Geophysical Research Letters

RESEARCH LETTER
10.1029/2021GL095663

Special Section:
Years of the Maritime Continent

An Observationally Trained Markov Model for MJO Propagation

Samson Hagos[1], L. Ruby Leung[1], Chidong Zhang[2], and Karthik Balaguru[1]

[1]Pacific Northwest National Laboratory, Richland, WA, USA, [2]NOAA Pacific Marine Environmental Laboratory, Seattle, WA, USA

Key Points:

- A stochastic Markovian model of Madden-Julian Oscillation (MJO) propagation is developed using 40 years of observations using Machine learning
- The model is used to examine the disruption of MJO propagation across the Maritime Continent region
- The simplicity of the model allows the generation of large ensembles for forecast skill and predictability studies

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Hagos,
samson.hagos@pnnl.gov

Citation:

Hagos, S., Leung, L. R., Zhang, C., & Balaguru, K. (2022). An observationally trained Markov model for MJO propagation. *Geophysical Research Letters*, *49*, e2021GL095663. https://doi.org/10.1029/2021GL095663

Received 16 AUG 2021
Accepted 8 DEC 2021

Author Contributions:

Conceptualization: Samson Hagos, Chidong Zhang
Formal analysis: Samson Hagos
Funding acquisition: Chidong Zhang
Investigation: Samson Hagos
Methodology: Samson Hagos, Chidong Zhang, Karthik Balaguru
Project Administration: Chidong Zhang
Software: Samson Hagos, Karthik Balaguru
Supervision: Chidong Zhang

Abstract A Markovian stochastic model is developed for studying the propagation of the Madden-Julian Oscillation (MJO). This model represents the daily changes in real time multivariate MJO (RMM) indices as random functions of their current state and background conditions. The probability distribution function of the RMM changes is obtained using a machine learning algorithm trained to maximize MJO forecast skills using observed daily indices of RMM and different modes of variability. Skillful forecasts are obtained for lead times between 8 and 27 days. Large ensemble simulations by the stochastic model show that with monsoonal changes in the background state, MJO propagation across the Maritime Continent (MC) is most likely to be disrupted in boreal spring and summer when MJO events propagate from favorable conditions over the Indian Ocean to unfavorable ones over the MC, and predictability is higher during spring and summer when MJO activity is away from the MC region.

Plain Language Summary The work demonstrates the application of machine learning in the development of reduced dimension stochastic models that can efficiently run and analyzed. As a simple form of such models a Markov model of Madden-Julian Oscillation (MJO) propagation is presented. The model is trained to maximize MJO forecast skill using 40 years of observations of MJO and the background, seasonal, El-Niño Southern Oscillation, Quasi-Biennial Oscillation and Indian Ocean Dipole state. The application of the model to the problem of MJO disruption over the maritime continent region and in the study of predictability using analysis of signal-to-noise ratio is discussed. The work highlights that, in addition to direct analysis of observations and numerical simulations, observationally trained reduced dimension models could be valuable tools of research in climate variability and multi-scale interactions.

1. Introduction

The Madden-Julian Oscillation (MJO) is well recognized as a source of potential predictability on subseasonal to seasonal timescales. This has motivated broad interests in understanding the MJO, particularly the mechanisms of its propagation, its variability and predictability. Several studies have examined how known modes of climate variability such as seasonality, El-Niño Southern Oscillation (ENSO), Quasi-Biennial Oscillation (QBO) and the Indian Ocean Dipole (IOD) affect the propagation of the MJO. In general, about half of MJO events weaken as they cross the Indo-Pacific Maritime Continent (MC) region and MJO events are twice more likely to weaken during El Niño years compared to La Niña years (Burleyson et al., 2018; Kerns & Chen, 2016). Recently, Hagos et al. (2019) showed that this variability of MJO propagation and strength is related to a slow eastward migration of zonal moisture flux convergence between the Asian and Australian monsoon convergence centers from boreal summer to winter and its modulation by ENSO. Strengths of individual MJO events change as they propagate across this zonally and seasonally varying monsoonal moisture convergence as well as the passage of a previous MJO event (Moum et al., 2016). The impacts of QBO on the MJO has also been a subject of recent interest (e.g., Liu et al., 2014; Marshall et al., 2017; Martin et al., 2021; Nishimoto & Yoden, 2017; Son et al., 2017; Yoo & Son, 2016; Zhang & Zhang, 2018). MJO activities in boreal winter are found to be stronger in easterly than westerly phases of QBO. The influence of the Indian Ocean Dipole (IOD) on the development and propagation was examined by Wilson et al. (2013) who showed that MJO events over the Indian Ocean and MC region are stronger during negative phases of IOD (warm surface in the western tropical Indian Ocean) and weaker during positive phases (warm surface in the eastern tropical Indian Ocean). The authors attributed this sensitivity to the modulation of local low-level moisture over the eastern Indian Ocean by sea surface temperature (SST).

Validation: Samson Hagos, Karthik Balaguru
Writing – review & editing: Chidong Zhang, Karthik Balaguru

These significant progress in understanding the influences of various modes of climate variability on the MJO have been made using composite analyses. However, composite analyses allow direct examinations of neither the impact of these modes of variability on individual MJO events nor the potential role of non-linear processes in the combined effects of multiple modes. In a complementary approach using physics-based models of various degrees of complexity, environmental conditions such as SSTs, upper-level wind, or insolation etc. (e.g., Back et al., 2020; Hagos et al., 2020) are modified to determine the impact of the phenomenon of interest on a case study basis. Such an approach assumes the model adequately represents MJO propagation and its relationship to the background state. Sufficient computational resources are needed to run many ensembles for more systematic analysis. Ensemble modeling is particularly important in predictability research where robust calculations of the signal-to-noise ratio are critical.

For robust and efficient examinations of the variability and predictability of MJO propagation, we propose a third approach through the development and application of reduced dimension stochastic models of MJO propagation trained by observations. As a demonstration of such an approach, we present a simple Markovian model trained by 40 years of daily indices of the climate state including the seasonal cycle, ENSO, QBO and IOD. The model development is described in the next section. Applications of the model to understand MJO disruption and predictability are elaborated in Section 3. Discussions are given in Section 4.

2. Model

We aim at developing a stochastic model of MJO propagation that is easy to train, run, analyze, and interpret. As a first step, a concise way of representing the MJO and the background state is sought. To this end we define a vector $\boldsymbol{S}$ whose elements include the indices of the states of the MJO and the background on a given day t. In this model the six-dimensional vector $\boldsymbol{S}$ is composed of the two real time multivariate MJO indices (RMMs), solar declination angle (SDA in radians) as well as ENSO, QBO and IOD indices. These sources and references for the indices are provided in Table S1 in the Supporting Information S1.

$$S_t = [\mathrm{RMM1}, \mathrm{RMM2}, \mathrm{SDA}, \mathrm{ENSO}, \mathrm{QBO}, \mathrm{IOD}] \tag{1}$$

A two-dimensional matrix function $\boldsymbol{P}$ is defined as the probability that changes in the state of the MJO from day t to day $t+1$ defined as ΔRMM1 and ΔRMM2 fall within one of 10 pre-defined bins drmm1 and drmm2. that is,

$$\boldsymbol{P}(\Delta\mathrm{RMM1} = \mathrm{drmm1}|S_t, \Delta\mathrm{RMM2} = \mathrm{drmm2}|\boldsymbol{S}_t) = f(\boldsymbol{S}_t) \tag{2}$$

where f is the transition function to be obtained from observations using machine learning. An important point to note is that Equation 2 represents a Markovian model, which is the simplest form in the general class of stochastic models (Gagniuc, 2017; Wilks, 1995), that is, the change in $\boldsymbol{S}_t$ is assumed to depend only on its previous state $\boldsymbol{S}_{t-1}$. The distribution of the daily changes in RMM1 and RMM2 from their 40 years of data between 1980 and 2019 is shown in Figure S1a in the Supporting Information S1. Part of the distribution that falls between −0.5 and 0.5 are partitioned into 10 bins such that each drmm1 and drmm2 bin has a width of 0.1. Values of drmm1 and drmm2 outside the range of −0.5 and 0.5 on 4% of the days are excluded to reduce the effect of small sampling on the training by requiring at least 500 days of training data for each bin. This limits the applicability of the model to this range of drmm1, drmm2. The (i,j) element of the 10×2 dimensional matrix $\boldsymbol{P}$ represents the probability that the change in ΔRMM1 and ΔRMM2 falls within the ith drmm1 bin and the jth drmm2 bin. The machine learning algorithm and the training process are described in some detail in the Supporting Information S1.

The model advances forward in time as follows. At a given time t, based on the initial and background state ($S(t)$), the optimized function $f(S)$ provides P from which the actual forecasts of (drmm1(t), drmm2(t) are randomly drawn. Then rmm1($t+1$) = rmm1(t) + drmm1(t) and rmm2($t+1$) = rmm2(t) + drmm2(t). A large independent ensemble of such optimizations is performed in parallel to limit the possibility of the training being trapped in a local minimum. To some extent the model developed here bears some resemblance to the Markov model of Jones (2009) who used a nine-state first-order Markov chain in which state 0 represents quiescent days and states 1–8 are the active phases of the MJO to estimate the transition probabilities based on the historical record of MJO events. The model proposed in the current study is more general in that it aims to predict the changes in the RMM index and therefore both amplitude and phase of the MJO are predicted using information of the variability of the states of the background as well as the MJO.

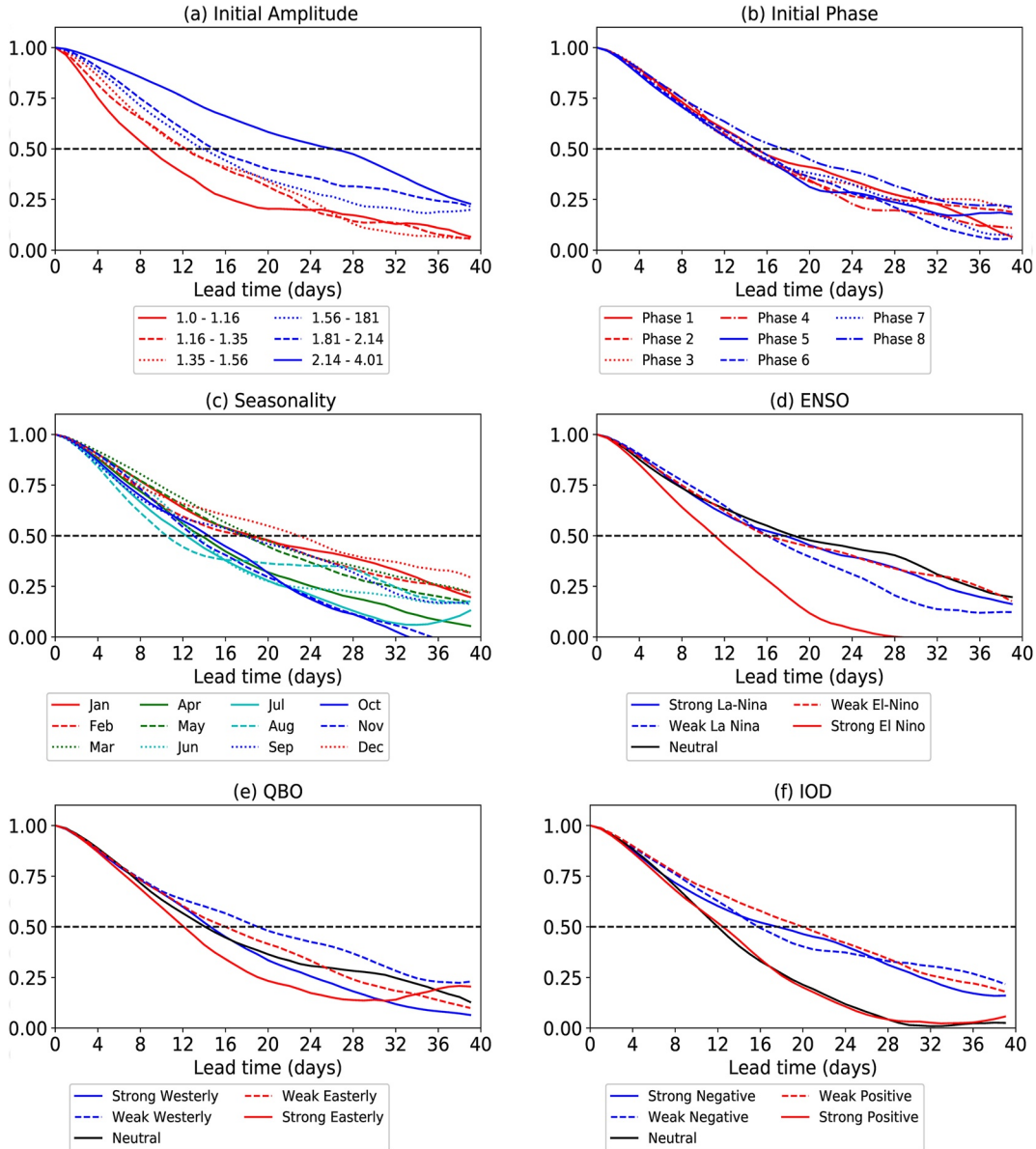

Figure 1. The BVCORR calculated from 20 ensemble member forecasts with various initial and boundary conditions. For the El-Niño Southern Oscillation, Quasi-Biennial Oscillation phases and Indian Ocean Dipole phases the five quintiles of their respective indices are used to define strong, weak, and neutral cases.

3. Results

Once the model is optimized, the variations of its forecast skill with a given initial amplitude and an initial phase as well as background conditions (i.e., seasonal cycle and the phases of the various interannual modes of variability) are examined. For each initial and background states of interest, simulations of a 20-member ensemble are performed. As in the training stage the forecasts are initialized on days when the RMM amplitude is greater than 1.0. Figure 1 summarizes the variation of the skill, defined as BVCORR, with initial amplitudes, initial phases, seasons as well as ENSO, QBO and IOD phases. The skill is found to be most sensitive to the initial MJO amplitude and to seasonality, especially during the first 10 days. Measured by BVCORR > 0.5, skillful predictions can be obtained with lead times as long as 27 days for the strongest 20% of MJO events and as short as 9 days for the weakest 20% of MJO events. Higher skills are found in boreal winter (up to 24 days in December) than in

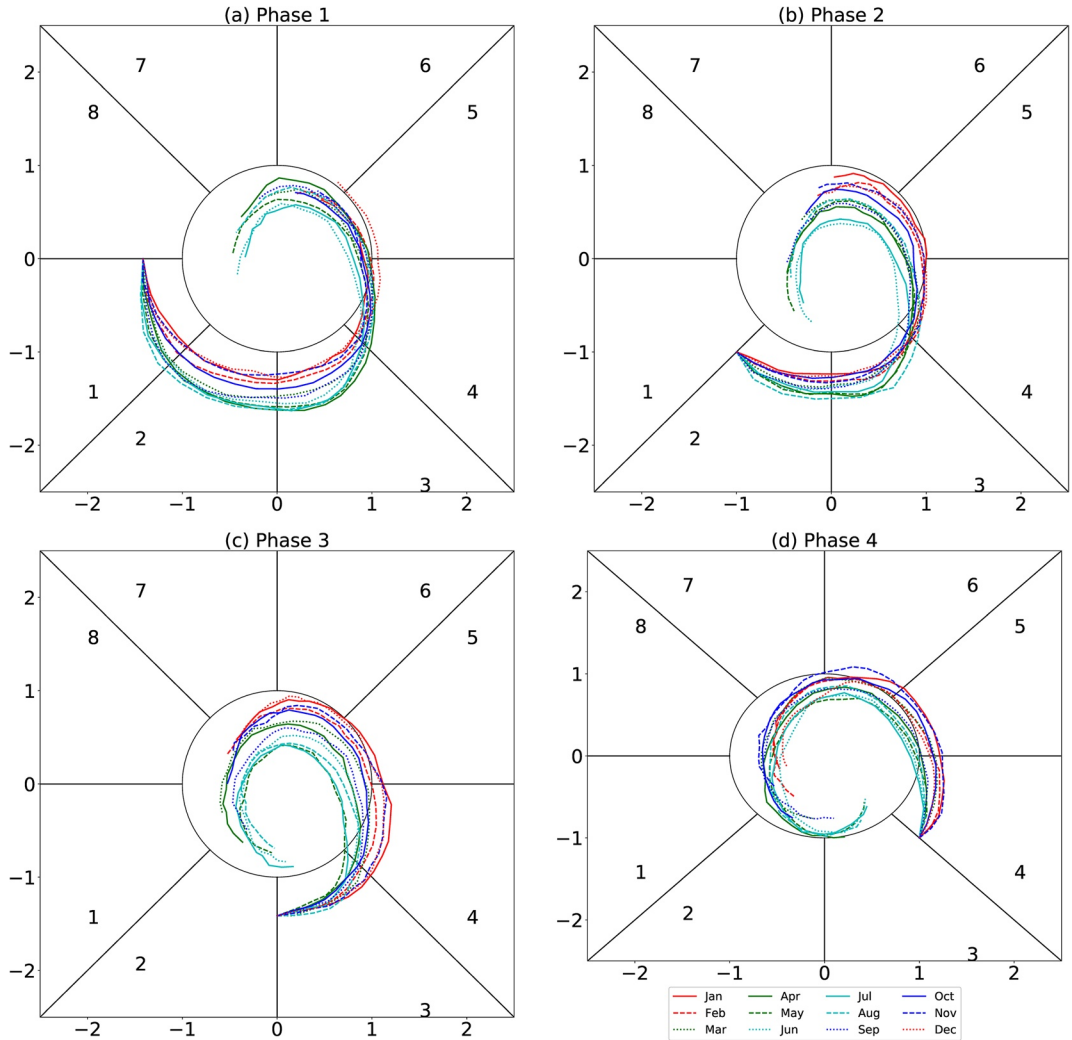

Figure 2. The averages of RMM diagrams of 100 member simulations of propagation of hypothetical Madden Julian Oscillation events initialized at (a) phase 1, (b) phase 2, (c) phase 3 and (d) phase 4 on the middle (15th day) of the month.

summer (12 days in August). As measured by the difference in lead time at which BVCORR crosses the 0.5 line, the variability in the skill associated with initial phase, ENSO, QBO and IOD is relatively smaller.

The model is used to examine the propagation characteristics of hypothetical MJO events initialized at various phases during different months. Figure 2 shows the seasonality of propagation of MJO events of the same initial amplitude of $\sqrt{2}$ but in different initial phases (phase 1 in Africa to phase 4 in MC) and initialized on different months (colors). All other indices are set to neutral. As part of the monsoonal transitions, from late summer to fall and winter, moisture convergence and associated convective activity propagate eastward from the Asian monsoon to the MC and the Australian monsoon longitudes. In contrast, from the end of boreal winter to spring and early summer, moisture convergence and convective activity propagate westward toward the Indian Ocean longitudes. In other words, in spring and summer, the Indian Ocean sector is favorable for convective activity while the Australian and MC regions are not, but in fall and winter the favorable environment shifts eastward (and southward) and the conditions are reversed. In Hagos et al. (2019) it was shown that the seasonal variability of MJO strength reflects this monsoonal variability. Specifically, during boreal spring and summer much of the MJO activities are in phases 2 and 3, that is, over the Indian Ocean, and during fall and winter they are in phases 4 and 5, that is, over the MC region. This has important implications for the role of initial conditions. For example, MJO events initialized in phases 2 and 3 (Figures 2b and 2c) in the middle of spring and summer months (green) are more likely to weaken as they would be propagating toward an unfavorable environment compared to MJO events initialized at

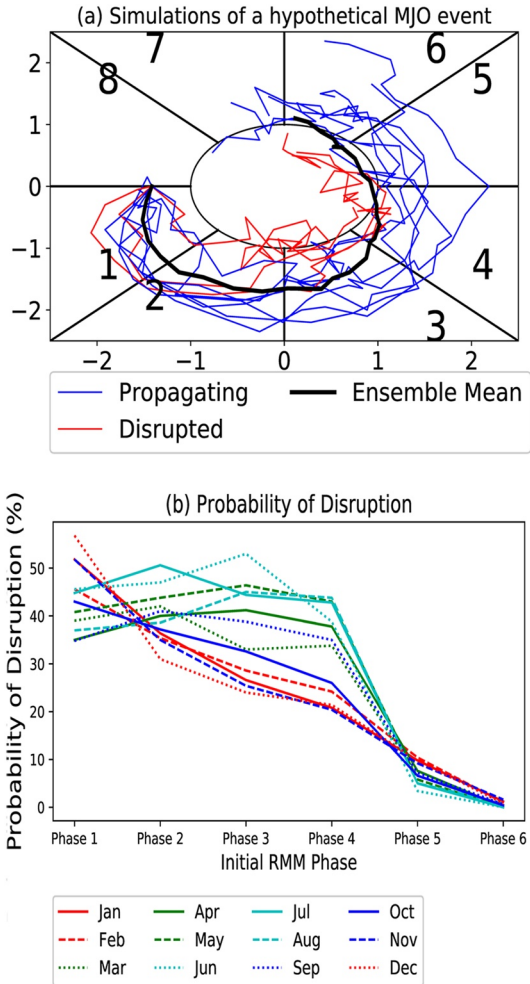

Figure 3. (a) A 10-member ensemble simulation of a hypothetical Madden Julian Oscillation (MJO) event. Events are considered disrupted if they don't an amplitude greater than 1.0 in phase six within 30 days of initialization. (b) The probability of disruption of hypothetical MJO events initiated at various phases and various months calculated from simulations of 1,000 member ensembles.

the same phase but in winter (red), as the latter events would propagate into a favorable environment. Sensitivity tests indicate this behavior shows little sensitivity to the other aspects of the climate state (the other indices).

The simplicity of the stochastic model allows many simulations to be generated for quantitatively assessing the seasonality of likelihood of disruption to MJO propagation. To that end we define propagating and disrupted events as follows. An MJO event is defined as propagating if on any day within 30 days of initialization it has an amplitude >1.0 in phase 6 (Western Pacific), otherwise it is defined as disrupted (by the MC). As a demonstration of this definition, RMM phase diagram of a 10-member ensemble simulation initialized at phase 1 is shown in Figure 3a. Three events (in red) meet the criteria for our definition of disrupted events. Similar simulations are performed with 1,000 ensemble members initialized at different seasons and phases so that the fraction of the ensemble members that are disrupted is defined as the probability of disruption. Figure 3b shows the dependence of this probability of disruption on the initial phase and the month of initialization. As can be expected from the seasonal zonal migration of convective activity between the Indian Ocean and MC region discussed above, MJO events entering the MC region in spring and summer are most likely to be disrupted as they are moving from a favorable to an unfavorable environment. Hence MJO events initialized in spring and summer (green) during phases 3 and 4 have higher probability of disruption relative to MJO events initialized in winter (red and blue) during the same phases 3 and 4 (Figure 3b). The higher chance of disruption remains the case even if the MJO events are still strong in phase 4.

Finally, the predictability of MJO amplitude is examined using large ensemble simulations by the stochastic model. Signal-to-noise ratio is defined as the ratio of the mean to the standard deviation of the ensemble for MJO amplitude. Figure 4 shows the variability of this signal-to-noise ratio with season of initialization (color), initial phase (panels a–h) and lead time (x-axis). Signal-to-noise ratio shows significant seasonality and dependence on initial phase with larger values in spring and summer. Furthermore, signal-to-noise ratio is enhanced when the active phase MJO events is away from the MC region such as within 15 days of initial phase 7 (Figure 4g), within 10 days of initial phase 8 (Figure 4h) etc.

4. Discussion

Much of our understanding of MJO propagation and variability is obtained from composite analysis, numerical model simulations and theoretical analyses. Each of these approaches has its own strengths and weaknesses. We introduce another approach that exploits the use of machine learning to build stochastic models trained to capture natural variability of the earth system using long-term observations. The first and key advantage of this approach relative to linear models is the machine learning piece in which the form of the non-linear function $\boldsymbol{F}(\boldsymbol{s})$ is determined by the observational data. The second advantage is that the stochasticity arises from the observation itself rather than imposed using arbitrary distribution. Thus, natural variability of the system is included which makes such models useful for predictability studies. Such stochastic models can produce a large number of ensemble members for robust statistics. In this work as a first version of stochastic models for demonstration, a Markovian model is developed and trained to maximize MJO forecast skill using 40 years of daily observation of indices of RMM, seasonality, ENSO, QBO and IOD. The variability of forecast skills with the initial and background states is documented. Specifically, the lead time of skillful forecasts varies between 8 and 27 days, primarily depending on the initial amplitude of the RMM index. Specifically, skillful predictions can be obtained with lead times as long as 27 days for the strongest 20% of MJO events and as short as 9 days for the weakest 20% of MJO events. The skill also shows seasonality with higher skills in boreal winter (up to 24 days in December) than in summer (12 days in August). The

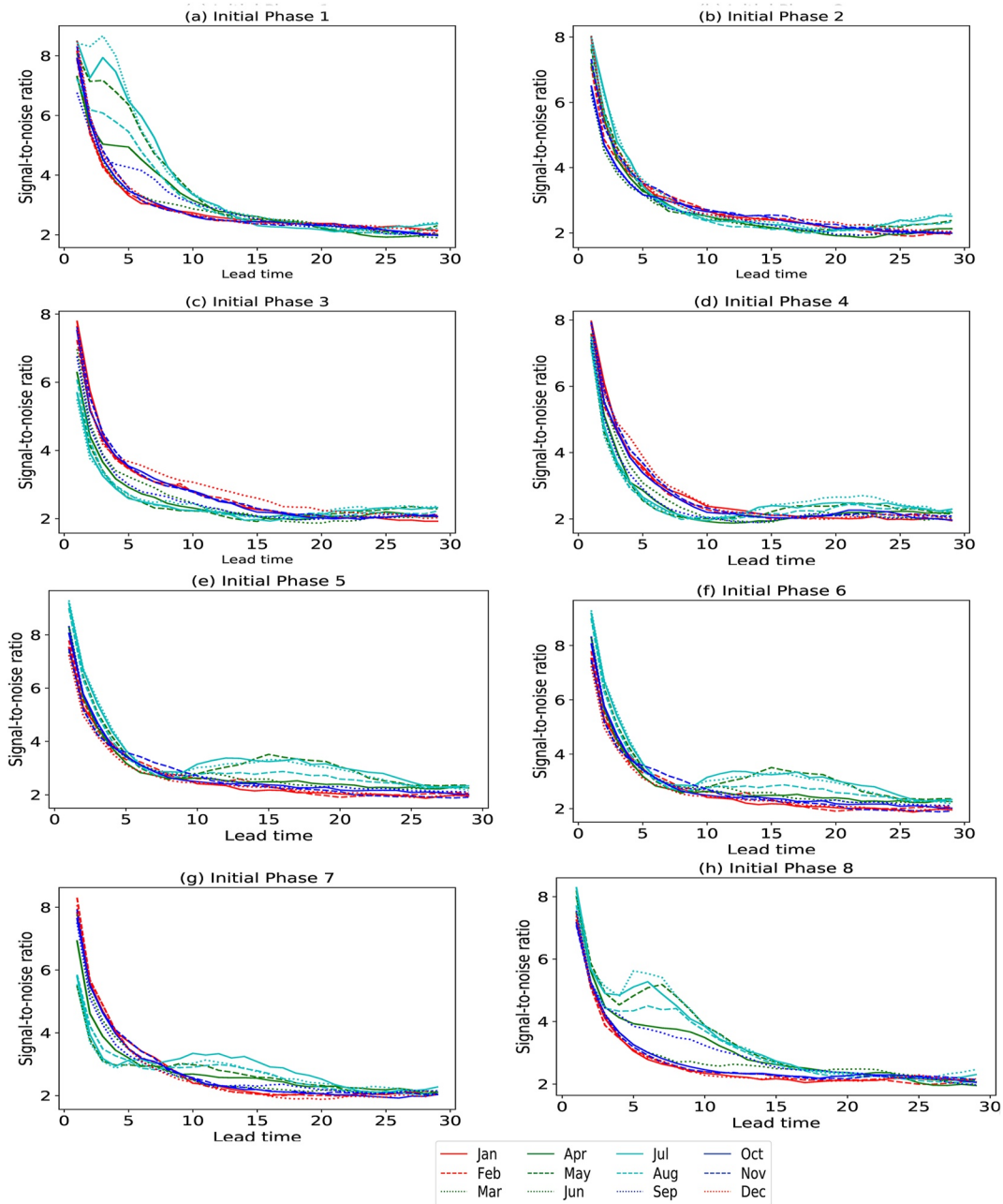

Figure 4. Signal-to-noise ratio calculated from 1,000 ensemble member simulations initialized at the eight RMM phases.

variability in the skill associated with initial phase, ENSO, QBO and IOD are relatively smaller. This is in contrast to the strong sensitivity of forecast skill of an numerical model to QBO phase reported by Lim et al. (2019).

As the model is computationally economic to run, large ensembles of simulations are used to examine the probability of MJO propagation being disrupted that may vary with the background state and initial phase. Because the background state favorable for convection migrates zonally with seasons, MJO events propagating across the MC region are most likely to be disrupted in boreal spring and summer when they propagate from a favorable environment over the Indian Ocean to an unfavorable one over the MC. Analysis of signal-to-noise ratio indicates potentially better predictability during spring and summer especially when MJO activities are away from the MC region. While the model has obvious room for improvement and generalization, such as introducing longer-term memory and additional indices representing the state of the extratropical atmosphere and ocean etc., this work

represents a promising demonstration of potential futures of machine learning approach. A more general class of reduced dimension models can complement existing approaches to addressing the MJO and other climate variability and predictability problems.

Data Availability Statement

Daily time series of the RMM index are available at http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt, accessed on 13 January 2016. The ENSO index data sets are available at https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php. The QBO index data are at https://www.cpc.ncep.noaa.gov/data/indices/qbo.u50.index. The IOD index data is available at https://psl.noaa.gov/gcos_wgsp/Timeseries/Data/dmi.had.long.data.

Acknowledgments
This work is supported by the National Oceanic and Atmospheric Administration (NOAA) Oceanic and Atmospheric Research, Climate Program Office (CPO), under NOAA grant number NA17OAR4310263, as well as the U.S. Department of Energy Office of Science Biological and Environmental Research as part of the Atmospheric Systems Research Program and Global and Regional Model Analysis program area. Computing resources are provided by the National Energy Research Scientific Computing Center (NERSC). Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RLO1830. This is Pacific Marine Environmental Laboratory contribution 5288.

References

Back, S.-Y., Han, J.-Y., & Son, S.-W. (2020). Modeling evidence of QBO-MJO connection: A case study. *Geophysical Research Letters*, *47*, e2020GL089480. https://doi.org/10.1029/2020GL089480

Burleyson, C. D., Hagos, S. M., Feng, Z., Kerns, B. W. J., & Kim, D. (2018). Large-scale environmental characteristics of MJOs that strengthen and weaken over the maritime continent. *Journal of Climate*, *31*(14), 5731–5748. https://doi.org/10.1175/jcli-d-17-0576.1

Gagniuc, P. A. (2017). *Markov chains: From theory to implementation and experimentation*. John Wiley & Sons.

Hagos, S., Zhang, C., Leung, L. R., Burleyson, C. D., & Balaguru, K. (2019). A zonal migration of monsoon moisture flux convergence and the strength of Madden-Julian Oscillation events. *Geophysical Research Letters*, *46*, 8554–8562. https://doi.org/10.1029/2019GL083468

Hagos, S., Zhang, C., Leung, L. R., Garuba, O., Burleyson, C. D., & Balaguru, K. (2020). Impacts of insolation and soil moisture on the seasonality of interactions between the Madden-Julian Oscillation and maritime continent. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD032382. https://doi.org/10.1029/2020jd032382

Jones, C. (2009). A homogeneous stochastic model of the Madden–Julian oscillation. *Journal of Climate*, *22*(12), 3270–3288. https://doi.org/10.1175/2008jcli2609.1

Kerns, B. W., & Chen, S. S. (2016). Large-scale precipitation tracking and the MJO over the Maritime Continent and Indo-Pacific warm pool. *Journal of Geophysical Research: Atmospheres*, *121*, 8755–8776. https://doi.org/10.1002/2015JD024661

Lim, Y., Son, S. W., Marshall, A. G., Hendon, H. H., & Seo, K.-H. (2019). Influence of the QBO on MJO prediction skill in the subseasonal-to-seasonal prediction models. *Climate Dynamics*, *53*, 1681–1695. https://doi.org/10.1007/s00382-019-04719-y

Liu, C., Tian, B., Li, K. F., Manney, G. L., Livesey, N. J., Yung, Y. L., & Waliser, D. E. (2014). Northern hemisphere mid-winter vortex-displacement and vortex-split stratospheric sudden warmings: Influence of the Madden-Julian oscillation and Quasi-Biennial oscillation. *Journal of Geophysical Research: Atmospheres*, *119*, 12599–12620. https://doi.org/10.1002/2014jd021876

Marshall, A. G., Hendon, H. H., Son, S. W., & Lim, Y. (2017). Impact of the quasi-biennial oscillation on predictability of the Madden–Julian oscillation. *Climate Dynamics*, *49*, 1365–1377. https://doi.org/10.1007/s00382-016-3392-0

Martin, Z., Son, S. W., Butler, A., Harry, H., Kim, H., Sobel, A., et al. (2021). The influence of the quasi- biennial oscillation on the Madden–Julian oscillation. *Nature Reviews Earth & Environment*, *2*, 477–489. https://doi.org/10.1038/s43017-021-00173-9

Moum, J., Pujiana, K., Lien, R. C., & Smyth, W. D. (2016). Ocean feedback to pulses of the Madden–Julian oscillation in the equatorial Indian ocean. *Nature Communications*, *7*, 13203. https://doi.org/10.1038/ncomms13203

Nishimoto, E., & Yoden, S. (2017). Influence of the stratospheric quasi-biennial oscillation on the Madden–Julian oscillation during austral summer. *Journal of the Atmospheric Sciences*, *74*, 1105–1125. https://doi.org/10.1175/JAS-D-16-0205.1

Son, S.-W., Lim, Y., Yoo, C., Hendon, H. H., & Kim, J. (2017). Stratospheric control of the Madden–Julian oscillation. *Journal of Climate*, *30*(6), 1909–1922. https://doi.org/10.1175/jcli-d-16-0620.1

Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences: An introduction* (p. 467). Academic Press.

Wilson, E. A., Gordon, A. L., & Kim, D. (2013). Observations of the Madden Julian oscillation during Indian Ocean Dipole events. *Journal of Geophysical Research: Atmospheres*, *118*, 2588–2599. https://doi.org/10.1002/jgrd.50241

Yoo, C., & Son, S.-W. (2016). Modulation of the boreal wintertime Madden–Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophysical Research Letters*, *43*, 1392–1398. https://doi.org/10.1002/2016GL067762

Zhang, C., & Zhang, B. (2018). QBO-MJO connection. *Journal of Geophysical Research: Atmospheres*, *123*, 2957–2967. https://doi.org/10.1002/2017JD028171