I would recommend performing one, as it might result in a more "focused" composite anomaly.

Response: Layerwise relevance propagation itself does not take into account significance, so we did not complete any significance testing on panels (a) and (b). A method for testing the significance of LRP heatmaps and optimal input fields is being developed separately, and will be usable in subsequent manuscripts. For this reason, we do not feel it is justified to filter the regression maps shown in subpanels (a) and (b) for significance. We agree that removing statistically insignificant regions from figures (a), (b), (c), and (d) may further limit the expanse of both the regression-based (panels a and b) and neural network-based (panels c and d) interpretations of the MJO.

Comment 8: The results presented in this manuscript are certainly a useful first step toward using neural network techniques for understanding and predicting the MJO. However, the greatest uncertainty in community understanding of the MJO is not the identification of MJO phase or seasonality, but the mechanisms for MJO genesis, intensification and propagation. For instance, why do some MJO events propagate across the Maritime Continent while others do not? Why are some MJO events stronger than others? The authors hint that their neural network techniques might be useful for addressing these challenges (L315), but I believe a more detailed discussion of this po-

tential would help the community to see the value in these techniques for understanding and predicting the MJO. As I am not an expert in neural network techniques, I cannot see a straightforward way to apply these techniques to understanding the propagation of the MJO or the mechanisms that drive that propagation. Can the authors add to this discussion in a revised manuscript?

Response: We have added a few lines of discussion on how a similar approach to the one used in this manuscript could be used for these specific hypotheses.

Comment 9: Throughout section 3.2.2, the authors discuss the atmospheric fields that are most “relevant” to the MJO. Perhaps this word has a precise definition in neural network analysis, but I struggled with the interpretation here. What does “relevant” mean? Does it mean that the atmospheric field controls MJO strength, or determine MJO phase? Is a “relevant” field simply a field that has a structure common to most MJO events in that phase, regardless of intensity?

Response: For clarity, we have changed the phrase “relevant” to “important for the identification of the MJO”, or something similar to that for all cases. The phrase “relevant” does not have specific meaning in the computer science community, aside from the concept that LRP identified aspects of the input are most relevant to the network’s associated output.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-152, 2020.