

A more predictable Madden-Julian Oscillation index derived from Koopman spectral analysis

Claire Valva^{1,*}, Edwin P. Gerber¹

¹Courant Institute School of Mathematics, Computing, and Data Science, New York University

*Now Computing and Mathematical Sciences, California Institute of Technology.

Correspondence: clairev@caltech.edu

Abstract. The Madden-Julian oscillation (MJO) is a major source of subseasonal-to-seasonal (S2S) predictability. The MJO is commonly defined and tracked with indices such as the Real-time Multivariate MJO (RMM) index. Although the RMM provides a useful description of the MJO, its evolution can be noisy and difficult to predict. We define an MJO index using a data-driven approximation of the Koopman operator. The Koopman index captures similar tropical circulation and convection patterns to the RMM but evolves more smoothly and predictably. Skillful prediction extends to 46 days for the Koopman index compared to 11 days for the RMM under the same prediction framework. While this new approach does not recover the RMM as well as operational S2S models, which provide skillful forecasts up to 35 days, the Koopman index could complement existing MJO diagnostics in evaluating and developing extended-range forecast systems.

Plain-language summary. The Madden-Julian oscillation (MJO) is an eastward moving pattern of enhanced and suppressed tropical rainfall that lasts, approximately, from 30 to 60 days. The MJO is connected with many aspects of global weather including tropical cyclones and the monsoon. Here, we use a data-driven method to identify an index of the MJO and show that this is a more predictable index than those traditionally used.

1 Introduction

The Madden-Julian oscillation (MJO) is the most prominent mode of intraseasonal variability in the tropical atmosphere. It is characterized by large-scale coupled patterns in atmospheric circulation and deep convection that propagate eastward with a phase speed of 5 ms^{-1} and a lifespan of 30-60 days (Madden & Julian, 1971; Zhang, 2005). The MJO is known to influence many aspects of global climate and weather, including the onset of global monsoons and tropical cyclones, as well as extreme temperature and precipitation events (Jiang et al., 2020; Jones et al., 2004). Due to its wide-ranging impacts on global weather and its quasi-periodic occurrence, the MJO is one of the primary sources of predictability for subseasonal-to-seasonal (S2S) forecasts (Vitart et al., 2017).

The MJO is often defined and tracked by indices based on winds, outgoing longwave radiation (OLR), or other convection proxies. The Real-time Multivariate MJO (RMM) index, initially derived in Wheeler and Hendon (2004), is the most used index for tracking the MJO (Kiladis et al., 2014; Straub, 2013). The RMM is based on an empirical orthogonal function (EOF) analysis, defining a 2D phase space where the index propagates counterclockwise, from which the amplitude and the phase are computed. While this index is widely used, its propagation is relatively noisy, making it more challenging to predict.

Koopman operator theory maps finite-dimensional nonlinear dynamical systems to linear, albeit infinite dimensional, systems (Koopman, 1931). Eigendecompositions of the Koopman operator can be used to identify dynamically meaningful oscillatory modes with associated evolution frequencies. Such modes have been used to analyze and predict elements of the climate system with particular success for oscillatory phenomena. This was done initially for the MJO in Lintner, Giannakis, Pike, and Slawinska (2023), but has also had success in El Niño Southern Oscillation (ENSO) and the Quasi-Biennial Oscillation (Wang et al., 2020; Froyland et al., 2021; Valva & Gerber, 2025; Lorenzo-Sánchez et al., 2025; Lorenzo-Sánchez et al., 2026).

Here, we define an alternative index of the MJO using a data-driven approximation of the Koopman operator. While the Koopman index and the RMM provide similar descriptions of the MJO when viewed through conditional statistics, the Koopman based index evolves more smoothly and can be predicted skillfully on intraseasonal timescales.

2 Extracting intraseasonal oscillations from data with Koopman operator methods

We provide an overview of Koopman operator theory, and how to use it to extract oscillatory modes from tropical winds and OLR. Consider a nonlinear dynamical system with states $\{x_s\}_{s \in \mathbb{R}}$ — in our case the tropical atmosphere — and a flow map φ^t so that $\varphi^t(x_s) = x_{s+t}$. The map φ^t can be thought of as a model that integrates forward the state of the atmosphere from time s to time $s+t$. The *Koopman operator* K^t acts on observables f (functions) of the state

$$K^t f(x_s) = f \circ \varphi^t(x_s) = f(x_{s+t}). \quad (1)$$

An example observable f is the zonal wind speed at a given location above the Indian Ocean. Then $f(x_s)$ will give the zonal wind speed measurement at that location at time s , and $(K^t f)(x_s)$ the wind speed at time $s+t$.

While the dynamics φ^t are nonlinear, the Koopman operator K^t is linear when viewed as an operator on a space of observables f . As such, the advantage of the Koopman operator formalism is that we can perform a spectral analysis of K^t , which can be interpreted as coherent feature extraction. If (ω_k, ζ_k) is an eigenfrequency-eigenfunction pair of K^t , then

$$K^t \zeta_k = e^{t\omega_k} \zeta_k. \quad (2)$$

The eigenfunctions (the coherent features) ζ_k evolve predictably, oscillating with frequency ω_k .

We can associate a spatio-temporally varying mode M_k to the eigenpair (ω_k, ζ_k) when we approximate the operator from data. This mode M_k is computed from a projection of the input data onto the eigenfunction ζ_k and represents the variability of the projected variables associated to that eigenfunction. We will identify a particular Koopman mode of variability M_k to the MJO. Then the eigenfunction ζ_k can serve as an MJO index.

We construct the approximate Koopman operator with ERA5 (Hersbach et al., 2020) outgoing long-wave radiation (OLR) and zonal winds at 200 and 850 hPa at 1.5° spatial resolution, meridionally averaged from 15° N to 15° S. We normalize the OLR and zonal wind fields to be mean zero with unit variance in each variable, following the RMM computation convention (Wheeler & Hendon, 2004) and previously defined Koopman MJO index (Lintner et al., 2023). The modes are defined using data from January 2000 to December 2019, and the later forecasting experiment will be validated on the period January 2021 to December 2024. We additionally *delay-embed* the data with an embedding window of $N_e = 64$ days before analysis. Delay embedding replaces the vector of input data $\mathbf{X}_j$ at a given time indexed as j with $\hat{\mathbf{X}}_j$ such that the j th time and the N_e previous times are included in that snapshot, i.e., $\hat{\mathbf{X}}_j = \{\mathbf{X}_j, \mathbf{X}_{j-1}, \dots, \mathbf{X}_{j-N_e}\}$. This step serves to insert “memory” and uniqueness to each state, a relevant step given the compressed nature of the meridionally averaged dataset (Ghil et al., 2002). We will refer to this final data array as $\mathbf{X} \in \mathbb{R}^{N_t \times N_d}$ where N_t is the number of delay embedded snapshots and N_d is the dimension of each snapshot.

We use the Koopman approximation algorithm described in Giannakis (2019). This algorithm assumes the dynamical system of interest has an invariant probability measure (a stationary climatology), resulting in the eigenfrequencies ω_k being entirely imaginary. The approximate Koopman operator can then be written as

$$K^t f \approx \sum_k e^{t\omega_k} \zeta_k c_k, \quad c_k = \langle \zeta_k, f \rangle. \quad (3)$$

The original dataset can be reconstructed from the corresponding Koopman modes $M_k \in \mathbb{C}^{N_t \times N_d}$,

$$\mathbf{X} \approx \sum_k M_k. \quad (4)$$

The goal is to find a small number of modes M_k that capture variability of interest.

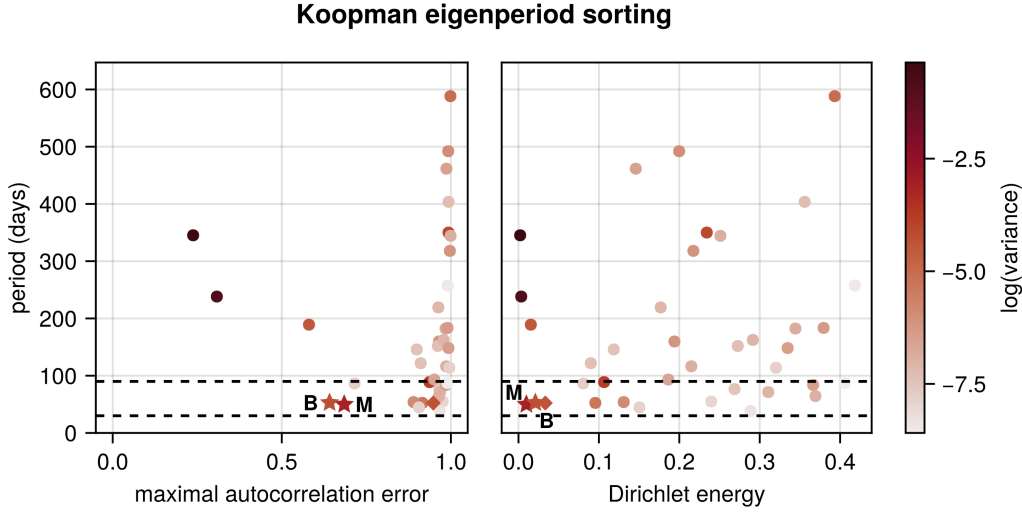


Figure 1. Koopman eigenperiods plotted according to two sorting metrics. The left panel sorts by a maximal autocorrelation error which ranges from 0 (best) to 1 (worst), while the right panel sorts by the Dirichlet energy metric where a smaller absolute value indicates a better, less noisy mode. The y-axis indicates the eigenperiod of each mode, while the x-axis is the value computed by each sorting metric. Marker color is associated with the variance of the mode. The horizontal lines indicate the period cutoff for intraseasonal modes (30-90 days). The stars indicate two intraseasonal modes of interest, interpreted to be the MJO (approx. 50 days period, labeled **M**) and the BSISO (approx. 53 days period, labeled **B**). The two points outside of the intraseasonal band that remain in good sorting positions are likely related to the annual cycle.

To use the resulting approximation to extract a dynamically meaningful representation of the MJO from data, we must address two questions. First, can we find a Koopman mode with the right frequency and spatiotemporal variability of the MJO? Second, is the resulting Koopman eigenpair (and associated Koopman mode) robust?

To begin answering these questions, we plot the periods of the obtained Koopman modes in figure 1. We can first target an eigenmode through its frequency ω_k . For example, to identify a representation of the annual cycle, we choose the eigenpair with $\omega_k = 1\text{yr}^{-1}$ and potentially higher harmonics. In the figure, we see a harmonic of the annual cycle around $\omega_k = 2\text{yr}^{-1}$ with a period of about 190 days. For the MJO, we will search for modes with an intraseasonal period, 30-90 days (dashed lines). There are many Koopman modes with periods within this intraseasonal range, but most of these eigenvalues represent a very small amount of variability (marker color) and so are insignificant.

To differentiate between the eigenpairs with the correct timescale of variability and meaningful variance, we assess their “trustworthiness” with two sorting metrics. First, we employ a metric of predictability called *maximal autocorrelation error* (proposed in Giannakis & Valva, 2024). This metric measures the distance of the eigenfunction ζ_k from a perfectly oscillatory eigenfunction with frequency ω_k ; a smaller score is better. Second, we sort by the Dirichlet energy of each eigenfunction, which measures of eigenfunction regularity; a lower energy indicates a more regular (better) eigenfunction. This serves as a proxy for numerical approximation error which has been used with empirical success in works including Das, Giannakis, and Slawinska (2021); Lintner et al. (2023); Valva and Gerber (2025). We precisely define these metrics in B.

Only two of the leading leading eigenpairs (marked by stars) in the intraseasonal range have favorable rankings under both metrics. To distinguish them, we examine their seasonality and propagation characteristics. One eigenfunction which we identify with the MJO (period ≈ 50 days) is more active during Boreal winter and has eastward-propagating variability. The other eigenmode (period ≈ 53 days) is more active during Boreal summer and exhibits meridional propagation, appearing to be associated with the Boreal Summer Intraseasonal Oscillation (BSISO). We leave analysis of

this possible BSISO mode to future work, and provide figures demonstrating the seasonality and propagation of this mode in the supplement. Here, we focus on the mode that is more active during Boreal winter, which we will call the MJO Koopman index.

3 The Koopman index captures key features of MJO variability

Comparison of bandpassed observations and the Koopman reconstruction

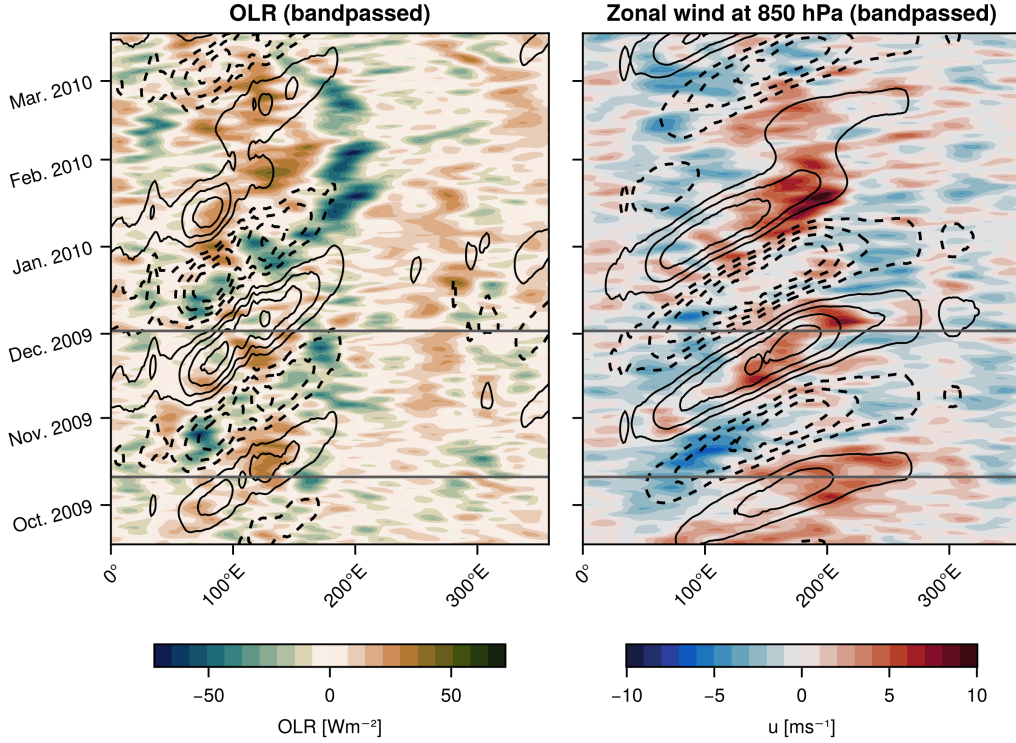


Figure 2. Intraseasonally bandpassed (20-96 days) OLR and zonal wind (filled contours) overlaid with the projection of the Koopman MJO (solid lines indicate positive values, dashed lines indicate negative values) during the YOTC period. Note the onset of two MJO events, seen by the initial band of negative OLR anomaly (enhanced convection): the first initiating in late October, the second in mid-December. Approximate initiation dates (October 25 and 16 December) are marked by lines on the figure. The Koopman projection tracks the large scale location and propagation of these events.

To establish the fitness of the Koopman index for characterizing and predicting the MJO, we ask whether the mode captures the evolution of MJO events. We first consider the behavior of the Koopman mode as compared to intraseasonally bandpassed OLR and zonal wind during the Year of Tropical Convection (YOTC) virtual experiment (Waliser et al., 2012), focusing on the period of October 2009 to March 2010.

Following past work (including Wheeler & Hendon, 2004), we bandpass OLR and zonal wind fields to retain variability between 20 and 96 days. Figure 2 overlays contours of the Koopman MJO mode on this bandpassed data. In the YOTC period, there were two MJO initiation events: the first in late October and the second in mid-December, which can be seen in the figure as negative OLR anomalies (indicating regions of enhanced convection) that propagate eastward at about 5 ms^{-1} followed by similar propagation of positive OLR anomalies (as interpreted in Kiladis et al. (2014)). These events are mirrored in the zonal wind with propagating westerlies followed by easterlies. The Koopman mode projected onto this data captures the initiation and propagation of both these events on the large scale, including the speed and location.

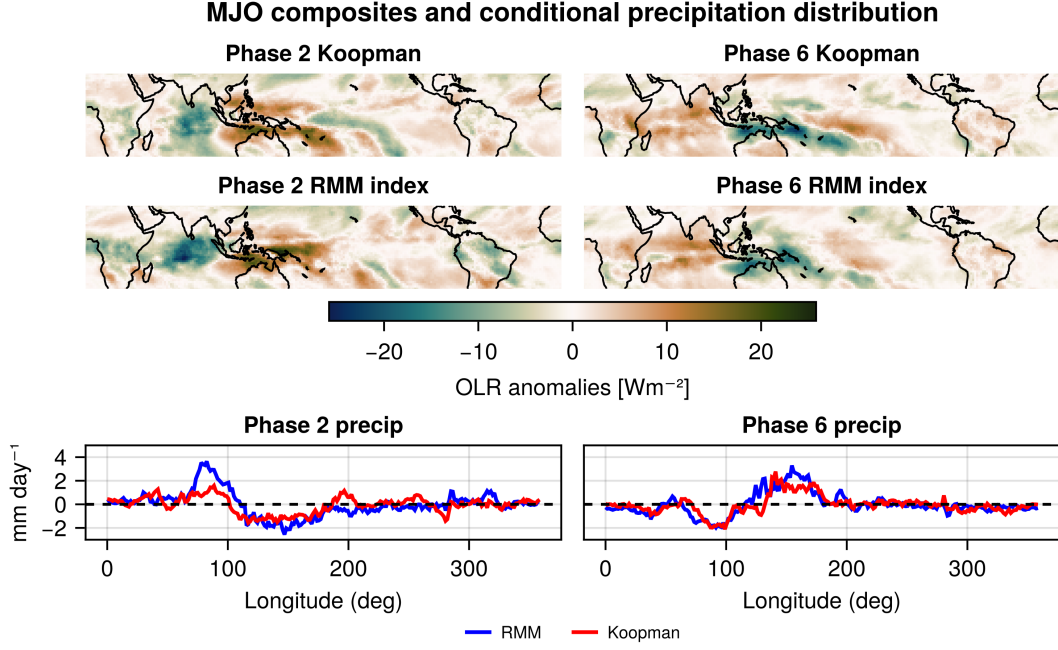


Figure 3. Comparison of the RMM and Koopman MJO representations. The top panel compares OLR composites from each of these indices in two phases. The bottom panel compares the conditional mean precipitation anomalies between the two indices. To focus on active MJO periods, we plot the precipitation anomalies (meridionally averaged between $15^\circ \text{ N} - 15^\circ \text{ S}$) associated with times when the amplitude of the MJO index exceeds the 75th percentile of the respective index.

We next examine the large-scale circulation and precipitation patterns associated with the Koopman index, comparing it with the RMM. We focus only on the extended Boreal Winter (DJFMA), as the selected Koopman index represents a boreal winter intraseasonal mode. The RMM is used as an all-season MJO index, but its boreal-summer variability also reflects aspects of the BSISO (Wheeler & Hendon, 2004).

So that we can directly compare the MJO phases between the two indices, we rotate the Koopman index to maximize the all-season correlation between the Koopman and RMM index, allowable as the complex phase of a Koopman eigenfunction is arbitrary. The DJFMA correlation between these two indices is 0.51 indicating that the indices have related, but not identical evolution. (The correlation can be improved somewhat if we intraseasonal bandpass the RMM, increasing to 0.58.) In the upper panel of figure 3 we plot the OLR composites for phases 2 and 6. The Koopman and RMM OLR composites show consistent representations of the MJO. Both exhibit similar large scale patterns of enhanced and suppressed convection that propagate in the same manner.

To reveal how these indices capture precipitation, we also show daily ERA5 precipitation anomalies associated with active MJO periods in phases 2 and 6 for the two respective indices in figure 3. Specifically, we plot the precipitation anomaly conditioned on an index exceeding a 75th percentile threshold to ensure that the index is identifying an active MJO. Both indices identify a positive precipitation anomaly over the Indian Ocean and a negative anomaly over the Western Pacific in phase 2. In phase 6, this pattern is reversed. The spatial patterns of positive and negative daily precipitation anomalies are quite similar, although the Koopman index exhibits slightly weaker amplitude. We conclude that the Koopman representation captures key features of MJO variability at both the event and composite levels, including its large-scale propagation and associated precipitation.

4 The Koopman MJO index is highly predictable

The Koopman formalism selects for spatiotemporal modes with approximately oscillatory behavior, yielding an index constructed to provide a more predictable representation of MJO-related behavior. We demonstrate this by comparing forecasts of the Koopman and RMM indices.

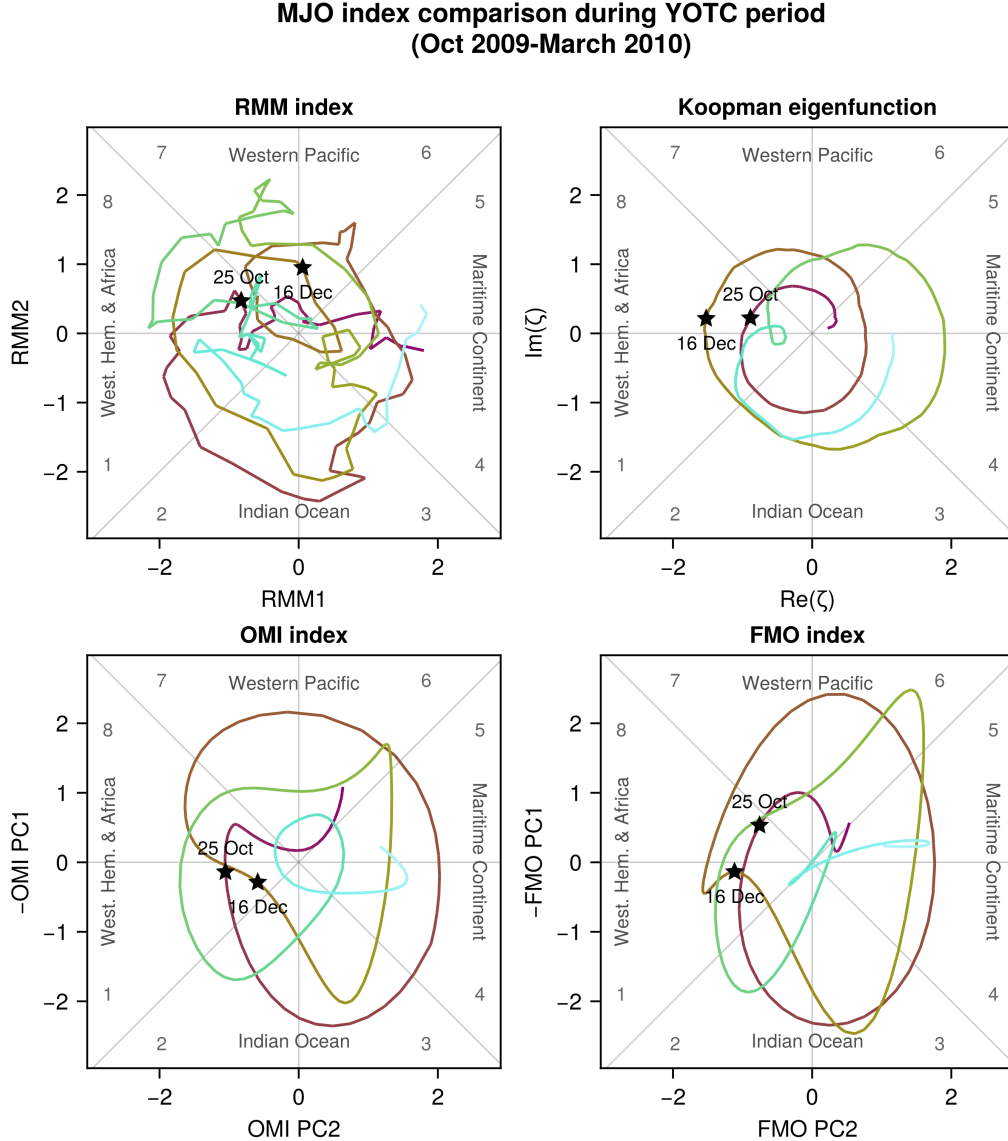


Figure 4. Comparison of the RMM, OMI, FMO and Koopman eigenfunctions during the YOTC period. Indices evolve (approximately) counterclockwise, with the color (red-to-blue) changing over time. The initiation dates of the two MJO events during this period (October 25th, December 16th) are marked by stars.

While the Koopman and RMM indices capture similar MJO variability on large spatiotemporal scales, their short term evolution differs: the Koopman index propagates more smoothly than the RMM. In figure 4, we plot the evolution of the Koopman MJO index as compared to the RMM during the YOTC period. While both track the counterclockwise propagation of the MJO, the evolution of the Koopman MJO index is much smoother and easier to interpret. The Koopman MJO exhibits two complete counterclockwise “loops” of the index in the complex plane, as well as a partial loop, which correspond to the two eastward propagating MJO events in this period and the initiation of the enhanced convection near the end of the YOTC period (Figure 2). While the RMM propagates

counterclockwise as well, it sketches a fairly noisy, rough pattern. One can make out 3-4 rotations, depending on interpretation of the noise. This suggests that the RMM may be potentially harder to predict.

The lower panels of Figure 4 allow comparison between the Koopman index to smoother MJO indices constructed from filtered data: the OMI (the OLR MJO Index) and the FMO (the Filtered OLR MJO index) (Kiladis et al., 2014). These indices evolve more smoothly like the Koopman MJO, but they are not “real-time” indices (i.e., cannot be computed instantaneously), and less useful for forecasting purposes.

We mark the two YOTC MJO initiation dates in the figure with stars. The Koopman index places both initiation dates in the same phase (8, corresponding to the Western Maritime Continent), while the RMM and FMO separates the phase of these dates. The OMI and Koopman phases assign similar dynamical states to the initiation of two distinct MJO events. The RMM exhibits more inconsistency for the second event in particular, progressing into the corresponding phase later than the Koopman index. This aligns with previously reported differences between RMM phase progression and MJO convective evolution (Kiladis et al., 2014).

Finally, we establish that the smooth, consistent evolution of the Koopman index translates into forecast skill. We make forecasts from meridionally averaged ERA5 OLR and zonal wind using the same 64-day delay embedding as in the data analysis. To ensure a fair test, for a lead time ℓ , the predictor consists of daily observations from $64 + \ell$ to ℓ days *before* the forecast target. We evaluate forecast skill over DJFMA 2021–2024, defining the season based on the calendar year of January, which is completely distinct from the 2000–2019 analysis period that we used to establish the Koopman index.

We use kernel analog forecasts (KAFs) to predict the Koopman and RMM indices. KAF is a non-parametric regression method in which predictions are constructed from the evolution of historical states, weighted according to their similarity to the current observation (Alexander et al., 2017). The idea is rather simple: a forecast is made by finding similar conditions in the past and then using the associated history as the forecast. A kernel function measures the similarity. We use a reduced kernel eigenfunction basis, as described in A, and use only data from the analysis period.

We construct three KAF forecasts: direct forecasts of both the Koopman and RMM indices and an indirect RMM forecast obtained by first forecasting the Koopman index and then estimating RMM from the predicted Koopman state. For comparison, we also evaluate a subset of the operational S2S forecasts of the RMM from the S2S Prediction Project (Vitart et al., 2017). Figure 5 compares these forecasts using the bivariate correlation as a metric. We consider a correlation of 0.5 or above a skillful forecast.

The smooth evolution of the Koopman MJO index makes it extremely predictable: simple analog forecasts maintain predictive skill for 46 days. This is well beyond the ability of the operational S2S models to forecast the MJO via the RMM, which are skillful for 15 to 35 days depending on the model (consistent with S2S forecast skill reported in previous studies: Lim et al., 2018; Zhou et al., 2024).

The difference in forecasting skill lies primarily in that the S2S models are aiming for a harder target, the RMM, as compared to the smoothly evolving Koopman index. To illustrate this, we include the two KAF forecasts of RMM derived from the Koopman indices. In figure 5, the dashed purple line shows how well we can forecast the RMM based on the Koopman MJO index alone: skillful forecasts are only possible for 11 days. The red dashed line shows a direct KAF forecast created using all the computed Koopman modes, but the improvement is minimal. The predictive skill of the RMM KAF forecasts decay at a similar rate to the S2S models, but begins with a rather low correlation value at day 1. The Koopman approach cannot capture the high-frequency variations of the RMM, as it selects for coherent, intraseasonal variability.

These results indicate that the Koopman index is substantially easier to forecast than the RMM. We emphasize that the Koopman and RMM indices do not precisely align. Nonetheless, the Koopman and RMM indices represent the same large scale variability associated with the MJO. Our main

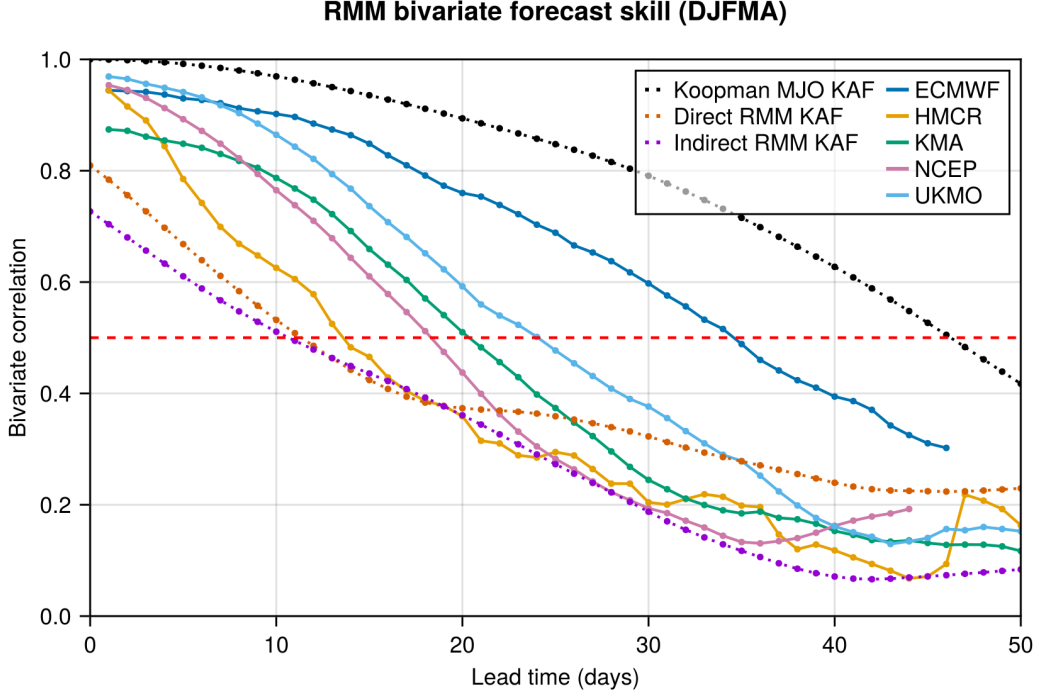


Figure 5. Comparison of S2S and KAF forecasts of RMM and the Koopman MJO Index. Solid lines are forecasts of the RMM from the S2S Prediction Project, while the dotted lines are KAF forecasts of the RMM and the Koopman MJO index. The KAF forecast of the Koopman MJO index is skillful for 46 days. In contrast, S2S forecasts of the RMM show skill ranging from 13 to 35 days, while the KAF forecast of RMM is only skillful for 11 days.

argument is that the Koopman index may be a better forecasting target in the subseasonal range.

5 Discussion

The Koopman approach provides a way to construct a data-driven representation of the MJO that is both physically interpretable and forecastable. The Koopman index captures primary features of the MJO variability, including its large-scale propagation and phase-dependent precipitation anomalies. Despite providing a similar MJO representation to the RMM, the two indices differ in their predictability. The Koopman index remains skillfully predictable up to 46 days with a simple kernel analog forecast compared to the RMM’s 11. By comparison, the operational RMM forecasts remain skillful to 35 days.

The ability to retain MJO characteristics while producing a more predictable index suggests that Koopman-based representations may provide useful alternative targets for extended-range forecasting. In addition, the construction of Koopman indices can be adapted to specific forecasting objectives. Here, we have chosen OLR and zonal winds as our input data to allow for direct comparison with the RMM. Including precipitation anomalies directly in the initial analysis step, however, could produce Koopman modes that more explicitly represent the precipitation variability associated with the MJO.

Open Research Statement

ERA5 zonal wind, outgoing longwave radiation, and precipitation data were obtained from the ECMWF Climate Data Store.

RMM indices from the S2S Prediction Project forecasts were obtained from <https://confluence.ecmwf.int/spaces/S2S/pages/40796953/Tools#Tools-MJORMMsindices>. We refer to Vitart et al.

(2017) for details of the S2S forecast database.

Observed RMM indices were obtained from the Australian Bureau of Meteorology: <http://www.bom.gov.au/climate/mjo/#tabs=Monitoring>.

The OMI and FMO values were obtained from <https://psl.noaa.gov/mjo/mjoindex/>.

AI Disclosure

The Julia code used in analysis was developed with the assistance of Codex (GPT 5.5 and 5.6). Authors take full responsibility for accuracy.

Acknowledgments

E.G. and C.V. acknowledge support from NSF OAC-2004572 and Schmidt Sciences, as part of the Virtual Earth System Research Institute (VESRI). C.V.'s postdoctoral work, during which the manuscript was completed, is supported by the DoD Vannevar Bush Faculty Fellowship N00014-22-1-2790.

References

- Alexander, R., Zhao, Z., Székely, E., & Giannakis, D. (2017, April). Kernel Analog Forecasting of Tropical Intraseasonal Oscillations. *Journal of the Atmospheric Sciences*, 74(4), 1321–1342. Retrieved from <https://journals.ametsoc.org/view/journals/atsc/74/4/jas-d-16-0147.1.xml> doi: 10.1175/JAS-D-16-0147.1
- Das, S., Giannakis, D., & Slawinska, J. (2021, September). Reproducing kernel Hilbert space compactification of unitary evolution groups. *Applied and Computational Harmonic Analysis*, 54, 75–136. Retrieved 2022-04-11, from <https://www.sciencedirect.com/science/article/pii/S1063520321000178> doi: 10.1016/j.acha.2021.02.004
- Froyland, G., Giannakis, D., Lintner, B. R., Pike, M., & Slawinska, J. (2021, November). Spectral analysis of climate dynamics with operator-theoretic approaches. *Nature Communications*, 12(1), 1–21. Retrieved 2021-11-12, from <https://www.nature.com/articles/s41467-021-26357-x> (Cc_license_type: cc_by Number: 1 Primary_atype: Research Subject_term: Applied mathematics;Climate sciences Subject_term_id: applied-mathematics;climate-sciences) doi: 10.1038/s41467-021-26357-x
- Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., ... Yiou, P. (2002). Advanced Spectral Methods for Climatic Time Series. *Reviews of Geophysics*, 40(1), 3–1–3–41. Retrieved 2026-03-06, from <https://onlinelibrary.wiley.com/doi/abs/10.1029/2000RG000092> doi: 10.1029/2000RG000092
- Giannakis, D. (2019, September). Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, 47(2), 338–396. Retrieved 2021-08-16, from <https://www.sciencedirect.com/science/article/pii/S1063520317300982> doi: 10.1016/j.acha.2017.09.001
- Giannakis, D., & Valva, C. (2024, June). Consistent spectral approximation of Koopman operators using resolvent compactification. *Nonlinearity*, 37(7), 075021. Retrieved 2024-07-10, from <https://dx.doi.org/10.1088/1361-6544/ad4ade> doi: 10.1088/1361-6544/ad4ade
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. (2020, July). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. Retrieved 2023-10-25, from <https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3803> doi: 10.1002/qj.3803
- Jiang, X., Adames, . F., Kim, D., Maloney, E. D., Lin, H., Kim, H., ... Klingaman, N. P. (2020). Fifty Years of Research on the Madden-Julian Oscillation: Recent Progress, Challenges, and Perspectives. *Journal of Geophysical Research: Atmospheres*, 125(17), e2019JD030911. Retrieved 2026-01-28, from <https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030911> doi: 10.1029/2019JD030911
- Jones, C., Waliser, D. E., Lau, K. M., & Stern, W. (2004, December). Global Occurrences of Extreme Precipitation and the Madden-Julian Oscillation: Observations and Predictability. *Journal of Climate*, 17(23), 4575–4589. Retrieved 2026-07-17, from <https://journals.ametsoc.org/view/journals/clim/17/23/3238.1.xml> doi: 10.1175/3238.1
- Kiladis, G. N., Dias, J., Straub, K. H., Wheeler, M. C., Tulich, S. N., Kikuchi, K., ... Ventrice, M. J. (2014, May). A Comparison of OLR and Circulation-Based Indices for Tracking the MJO. Retrieved 2024-06-21, from <https://journals.ametsoc.org/view/journals/mwre/142/5/mwr-d-13-00301.1.xml> doi: 10.1175/MWR-D-13-00301.1
- Koopman, B. O. (1931, May). Hamiltonian Systems and Transformation in Hilbert Space. *Proceedings of the National Academy of Sciences*, 17(5), 315–318. Retrieved 2026-07-17, from <https://www.pnas.org/doi/abs/10.1073/pnas.17.5.315> doi: 10.1073/pnas.17.5.315
- Lim, Y., Son, S.-W., & Kim, D. (2018, May). MJO Prediction Skill of the Subseasonal-to-Seasonal Prediction Models. Retrieved 2025-05-27, from <https://journals.ametsoc.org/view/journals/clim/31/10/jcli-d-17-0545.1.xml> doi: 10.1175/JCLI-D-17-0545.1
- Lintner, B. R., Giannakis, D., Pike, M., & Slawinska, J. (2023). Identification of the Madden-Julian Oscillation With Data-Driven Koopman Spectral Analysis. *Geophysical Research Letters*, 50(10), e2023GL102743. Retrieved 2023-11-22, from <https://onlinelibrary.wiley.com/doi/abs/10.1029/2023GL102743> doi: 10.1029/2023GL102743

- Lorenzo-Sanchez, P., Colbrook, M. J., & Navarra, A. (2026, June). *Residual Pseudospectra Reveal a Physics-Informed Koopman Backbone for Tropical Pacific Variability and ENSO Prediction*. arXiv. Retrieved 2026-07-13, from <http://arxiv.org/abs/2606.09369> (arXiv:2606.09369 [math.NA]) doi: 10.48550/arXiv.2606.09369
- Lorenzo-Sánchez, P., Newman, M., Albers, J., Subramanian, A. C., & Navarra, A. (2025, November). Koopman Theory for Enhanced Tropical Pacific SST Forecasting. *Artificial Intelligence for the Earth Systems*, 4(4). Retrieved 2026-07-17, from <https://journals.ametsoc.org/view/journals/aies/4/4/AIES-D-24-0088.1.xml> doi: 10.1175/AIES-D-24-0088.1
- Madden, R. A., & Julian, P. R. (1971, July). Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*, 28(5), 702–708. Retrieved 2026-07-17, from https://journals.ametsoc.org/view/journals/atsc/28/5/1520-0469_1971_028_0702_doadoi_2_0_co_2.xml doi: 10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2
- Straub, K. H. (2013, February). MJO Initiation in the Real-Time Multivariate MJO Index. *Journal of Climate*, 26(4), 1130–1151. Retrieved 2026-01-26, from <https://journals.ametsoc.org/view/journals/clim/26/4/jcli-d-12-00074.1.xml> doi: 10.1175/JCLI-D-12-00074.1
- Valva, C., & Gerber, E. P. (2025, July). The QBO, the Annual Cycle, and Their Interactions: Isolating Periodic Modes with Koopman Analysis. *Journal of Climate*, 38(14), 3437–3452. Retrieved 2025-11-18, from <https://journals.ametsoc.org/view/journals/clim/38/14/JCLI-D-24-0411.1.xml> doi: 10.1175/JCLI-D-24-0411.1
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., ... Zhang, L. (2017, January). The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, 98(1), 163–173. Retrieved 2026-07-15, from <https://journals.ametsoc.org/view/journals/bams/98/1/bams-d-16-0017.1.xml> doi: 10.1175/BAMS-D-16-0017.1
- Waliser, D. E., Moncrieff, M. W., Burridge, D., Fink, A. H., Gochis, D., Goswami, B. N., ... Yuter, S. (2012, August). The “Year” of Tropical Convection (May 2008–April 2010): Climate Variability and Weather Highlights. *Bulletin of the American Meteorological Society*, 93(8), 1189–1218. Retrieved 2026-07-13, from <https://journals.ametsoc.org/doi/10.1175/2011BAMS3095.1> doi: 10.1175/2011BAMS3095.1
- Wang, X., Slawinska, J., & Giannakis, D. (2020, February). Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports*, 10(1), 2636. Retrieved 2026-07-17, from <https://www.nature.com/articles/s41598-020-59128-7> doi: 10.1038/s41598-020-59128-7
- Wheeler, M. C., & Hendon, H. H. (2004, August). An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. Retrieved 2024-06-21, from https://journals.ametsoc.org/view/journals/mwre/132/8/1520-0493_2004_132_1917_aarmmi_2_0_co_2.xml
- Zhang, C. (2005). Madden-Julian Oscillation. *Reviews of Geophysics*, 43(2). Retrieved 2023-09-05, from <https://onlinelibrary.wiley.com/doi/abs/10.1029/2004RG000158> doi: 10.1029/2004RG000158
- Zhou, X., Wang, L., Hsu, P.-c., Li, T., & Xiang, B. (2024, September). Understanding the Factors Controlling MJO Prediction Skill across Events. *Journal of Climate*, 37(20), 5323–5336. Retrieved 2026-01-28, from <https://journals.ametsoc.org/view/journals/clim/37/20/JCLI-D-23-0635.1.xml> doi: 10.1175/JCLI-D-23-0635.1

A Kernel basis and kernel analog forecasting

Both the Koopman approximation and kernel analog forecasting used in this study employ a reduced basis constructed from eigenfunctions of a kernel integral operator. Let $\mathbf{X}_j \in \mathbb{R}^{N_d}$ denote the delay-embedded observations defined in Section 2, and let $k : \mathbb{R}^{N_d} \times \mathbb{R}^{N_d} \rightarrow \mathbb{R}$ denote the normalized kernel used in the analysis. The corresponding empirical kernel operator $\mathcal{K}_{N_t}$ acts on functions g as

$$(\mathcal{K}_N g)(\mathbf{Y}) = \frac{1}{N_t} \sum_{j=0}^{N_t-1} k(\mathbf{Y}, \mathbf{X}_j) g(\mathbf{X}_j), \quad (5)$$

with eigenfunctions ϕ_j and eigenvalues λ_j satisfying

$$\mathcal{K}_{N_t} \phi_j = \lambda_j \phi_j. \quad (6)$$

For a new datapoint $\mathbf{Y} \in \mathbb{R}^{N_d}$, an eigenfunction with $\lambda_j \neq 0$ can be evaluated using the Nyström extension,

$$\phi_j(\mathbf{Y}) = \frac{1}{N \lambda_j} \sum_{n=0}^{N-1} k(\mathbf{Y}, \mathbf{X}_n) \phi_j(\mathbf{X}_n). \quad (7)$$

The value of $\phi_j(\mathbf{Y})$ is determined by a kernel-weighted average over training states, with larger contributions from states $\mathbf{X}_n$ that are close to $\mathbf{Y}$ in the kernel-induced geometry.

Kernel analog forecasting (KAF) uses this basis to estimate the future value of a forecast observable. Let $f : \mathbb{R}^{N_d} \rightarrow \mathbb{C}$ denote a scalar- or complex-valued observable, so that $f(\mathbf{X}_n)$ is its value at time n . For a forecast lead ℓ , we approximate the map from a current observation Y to the future value of the observable using the first L kernel eigenfunctions,

$$\hat{f}_\ell(\mathbf{Y}) = \sum_{j=0}^{L-1} c_{j,\ell} \phi_j(\mathbf{Y}), \quad c_{j,\ell} = \frac{1}{N - \ell} \sum_{n=0}^{N-\ell-1} \phi_j(\mathbf{X}_n)^* f(\mathbf{X}_{n+\ell}). \quad (8)$$

Thus, $\hat{f}_\ell(\mathbf{Y})$ estimates the value of f at lead ℓ from the current observation Y .

We use this construction for three forecasts. First, we directly forecast the Koopman MJO index (with data $\mathbf{Z}_n \in \mathbb{C}$) and the RMM index (with data $\mathbf{R}_n = \text{RMM1}_n + i \text{RMM2}_n$).

We also make an indirect forecast, where we first predict the Koopman index using KAF and then regress the RMM onto the Koopman index to estimate the RMM index. The direct and indirect RMM experiments distinguish between predicting the RMM field directly from the OLR and zonal wind fields and predicting it via the intermediate Koopman representation.

B Eigenvalue sorting

B.1 Autocorrelation sorting

We sort the eigenfunctions by the *maximal autocorrelation error*, which measures the extent to which a computed eigenfunction ζ_k departs from ideal Koopman eigenfunction behavior over a specified time interval, providing an estimate of predictive skill.

Let $C_f(t)$ denote the time-autocorrelation function of a function f with zero-mean, i.e. $C_f(t) = \langle f, U^t f \rangle$. If ζ is a perfect Koopman eigenfunction with eigenfrequency $\omega \in \mathbb{C}$ then $C_\zeta(t) = e^{\omega t}$. This, as well as a desire to favor eigenfunctions whose autocorrelation remains coherent over the timescale of interest, motivates the use of $\epsilon_{T_c}(\omega, \zeta)$ as a sorting parameter

$$\epsilon_{T_c}(\omega, \zeta) = \max_{t \in [-T_c, T_c]} \text{Re}(1 - e^{-\omega t} C_\zeta(t)), \quad (9)$$

where T_c is the time window of interest (here, 45 days). A perfect Koopman eigenfunction-eigenvalue pair will have $\epsilon_{T_c}(\omega, \zeta) = 0$ and all others will have $\epsilon_{T_c}(\omega, \zeta) \in (0, 1]$. As such $\epsilon_{T_c}(\omega, \zeta)$ quantifies the departure of a function from Koopman eigenfunction behavior over the interval $[-T_c, T_c]$, and is closely related to the ϵ -approximate spectrum of the Koopman operator (see Giannakis & Valva, 2024).

B.2 Dirichlet energy

We also sort the computed Koopman eigenfunctions by their Dirichlet energy, a measure of eigenfunction regularity. Let λ_k and ϕ_k denote the kernel eigenvalues and corresponding orthonormal eigenfunctions as defined in eq. 6. Associated with each ϕ_k is the Dirichlet energy

$$\eta_k = \frac{1}{\epsilon} \left(\frac{1}{\lambda_k} - 1 \right), \quad (10)$$

where ϵ is the kernel bandwidth parameter. The kernel eigenvalues are ordered from largest to smallest, η_k increases with k .

Each computed Koopman eigenfunction ζ_j is represented in this basis as

$$\zeta_j = \sum_{k=0}^{L-1} c_{j,k} \phi_k. \quad (11)$$

Because the ϕ_k are orthonormal eigenfunctions of the corresponding diffusion operator, the Dirichlet energy of ζ_j is

$$E(\zeta_j) = \sum_{k=0}^{L-1} \eta_k |c_{j,k}|^2 = \frac{1}{\epsilon} \sum_{k=0}^{L-1} \left(\frac{1}{\lambda_k} - 1 \right) |c_{j,k}|^2. \quad (12)$$

Eigenfunctions whose expansion places more weight on high-energy kernel basis functions have larger Dirichlet energy and are less regular.