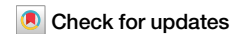




Incorrect computation of Madden-Julian oscillation prediction skill

Tamaki Suematsu, Zane K. Martin, Elizabeth A. Barnes, Charlotte A. DeMott, Samson Hagos, Yoo-Geun Ham, Daehyun Kim, Hyemi Kim, Tieh-Yong Koh & Eric D. Maloney



The Madden–Julian oscillation (MJO) is a major tropical weather system and one of the largest sources of predictability for subseasonal-to-seasonal weather forecasts. Skillful prediction of the MJO has been a highly active area of research due to its large socio-economic impacts. Silini et al., herein S21, developed a machine learning model to predict the MJO, which they claimed to have an MJO prediction skill of 26–27 days over all seasons and 45 days for December–February (DJF) winter. If true, this would make the skill of their model competitive with that of the state-of-the-art dynamical MJO prediction systems at 20–35 days. However, here we show that the MJO prediction was calculated incorrectly in S21, which spuriously increased the performance of their model. Correctly computed skill of their model was substantially lower than that reported in S21; the skill for all seasons drops to 11–12 days and the skill for forecasts initialized during DJF drops to 15 days. Our findings clarify that the S21 machine learning model is not competitive with state-of-the-art numerical weather prediction models in predicting the MJO.

Prediction of the Madden–Julian oscillation (MJO) has been an active area of research since the 1990s^{1,2} because of the recognized importance of the MJO as a source of predictability for subseasonal-to-seasonal weather forecasts. There have been efforts to improve the forecasts of the MJO and standardize its assessments³, including those of the WGNM MJO Task Force^{4,5}. These coordinating bodies and studies have developed frameworks and metrics for fair comparison of model performances regarding MJO simulation across different models. Recent studies using these comparison frameworks have documented impressive improvement in MJO prediction^{1,3}, largely due to improved dynamical models optimized for MJO prediction⁵. The best dynamical

models can skillfully predict the MJO beyond 30 days¹, with most showing 20–25 days of prediction skill¹.

Silini et al. (2021; hereafter S21)⁶ introduced a set of machine learning models to predict the Real-time Multivariate MJO index (RMM)⁷, which is an index commonly used to identify the MJO. The RMM represents the state of an MJO with RMM1 and RMM2, that correspond to the leading pair of empirical orthogonal functions of the outgoing longwave radiation, and zonal winds at 850 and 200 hPa. S21 trained different types of machine learning models to forecast the RMM indices using past values as inputs and outputs. The details are described in S21.

A key metric for assessing the prediction skill of MJO forecasts that has widely been used in the weather and climate community is the bivariate correlation coefficient (COR)^{1,4}. COR is calculated via Eq. (1) (identical to Eq. (1) in S21):

$$\text{COR}(\tau) = \frac{\sum_{t=1}^{t=N} [a_1(t)b_1(t, \tau) + a_2(t)b_2(t, \tau)]}{\sqrt{\sum_{t=1}^{t=N} [a_1^2(t) + a_2^2(t)]} \sqrt{\sum_{t=1}^{t=N} [b_1^2(t, \tau) + b_2^2(t, \tau)]}} \quad (1)$$

Here N is the total number of forecast cases, and τ is the lead time of the forecast. $a_{1,2}$ are the observed values of RMM1 and RMM2 at time t , respectively. $b_{1,2}$ are the corresponding model predicted values of RMM1 and RMM2, respectively, at a lead time τ .

While S21 states that Eq. (1) is employed to calculate the prediction skill of their model, in attempting to reproduce their results, we found that this was not the case. Instead, it was found that S21 erroneously used the following incorrect equation to compute the prediction skill:

$$\text{incorrect COR}(\tau) = \frac{\sum_{i=0}^{i=\tau} \sum_{t_0=0}^{t_0=N} [a_1(t_0+i)b_1(t_0, i) + a_2(t_0+i)b_2(t_0, i)]}{\sqrt{\sum_{i=0}^{i=\tau} \sum_{t_0=0}^{t_0=N} [a_1^2(t_0+i) + a_2^2(t_0+i)]} \sqrt{\sum_{i=0}^{i=\tau} \sum_{t_0=0}^{t_0=N} [b_1^2(t_0, i) + b_2^2(t_0, i)]}} \quad (2)$$

Note also that here, t_0 denotes the initial date of the forecast. In Eq. (2), there is an extra summation over τ compared to Eq. (1). S21 erroneously computed the MJO prediction skill of their model using the covariance of all lead times from zero to τ days, rather than computing the model skill at the lead time of τ days. What S21 claim to be the model skill at a lead time of 7 days, for example, incorporates information of the forecast skills at all lead times from 0 to 7 days. In the correct formulation given in Eq. (1), $b_{1,2}(t, \tau)$ represent *only* the model values of RMM1 and RMM2 at a lead time τ .

Personal communication with the authors of S21 confirmed that this error was unintentionally made during the calculation of COR. We emphasize that Eq. (2) is not an alternative metric employed by preceding studies regarding MJO predictions. Although Eq. (2) is the actual equation

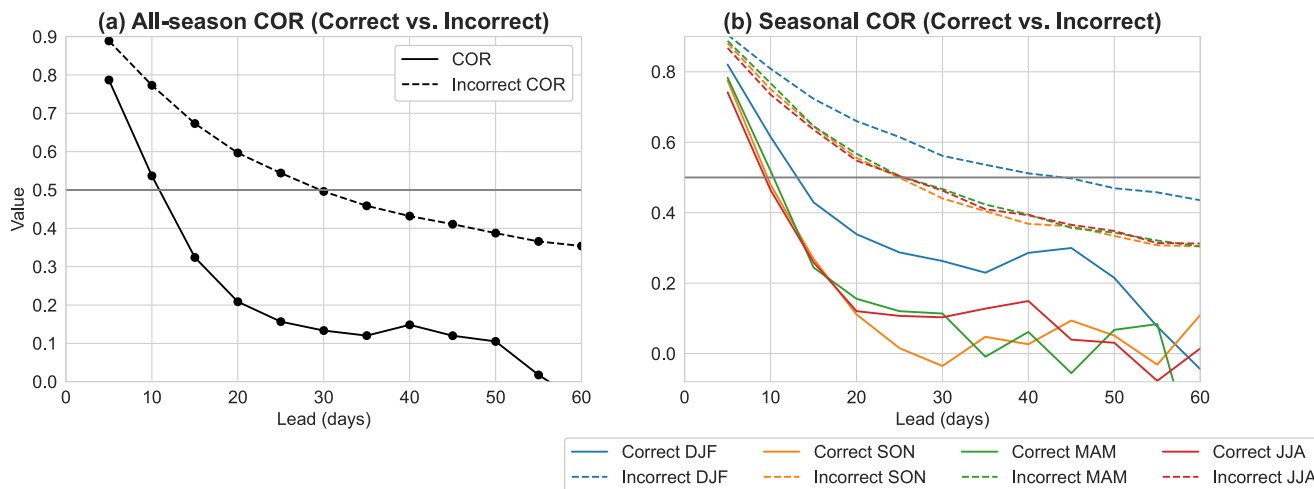


Fig. 1 | Comparison of correctly and incorrectly calculated COR. The COR values calculated correctly (solid lines) per Eq. (1), and incorrectly (dashed) as in S21, following Eq. (2), for all season (a) and for each season (b). Note that, compared to

Figs. 1 and 3 in S21, the incorrect results agree very well with the published results of S21, while the correct results show substantially lower skill in both panels.

employed to evaluate the MJO prediction skill in S21, S21 states inaccurately that Eq. (1) is used. Thus, statements regarding the prediction skill in S21 are neither correct nor comparable with any previous MJO prediction skill assessment studies. Since COR tends to decrease with time, the extra summation over the period between zero and τ in Eq. (2) tends to make the value of the incorrect COR greater than that of the correct COR.

To quantify the impact of this error on the results in S21, we compare the prediction skill of the S21 model evaluated by the correct COR via Eq. (1) with that evaluated by the incorrect COR via Eq. (2) (Fig. 1). Here, the results are shown only for the feed-forward neural network model in S21 (and not for the autoregressive recurrent neural network). However, given the nature of the error, we are confident that findings here hold for both types of models demonstrated in S21.

In Fig. 1, the incorrect skill calculated via Eq. (2) agrees almost identically with the skill published in S21 (see Figs. 1 and 3 in Silini et al.⁶). However, the correct skill computed via Eq. (1) indicates that the actual prediction skill of S21 model is much less than what was reported. The skill for all season drops from the reported 30 days to 11–12 days, and prediction skill of simulations initiated during December–February (DJF) drops from the reported 45 days to 15 days. Similar reduction in prediction skill is seen for simulations initiated during March–May (MAM), June–July (JJA), and September–November (SON), as well. Reproduction of other Figures (e.g., S21 Fig. 5) demonstrate similarly large reductions of model prediction skill (not shown). Considering that the MJO prediction skills of the best performing dynamical models are currently at 4–5 weeks, the reduction of prediction skill by 1–2 weeks is substantial.

The discussion in S21 regarding the relative merits of their model is invalid because skill was computed incorrectly. It turned out that the machine learning framework for MJO prediction in S21 is not competitive with the state-of-the-art dynamical models. Their models show skill comparable to classical linear models of the MJO^{1,8}, at approximately 1–2 weeks, which use simple linear regression techniques to forecast the RMM. This implies that current state-of-the-art dynamical models are still the optimal method for MJO prediction.

We also note that there have been other machine learning models that demonstrated performances exceeding that of S21^{9–11}. This suggests that

machine learning framework of S21, which employs only RMM as the input of the machine learning, may be less advantageous than existing machine learning approaches that use additional input variables. A more detailed comparison would be needed to fully evaluate this claim. Evaluation of S21 model using the correct equation highlights that their models do not outperform other existing machine learning approaches for MJO prediction.

We do not want to discourage or dissuade continued research on prediction of the MJO using machine learning. However, carefully adhering to the best practices established by the MJO community and correctly employing standardized metrics to evaluate model performances are necessary for fair comparison across different models. This is especially true given the important implications that MJO prediction has on subseasonal-to-seasonal forecasting in both scientific and applied domains.

Preceding studies have suggested that inherent MJO predictability is up to 6–7 weeks^{12,13}. The initial claim of S21 that MJO prediction skill exceeds 60 days in certain phases or seasons seemed especially difficult to reconcile with this predictability estimates. The correction of the equation to evaluate their model indeed reduced the model performance down below the reported predictability estimates. However, both dynamical and machine learning approaches for MJO prediction should continue to endeavor to meet these predictability limits. Future research in this area will continue to illuminate whether any modeling framework might reach this goal.

Data availability

RMM data used for this study is available from the Australian Government Bureau of Meteorology: <http://www.bom.gov.au/climate/mjo/>.

Code availability

Keras TensorFlow trained feed-forward neural network of S21 can be found at repository indicated in S21¹⁴. Codes to reproduce the results of S21 and the correct computation of COR is available at <https://doi.org/10.5281/zenodo.10930830>.

Tamaki Suematsu¹ ✉, Zane K. Martin², Elizabeth A. Barnes², Charlotte A. DeMott², Samson Hagos³, Yoo-Geun Ham⁴, Daehyun Kim⁵, Hyemi Kim^{6,7}, Tieh-Yong Koh⁸ & Eric D. Maloney²

¹RIKEN Center for Computational Science, Kobe, Japan. ²Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA. ³Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA. ⁴Department of Environmental Planning, Graduate School of Environmental Studies, Seoul National University, Seoul, Republic of Korea. ⁵School of Earth and Environmental Sciences, Seoul National University, Seoul, Republic of Korea. ⁶Department of Science Education, Ewha Womans University, Seoul, Republic of Korea. ⁷School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA. ⁸Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. ✉e-mail: tamaki.suematsu@riken.jp

Received: 8 January 2024; Accepted: 9 June 2024;
Published online: 17 June 2024

References

- Kim, H., Vitart, F. & Waliser, D. E. Prediction of the Madden–Julian oscillation: a review. *J. Clim.* **31**, 9425–9443 (2018).
- Waliser, D. in *Intraseasonal Variability in the Atmosphere–Ocean Climate System* (eds Lau, W. K. M. & Waliser D. E.) 389–423 (Springer, 2005).
- Gottschalck, J. et al. A framework for assessing operational Madden–Julian oscillation forecasts: a CLIVAR MJO Working Group Project. *Bull. Am. Meteorol. Soc.* **91**, 1247–1258 (2010).
- Vitart, F. et al. The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bull. Am. Meteorol. Soc.* **98**, 163–173 (2017).
- WGNE. MJO Task Force. <https://wgne.net/activities/on-going-activities/wgne-mjo-task-force/> (2024).
- Silini, R., Barreiro, M. & Masoller, C. Machine learning prediction of the Madden–Julian oscillation. *npj Clim. Atmos. Sci.* **4**, 1–7 (2021).
- Wheeler, M. C. & Hendon, H. H. An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Mon. Weather Rev.* **132**, 1917–1932 (2004).
- Kang, I. S. & Kim, H. M. Assessment of MJO predictability for boreal winter with various statistical and dynamical models. *J. Clim.* **23**, 2368–2378 (2010).
- Hagos, S., Leung, L. R., Zhang, C. & Balaguru, K. An observationally trained Markov model for MJO propagation. *Geophys. Res. Lett.* **49**, e2021GL095663 (2022).
- Martin, Z. K., Barnes, E. A. & Maloney, E. Using simple, explainable neural networks to predict the Madden–Julian oscillation. *J. Adv. Model. Earth Syst.* **14**, e2021MS002774 (2022).
- Shin, N. Y., Kim, D., Kang, D., Kim, H. & Kug, J. S. Deep learning reveals moisture as the primary predictability source of MJO. *npj Clim. Atmos. Sci.* **7**, 1–8 (2024).
- Neena, J. M., Lee, J. Y., Waliser, D., Wang, B. & Jiang, X. Predictability of the Madden–Julian oscillation in the intraseasonal variability hindcast experiment (ISVHE). *J. Clim.* **27**, 4531–4543 (2014).
- Mengist, C. K., Seo, K. H., Ding, R. & Son, S. W. Potential predictability of the MJO during easterly and westerly phases of the QBO. *Clim. Dyn.* **57**, 717–726 (2021).
- GitHub. [riccardosilini/MJO-prediction-networks](https://github.com/riccardosilini/MJO-prediction-networks). <https://github.com/riccardosilini/MJO-prediction-networks> (2024).

Acknowledgements

All members of the WGNE MJO Task Force are appreciated for discussion of these results and formulating this response. T.S. was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI grant 21K13991. Z.K.M. was supported by National Science Foundation under Award No. 2020305. This work was also supported, in part, by NOAA grant NA19OAR4590151, NSF Grant AGS-1841754, NOAA Grant NA22OAR4590168, and NOAA Grant NA22OAR4310608. D.K. was supported by New Faculty Startup Fund from Seoul National University. H.K. was supported by the National Research Foundation of Korea (NRF) grant (NRF-RS-2023-00278113).

Author contributions

Z.K.M., E.A.B., and E.D.M. initiated the work through discussion regarding S21's initial findings. Z.K.M. conducted the analysis, drafted the initial document. Z.K.M. and T.S. coordinated editing. All authors contributed to discussion of results and drafting the final document.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Tamaki Suematsu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024, corrected publication 2024