Geophysical Research Letters®

Author Contributions:

Conceptualization: Benjamin R. Lintner, Dimitrios Giannakis, Joanna Slawinska
Data curation: Max Pike
Formal analysis: Benjamin R. Lintner, Max Pike
Funding acquisition: Benjamin R. Lintner, Dimitrios Giannakis
Investigation: Benjamin R. Lintner
Methodology: Dimitrios Giannakis, Joanna Slawinska
Software: Dimitrios Giannakis
Visualization: Benjamin R. Lintner

Identification of the Madden–Julian Oscillation With Data-Driven Koopman Spectral Analysis

Benjamin R. Lintner[1,2] , **Dimitrios Giannakis[3]** , **Max Pike[1]**, and **Joanna Slawinska[3]**

[1]Department of Environmental Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA, [2]Rutgers Institute of Earth, Ocean, and Atmospheric Sciences, New Brunswick, NJ, USA, [3]Department of Mathematics, Dartmouth College, Hanover, NH, USA

Abstract The Madden-Julian Oscillation (MJO), the dominant mode of tropical intraseasonal variability, is commonly identified using the realtime multivariate MJO (RMM) index based on joint empirical orthogonal function (EOF) analysis of near-equatorial upper and lower level zonal winds and outgoing longwave radiation. Here, in place of conventional EOFs, we apply an operator-theoretic formalism based on dynamic systems theory (the Koopman operator) to extract an analog to RMM that exhibits certain features that refine the characterization and predictability of the MJO. In particular, the spectrum of Koopman operator eigenfunctions, with eigenvalues corresponding to mode periods, contains a leading intraseasonal mode with period of $\sim$50 days. Moreover, the amplitude of this leading intraseasonal eigenfunction exhibits a seasonal modulation clearly peaked in boreal winter. Finally, the phase space formed by the complex Koopman MJO eigenfunction exhibits a smoother temporal evolution and higher degree of autocorrelation than RMM, which may contribute to enhanced predictive skill.

Plain Language Summary The Madden-Julian Oscillation (MJO) is a key driver of tropical variability on timescales of a few weeks to months. Conventionally, the MJO is identified from observations through a widely used geophysical data analysis technique known as empirical orthogonal function analysis. Here, a fundamentally different approach, rooted in theoretical understanding of dynamical systems, is used to extract the MJO from a combination of daily outgoing longwave radiation and zonal wind data. The principal features of this representation of the MJO, such as its propagation speed, are shown to compare favorably to the standard representation while potentially providing enhanced predictability, which could be exploited to improve future forecasts of MJO conditions.

1. Introduction

First identified in the 1970s, the Madden-Julian Oscillation (MJO) is now recognized as the dominant mode of tropical intraseasonal variability (Madden & Julian, 1971). Characterized by anomalous precipitating deep convection and associated atmospheric circulation structures, the MJO propagates eastward along the equator with phase speeds of $\sim$5 m s^{-1}. MJO events frequently initiate over the Indian Ocean and may persist for 30–60 days (Zhang, 2005). Such events are most prevalent during boreal winter, although an analogous intraseasonal mode exhibiting more poleward propagation, the Boreal Summer Intraseasonal Oscillation (BSISO), occurs in boreal summer (Madden, 1986). While much progress has been made to advance understanding of the physics of the MJO, fundamental questions remain regarding its genesis, propagation, and seasonality, among other characteristics (Jiang et al., 2020).

The MJO is frequently described via indices derived from outgoing longwave radiation (OLR), winds, and/or rainfall as inputs (Schreck et al., 2013). Among the most widely used MJO indices is the real-time multivariate MJO (or RMM) index, first introduced by Wheeler and Hendon (2004). Although the precise details of the RMM calculation may differ, it is fundamentally based on joint empirical orthogonal function (EOF) analysis of OLR and upper and lower level zonal winds (u, at 200 and 850 mb, respectively), after some filtering to isolate intraseasonal band variability by removing the annual cycle and its harmonics, El Niño/Southern Oscillation (ENSO), and lower frequency variability. From this joint EOF, the two leading modes, accounting for a similar amount of explained variance and comprising a pair in quadrature, are identified as the MJO; these modes, conventionally referred to as RMM1 and RMM2, define the familiar 2D phase space used to index MJO amplitude and phase (see Figure 1 below).

Writing – original draft: Benjamin R. Lintner, Max Pike
Writing – review & editing: Benjamin R. Lintner, Dimitrios Giannakis

The present study addresses MJO identification through application of an operator-theoretic technique for dynamical systems based on data-driven approximations of the Koopman operator. First introduced in the early 1930s (Koopman, 1931; Koopman & von Neumann, 1932), operator-theoretic approaches study dynamical systems through their induced action on spaces of observables, that is, functions of the state that represent physical quantities that can be measured or observed. A key observation of Koopman (1931) was that a general (nonlinear) dynamical system acts on such spaces by means of intrinsically *linear* transformations, which have since come to be known as Koopman operators. This perspective enables the use of linear operator theory to study nonlinear dynamics, with the Koopman operator's spectral decomposition fully characterizing the evolution of observables under the nonlinear system. Closely related to Koopman operator techniques are methods based on the transfer operator, which is the adjoint of the Koopman operator acting on spaces of measures (e.g., probability densities).

In recent years, there has been considerable interest in developing data analysis and forecasting techniques based on Koopman and transfer operators (Brunton et al., 2022; Dellnitz & Junge, 1999; Mezić, 2005), which aim to provide more dynamically relevant and physically interpretable results than dynamics-agnostic techniques such as EOF analysis. Applications of operator-theoretic approaches to climate dynamics include identification (Froyland et al., 2021) (hereafter, F21) and forecasting (Navarra et al., 2021; Wang et al., 2020) of ENSO, forecasting of sea ice concentration (Hogg et al., 2020), and analysis of tropical convective variability on diurnal to seasonal timescales (Giannakis et al., 2015).

The data-driven Koopman framework employed in this work (Berry et al., 2020) leverages kernel methods for machine learning to build approximations of the Koopman operator from time series data and to obtain knowledge of the underlying dynamics that govern the evolution of observables. Eigendecomposition of the approximate Koopman operator yields a set of eigenfunctions $\{z_k\}$ that exhibit some properties rendering them useful for diagnosing and interpreting propagating, (quasi)periodic climate signals, as we illustrate below in the context of the MJO.

2. Methods and Data

The Koopman operator U^t can be thought of as evolving an observable from a given time to a time t later, that is, for an observable f that is a function of the dynamical state x at a given time:

$$U^t f(x) = f\left(\Phi^t x\right) \tag{1}$$

where Φ^t represents the flow map along which the system evolves in continuous time. The infinitesimal generator of the Koopman operator, V, is defined such that

$$V f = \lim_{t \to 0} \frac{U^t f - f}{t} \tag{2}$$

In the present study, we seek to construct and solve an eigenvalue equation for V, namely, $V z_k = i\omega_k z_k$, where $z_k(x)$ is a complex-valued eigenfunction and ω_k is a real eigenfrequency that represents a characteristic timescale associated with the eigenfunction. Further information about the methodology is presented in the Supporting Information S1 (Text S1).

There are several properties of the Koopman eigenfunctions that render them ideal for the description of quasiperiodic, recurrent climate phenomena like the MJO. First and foremost, the appearance of complex pairs of eigenfunctions in the Koopman spectrum directly indicates cyclic behavior, with eigenvalues corresponding to cycle periodicity. Although paired ("in quadrature") EOFs emerge in the standard RMM depiction of the MJO, as Roundy (2015) notes, it is possible that EOFs fail to identify cyclic behavior if modes reflect very different amounts of variance. In this regard, we note that the leading two EOFs of Pacific region interannual sea surface temperature variability exhibit unequal amounts of variance (>50% for mode 1 vs. ~10% for mode 2), which may be interpreted as reflecting distinct behavior (Ashok et al., 2007). On the other hand, these EOF modes appear to be captured by a single Koopman eigenfunction (Pike, 2022).

As F21 point out, two further properties inherent to Koopman eigenmodes, dynamical rectification and phase equivariance, may facilitate interpretation. Essentially, dynamical rectification refers to a representation of an oscillatory process in analogy to simple harmonic oscillation. In more detail, the time series $z_k(t) \equiv z_k(x(t))$ obtained
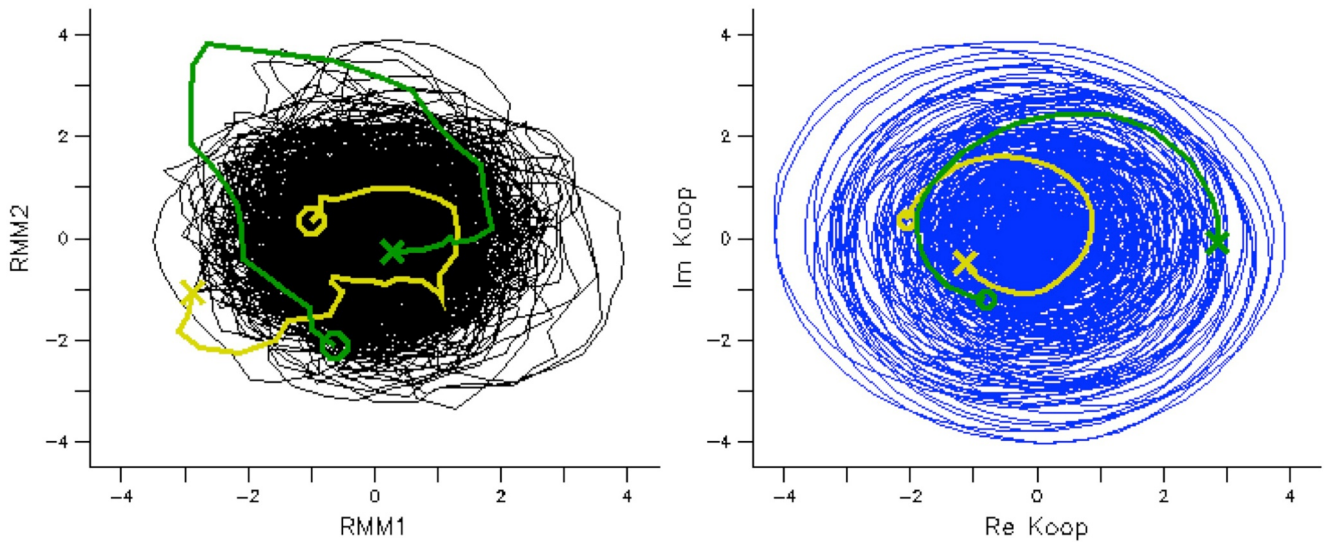
Figure 1. 2D phase spaces for the standard realtime multivariate Madden-Julian Oscillation (RMM) (left) and Koopman RMM (right). Two example Madden-Julian Oscillation events are highlighted: 16 October to 15 November 2011 (yellow) and 1–to 30 March 2015 (green), with the initial date denoted by a cross and the final date by a circle.

by sampling a Koopman eigenfunction z_k along the trajectory evolves as a complex phase with *fixed* oscillatory frequency ω_k, that is,

$$z_k(t) = z_k(0)e^{i\omega_k t} \tag{3}$$

This property holds even if the state-space process evolves at a variable frequency (see Figure 4 in F21 for an example). Even when the underlying flow Φ^t is chaotic, the leading Koopman eigenfunctions are adept at identifying approximately cyclical observables that remain correlated on timescales significantly longer than the Lyapunov timescale of the system. In such cases, Equation 3 holds approximately, and $z_k(t)$ may exhibit amplitude and/or phase modulation (e.g., F21, Figures 2 and 5). Phase equivariance relates to partitioning of the complex plane into angular sectors ("wedges") which identify phases of the cycle associated with z_k that evolve compatibly with the dynamics. For a partitioning of a cycle of frequency ω_k into S phases, evolving the samples belonging in a given phase over a time interval $2\pi/S\omega_k$ should consistently map them into the next phase. The fact that z_k evolves at a constant frequency ω_k means that phase equivariance can simply be obtained by choosing the wedges to have equal angular extent. This facilitates the construction of phase composites while aiding phase predictability.

Because we have discrete, rather than continuous, observations, we can only approximate the solution of the Koopman generator defined in the limit of Equation 2. We therefore follow an approach in which we first apply nonlinear Laplacian spectral analysis (NLSA) (Giannakis & Majda, 2012) to the input data. In this step, we construct and solve a kernel eigenvalue problem that approximates the eigenfunctions of the Laplace-Beltrami operator on the nonlinear manifold sampled by the data. The set of leading L NLSA eigenfunctions, which we denote as $\{\phi_1, \ldots, \phi_L\}$, can be thought of as nonlinear principal components which may be used to represent the generator V by an $L \times L$ matrix (Section 2 in Text S1 in Supporting Information S1). Solving the eigenvalue problem for this matrix yields the eigenfrequencies ω_k as the imaginary part of the eigenvalues of the generator matrix and the corresponding Koopman eigenfunctions as linear combinations of the NLSA basis functions, that is, $z_k = \sum_j c_{jk}\phi_j$, where the expansion coefficients c_{jk} are determined from the eigenvectors of the generator matrix (Section 3 in Text S1 in Supporting Information S1). We order the z_k in decreasing order of the real parts γ_k of the corresponding eigenvalues; by construction of our generator approximation, the γ_k are non-positive, which means that the eigenfunctions are arranged in decreasing order of decorrelation timescale under the dynamics induced by the approximate generator. Analogously to extended EOF (EEOF) analysis or singular spectrum analysis (SSA) (Ghil et al., 2002), NLSA operates on delay-embedded sequences of data snapshots (Section 1 in Text S1 in Supporting Information S1). This results in basis functions ϕ_j that span approximately invariant subspaces

under the Koopman operator (Giannakis, 2021), which may be exploited to mitigate numerical errors arising from approximation of the Koopman generator.

Several parameters are introduced in the numerical solution of the Koopman generator eigenvalue problem. Two of the notable parameters are embedding length and nearest neighbors. Embedding length refers to inputs comprising temporal sequences of data, that is, each available instance in a space time field of input data is expanded to a sequence equal to the embedding length. The underlying idea here is that the dynamics should induce some "regularity" that is captured by such time sequences, although in principle, results should be asymptotically independent of embedding length. The nearest neighbors parameter is applied to the NLSA kernel as an effective weighting factor to account for local similarity in inputs (see Section 2 in Text S1 in Supporting Information S1). A well-tuned kernel should balance between locality (rate of decay), which improves the quality of the corresponding eigenfunctions ϕ_j, and robustness against sampling errors (which decreases with increasing kernel locality at a fixed number of samples). To define parameters appropriate for solution of the Koopman generator on the RMM input, we leverage insights from the prior NLSA analysis of brightness temperature of Székely et al. (2016) and Tung et al. (2014). In particular, we choose an embedding length of 64 days and a nearest neighbors value of approximately 20% of the input samples. While there is some sensitivity to these parameter choices, similar qualitative behavior is captured over a broad range of parameter values. See Section 4 in Text S1 in Supporting Information S1 for further details on parameter selection and Table S1 in Supporting Information S1 for a listing of the parameter values employed in the Koopman MJO analysis described below.

We also consider application of EEOF analysis performed on sequences of 64 days. In contrast to conventional EOF analysis, which is insensitive to sequencing (i.e., the same spatial patterns would be obtained for an input dataset under permutation of the time sequencing), EEOFs inherit some information about the temporal evolution. Indeed, we include here some results of EEOF analysis as a bridge between the standard RMM results and those obtained from the Koopman generator.

We analyze here daily interpolated National Oceanic and Atmospheric Administration OLR (Liebmann & Smith, 1996) along with ERA5 (Hersbach et al., 2020) zonal winds at 200 and 850mb winds. All input data are first regridded to a common 2.5° × 2.5° grid and meridionally averaged over 15°S–15°N.

3. Results and Discussion

Among the first 20 RMM Koopman eigenfunctions, there appear four pairs with periods in the intraseasonal band (see Table S2 in Supporting Information S1). In what follows, we focus on the first pair of these intraseasonal eigenfunctions, which we will demonstrate possess characteristics consistent with the MJO. The eigenperiod for the leading intraseasonal mode is 49.8 days; using ERA5 OLR, the corresponding eigenfunction pair exhibits a slightly longer eigenperiod (52.5 days). Varying the length of the embedding window over a range of 45–180 days (not shown) indicates stability of the mode behavior, with the estimated eigenperiod varying by only ~2 days over this range of embedding lengths.

Figure 1 depicts the 2D phase space representation of RMM index, with RMM2 on the vertical axis against RMM1 on the horizontal axis. For comparison, we also present the phase space formed by the real and imaginary components of the Koopman RMM eigenfunction, where we have first rotated the eigenfunction through multiplication by a phase $e^{i\alpha}$, where the phase angle α is chosen such that covariance of the rotated Koopman RMM eigenfunction with respect to RMM1 and RMM2 (written in complex notation as RMM1 + iRMM2) is maximized. Further details on this procedure can be found in Text S2 in Supporting Information S1. In essence, the rotation aligns the real and imaginary components of the Koopman RMM eigenfunction with RMM1 and RMM2, respectively. Upon rotation, the real and imaginary components of the Koopman RMM mode exhibit a moderate, but statistically significant (according to a two-tailed Student t-test), correlation with RMM1 and RMM2 (Pearson r values of 0.57 and 0.62, respectively). Note that the (rotated) Koopman RMM1 and RMM2, like the standard RMM1 and RMM2, are normalized to unit variance. Two MJO events, one that occurred during the CINDY/DYNAMO field campaign from mid-October 2011 to mid-November 2011 (Kikuchi et al., 2018) and another with record setting RMM amplitude that occurred in March 2015 (Marshall et al., 2016), are highlighted.

As in the standard RMM, MJO propagation is represented by counterclockwise motion around the Koopman RMM 2D phase space. Qualitatively, the phase space for the Koopman RMM exhibits a smoother appearance than the standard RMM, which as we discuss further below may be partially attributed to the use of delay

embedding. For the two example MJO events, both the standard RMM and Koopman RMM phase progressions broadly agree, but the evolution of the standard RMM depiction is more irregular, with more significant radial motions evident. In the results that follow, we explore this phase space behavior more fully, but for now we note that the smoother phase space orbits inherent in the Koopman RMM may imply greater predictive skill. F21 recently demonstrated analogous behavior comparing a Koopman generator-based 2D phase space for ENSO to that of a conventional ENSO index (NINO3.4) plotted against itself with a lag corresponding to the characteristic timescale of the Koopman ENSO eigenmode.

Figure 2a illustrates the power spectra of the standard RMM1 with its optimally rotated Koopman RMM counterpart. While the power spectrum of RMM1 exhibits somewhat higher amplitude at intraseasonal frequencies, the intraseasonal peak is more pronounced in the corresponding Koopman mode. In other words, the Koopman mode more clearly distinguishes the intraseasonal behavior from other timescales. At higher frequencies, the Koopman mode appears to share much of the timescale separation relative to the standard RMM with the underlying NLSA eigenfunctions onto which it projects (not shown). However, at lower frequencies, the Koopman eigenfunctions exhibit lower amplitude and spectral power than their NLSA analogs, from which we conclude that the Koopman operator approach accounts for at least part of the clearer intraseasonal behavior evident in Figure 2a.

Amplitudes for the standard and Koopman RMM modes for a 5-year period spanning 1 January 2010 to 31 December 2015 are depicted in Figure 2b. The EEOF version of RMM is included here for comparison, specifically to assess how application of embedding in an EOF framework can impact amplitude smoothness. Like the Koopman RMM amplitude, the EEOF RMM amplitude is also smoother than the standard RMM, underscoring that time sequence inputs serve to filter resultant behavior. On the other hand, the Koopman RMM amplitude reflects a more seasonal amplitude modulation than is associated with either the standard or the EEOF RMM indices: in particular, Koopman RMM amplitudes are larger during boreal winter than during boreal summer. The seasonal amplitude modulation appears to depend on the choice of nearest neighbor parameter. Indeed, with larger nearest neighbor parameter values, which less strongly localize the NLSA kernel, the seasonality of the Koopman RMM is diminished. (For further discussion of this point, please see Section 4 in Text S1 in Supporting Information S1).

To further demonstrate that the Koopman RMM mode captures MJO characteristics, we consider the eigenfunctions projected onto the embedded input data. The resultant projections for the Koopman counterpart of RMM1, as a function of longitude and embedding time, are illustrated in Figure 3 for OLR, u850mb, and u200mb (filled contours). For comparison, we have also included projections based on lagged regression of RMM1 onto the input data (line contours).

For each field, eastward phase propagation is clearly evident, particularly over longitudes corresponding to the eastern Indian Ocean through western Pacific. Here, the estimated propagation speed is ~4.5 m s^{-1}, which is in agreement with previously reported values (Chen & Wang, 2020). While the signal in OLR (Figure 3c) is largely confined to the eastern Indian Ocean-western Pacific, the signal in both lower and upper level zonal winds (Figures 3a and 3b, respectively) is seen to extend further eastward. Interestingly, the estimated phase speed based on the zonal wind component (especially for 850 mb) increases moving into the Central and Eastern Pacific. Since the thermodynamic conditions over the Central and Eastern Pacific are less favorable to precipitating deep convection, convective coupling in the MJO is expected to weaken (Adames & Kim, 2016). With a reduction in latent heating, simple theories of equatorial wave propagation suggest increased phase speed, which is consistent with what is seen here.

For a perfectly coherent propagation, as in a mode characterized by sinusoidal oscillation, the 2D phase orbits would lie on a circle. To compare the coherence of the standard and Koopman RMM representations of the MJO, we compute the autocorrelations and cross-correlations of the components of the 2D phase space, for lags of up to 48 days. Figure 4 illustrates the resultant scatterplot of these lagged correlations, along with a reference unit circle of a sinusoidal oscillation. Note that each scatterplot consists of two branches, depending on whether the first component leads or lags the second.

While both the Koopman and standard RMM correlations decrease with increasing lag, as expected, the components of the Koopman RMM (blue) lose coherence more slowly than do RMM1 and RMM2 (black), that is, the Koopman component correlation curve remains closer to the reference unit circle. It is also worth noting that Koopman RMM component autocorrelations and cross-correlations tend to exceed those for the EEOF version of RMM (red), so it is not simply the smoothing introduced by embedding that enhances cycle coherence. Interestingly, the two branches of the EEOF are visibly asymmetric, meaning that one component is more strongly
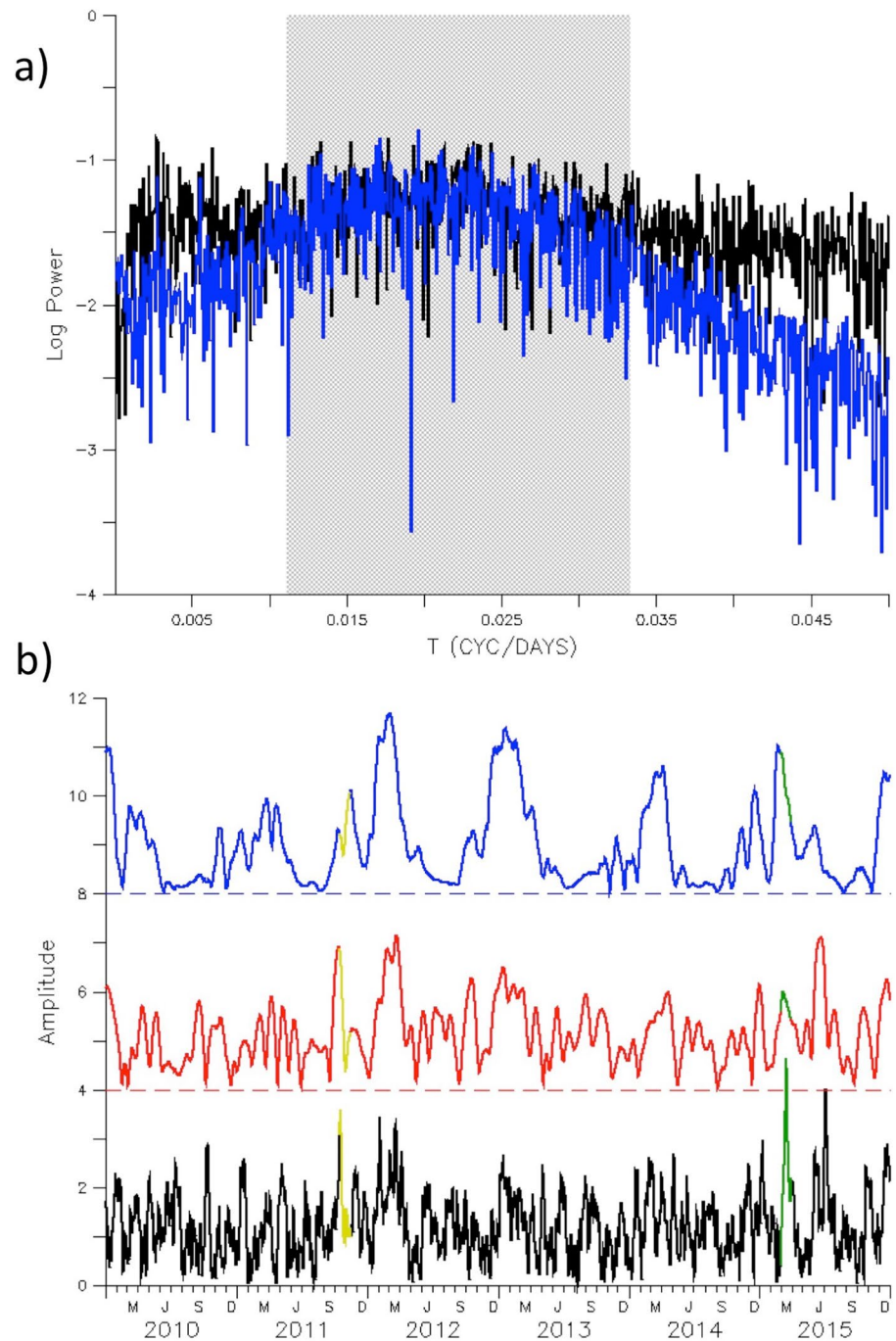
Figure 2. (a) Power spectral densities for the standard realtime multivariate Madden-Julian Oscillation (RMM) (black) and Koopman RMM (blue). The gray shading delimits the 30–90-day intraseasonal band. (b) Amplitudes for RMM (black), extended empirical orthogonal function (EEOF) RMM (red), and Koopman RMM (blue) over the period 1 January 2010 to 31 December 2015. The amplitudes are dimensionless and reflect normalization of the standard RMM components (or their EEOF and Koopman analogs) to unit variance, with offsets introduced for visual clarity (i.e., dashed lines denote the zeros of EEOF and Koopman RMM amplitudes). Amplitudes during two example Madden-Julian Oscillation events are highlighted: 16 October to 15 November 2011 (yellow) and 1–30 March 2015 (green).
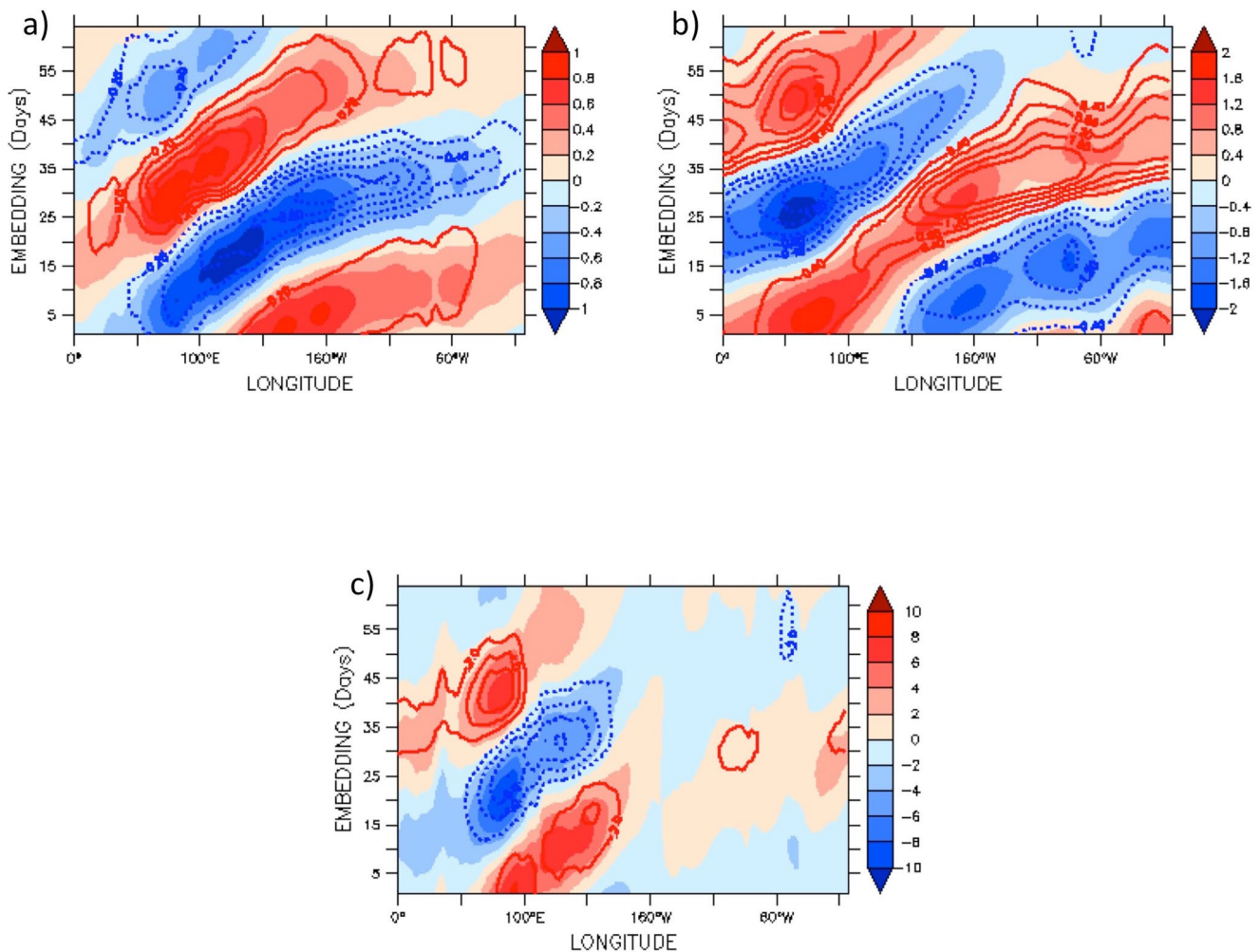
Figure 3. Projections of Koopman RMM1 (filled contours) onto (a) 850 mb zonal wind (in units of m s^{-1}), (b) 200 mb zonal wind (in units of m s^{-1}), and (c) outgoing longwave radiation (in units of W m^{-2}). Line contours are the same projections but for the standard RMM1.

autocorrelated than the other. Although the source of this asymmetry is unclear, the more symmetric Koopman behavior is consistent with the rectification introduced by Koopman, as discussed in Froyland et al. (2021).

For further insight into the potential relevance of our results to MJO predictability, we briefly compare the phase progression of the standard and Koopman RMM indices. Conventionally, MJO phases are reported in terms of index values ranging from 1 to 8, depending on where (in longitude) the active center of the MJO is located at a given time (Wheeler & Hendon, 2004); these 8 phases map onto 45° wedges in the 2D RMM phase space. Note that for a point in the RMM phase to be considered as part of an MJO event, it is subject to the condition that its amplitude equal or exceed unity.

As explained in Section 2, if the mean orbit over the phase space for an MJO event occurs over a time T, we would expect progression from a given MJO phase to the subsequent one to occur after ~T/8. In the analysis here, we first identify all points in the standard and Koopman RMM phase spaces at or above the event amplitude threshold, grouped by the phase index. We then determine where in the phase spaces these points are located 6 days later, that is, ~1/8 of the Koopman RMM eigenperiod. If a point continues to exceed the amplitude threshold and it is in the next phase, we assign a value of 1 to the initial point; otherwise, we assign a value of 0. Essentially, we are using the expected phase progression from an initial point to predict the phase space location 6 days later, subject to amplitude remaining above a threshold (i.e., the event persists after 6 days).

Following application of the method above, the binary index is averaged over all phases to yield a mean percentage of event progression from one phase to the next. For the standard RMM, on average, we find that ~38% of the
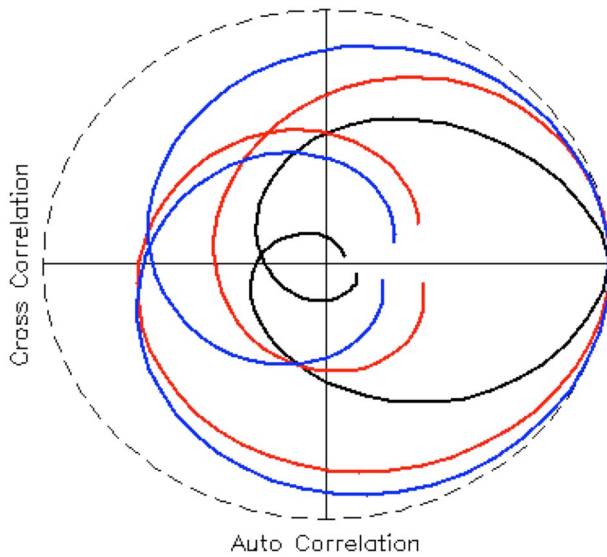
Figure 4. Scatterplots of auto versus cross-correlations for the components of the standard realtime multivariate MJO (RMM) (black), extended empirical orthogonal function RMM (red), and Koopman RMM (blue). Here, lags of up to 48 days are included. For reference, the dashed curve depicts a unit circle, which would correspond to the auto versus cross-correlation of a purely sinusoidal mode.

initial points "arrive" at the expected (target) phase after 6 days. By contrast, for the Koopman RMM, nearly twice the percentage of initial points (71%) appear in the target phase at day 6. While we suggest that this result implicates greater predictability inherent to the Koopman RMM, some important caveats should be mentioned. First, the selection of timescale here is based on a mean event cycle, but known event to event differences in MJO duration or propagation speed (Chen & Wang, 2020) would impact the phase progression for different events. By mathematical construction, as we have noted, the Koopman eigenfunctions exhibit more regularity in this regard. Our quick sensitivity assessment by varying the timescale for expected phase progression demonstrates that the Koopman-based RMM maintains a higher percentage of points reaching the target phase over 1–10 days compared to the standard RMM.

A further caveat is that use of the same event amplitude threshold for both the standard and Koopman RMM yields a larger number of initial points for the standard RMM. This arises in part as a consequence of the seasonal amplitude modulation of the Koopman eigenfunctions, which leads to a more leptokurtic distribution of values relative to RMM1/RMM2 (which are effectively Gaussian-distributed). However, variation of the event threshold, for example, by using a higher value for the standard RMM and/or a lower value for Koopman, does not appear to alter the results qualitatively, as the mean progression of Koopman RMM to the target phase continues to exceed the standard RMM over a wide range of thresholds.

4. Summary and Conclusions

In this study, we illustrate the application of an operator-theoretic formalism, the Koopman operator (or its infinitesimal generator), as a basis for obtaining a representation of the MJO analogous to the widely-used, EOF-based RMM. Our results demonstrate that the Koopman RMM manifests comparable behavior to the standard RMM in terms of MJO event identification, propagation, and large-scale spatial structure. Relative to the standard RMM, the Koopman RMM captures a more seasonally-modulated MJO amplitude peaking in boreal winter. Numerically, this seasonality appears to largely depend on the locality of the kernel employed for operator approximation (achieved here via nearest neighbor truncation), which aides in constraining events based on similarity.

Like the standard RMM, the Koopman RMM can be updated in (near) real-time. That is, continuously updated sequences of daily anomalies of OLR and upper and lower level zonal winds equal to the embedding length can be projected onto the optimally rotated spatial patterns (as in Figure 3) to obtain index values for Koopman RMM1 and RMM2. For example, for an embedding length of 64 days, updating for a 1-day forecast would require the preceding 63 days (Sabeerali et al., 2017). In principle, it is possible to employ a continuously updated Koopman RMM as a diagnostic for operational model forecast evaluation (Gottschalck et al., 2010); however, a long embedding length means that the index values will largely reflect phenomena that decorrelate on timescales comparable to the embedding window, and thus may not well represent daily-scale events. On the other hand, the Koopman RMM yields a more coherent and persistent MJO depiction compared to the standard RMM, or for that matter, a smooth (EEOF) version thereof. In this sense, there may be greater predictability inherent to the MJO than is inferred by the standard RMM approach, although whether such predictability could be realized as enhanced forecast skill warrants further exploration. Finally, one potentially desirable feature of the Koopman approach is that it yields a single explicit timescale (an eigenperiod) that we suggest could be useful as simple benchmarking metric for multimodel evaluation of MJO simulation. To the extent that this timescale varies within an ensemble of models, it may have some utility as a basis for grouping models to aide in evaluation of process or parameter sensitivity or diagnosis of model bias (Waliser et al., 2009).

Data Availability Statement

The data analyzed in this study may be obtained through the following links: https://www.psl.noaa.gov/data/gridded/data.olrcdr.interp.html, https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview.

References

Adames, Á. F., & Kim, D. (2016). The MJO as a dispersive, convectively coupled moisture wave: Theory and observations. *Journal of the Atmospheric Sciences*, *73*(3), 913–941. https://doi.org/10.1175/jas-d-15-0170.1

Ashok, K., Behera, S. K., Rao, S. A., Weng, H., & Yamagata, T. (2007). El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research*, *112*(C11), C11007. https://doi.org/10.1029/2006JC003798

Berry, T., Giannakis, D., & Harlim, J. (2020). Bridging data science and dynamical systems theory. *Notices of the American Mathematical Society*, *67*(9), 1336–1349. https://doi.org/10.1090/noti2151

Brunton, S. L., Budisic, M., Kaiser, J., & Kutz, J. N. (2022). Modern Koopman theory for dynamical systems. *SIAM Review*, *64*(2), 229–340. https://doi.org/10.1137/21M1401243

Chen, G., & Wang, B. (2020). Circulation factors determining the propagation speed of the Madden–Julian Oscillation. *Journal of Climate*, *33*(8), 3367–3380. https://doi.org/10.1175/jcli-d-19-0661.1

Dellnitz, M., & Junge, O. (1999). On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis*, *36*(2), 491–515. https://doi.org/10.1137/S0036142996313002

Froyland, G., Giannakis, D., Lintner, B. R., Pike, M., & Slawinska, J. (2021). Spectral analysis of climate dynamics with operator-theoretic approaches. *Nature Communications*, *12*(1), 6570. https://doi.org/10.1038/s41467-021-26357-x

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., et al. (2002). Advanced spectral methods for climatic time series. *Reviews of Geophysics*, *40*, 1003. https://doi.org/10.1029/2000RG000092

Giannakis, D. (2021). Delay-coordinate maps, coherence, and approximate spectra of evolution operators. *Research in the Mathematical Sciences*, *8*(1), 8. https://doi.org/10.1007/s40687-020-00239-y

Giannakis, D., & Majda, A. J. (2012). Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proceedings of the National Academy of Sciences*, *198*(7), 2222–2227. https://doi.org/10.1073/pnas.1118984109

Giannakis, D., Slawinska, J., & Zhao, Z. (2015). Spatiotemporal feature extraction with data-driven Koopman operators. *Journal of Machine Learning Research Proceedings*, *44*, 103–115.

Gottschalck, J., Wheeler, M., Weickmann, K., Vitart, F., Savage, N., Lin, H., et al. (2010). A framework for assessing operational Madden–Julian oscillation forecasts. *Bulletin America Meteorology Society*, *91*(9), 1247–1258. https://doi.org/10.1175/2010BAMS2816.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hogg, J., Fonoberova, M., & Mezić, I. (2020). Exponentially decaying modes and long-term prediction of sea ice concentration using Koopman mode decomposition. *Scientific Reports*, *10*(1), 16313. https://doi.org/10.1038/s41598-020-73211-z

Jiang, X., Adames, Á. F., Kim, D., Maloney, E. D., Lin, H., Kim, H., et al. (2020). Fifty years of research on the Madden-Julian Oscillation: Recent progress, challenges, and perspectives. *J. Geophys. Res.-Atmos.*, *125*(17), e2019JD030911. https://doi.org/10.1029/2019JD030911

Kikuchi, K., Kiladis, G. N., Dias, J., & Nasuno, T. (2018). Convectively coupled equatorial waves within the MJO during CINDY/DYNAMO: Slow Kelvin waves as building blocks. *Climate Dynamics*, *50*(11–12), 4211–4230. https://doi.org/10.1007/s00382-017-3869-5

Koopman, B. O. (1931). Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, *17*(5), 315–318. https://doi.org/10.1073/pnas.17.5.315

Koopman, B. O., & von Neumann, J. (1932). Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences*, *18*(3), 255–263. https://doi.org/10.1073/pnas.18.3.255

Liebmann, B., & Smith, C. A. (1996). Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society*, *77*, 1275–1277.

Madden, R. A. (1986). Seasonal variations of the 40–50 day oscillation in the tropics. *Journal of the Atmospheric Sciences*, *43*(24), 3138–3158. https://doi.org/10.1175/1520-0469(1986)043<3138:svotdo>2.0.co;2

Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, *28*(5), 702–708. https://doi.org/10.1175/1520-0469(1971)028<0702:doadoi>2.0.co;2

Marshall, A. G., Hendon, H. H., & Wang, G. (2016). On the role of anomalous ocean surface temperatures for promoting the record Madden-Julian Oscillation in March 2015. *Geophysical Research Letters*, *43*(1), 472–481. https://doi.org/10.1002/2015GL066984

Mezić, I. (2005). Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, *41*(1–3), 309–325. https://doi.org/10.1007/s11071-005-2824-x

Navarra, A., Tribbia, J., & Klus, S. (2021). Estimation of Koopman transfer operators for the equatorial Pacific SST. *Journal of the Atmospheric Sciences*, *78*(4), 1227–1244. https://doi.org/10.1175/JAS-D-20-0136.1

Pike, M. (2022). *Exploring the climate variability of the tropical indo-pacific via advanced data sorting and signal processing techniques*. (Ph.D. dissertation) (139 pp.). Rutgers, The State University of New Jersey.

Roundy, P. E. (2015). On the interpretation of EOF analysis of ENSO, atmospheric Kelvin waves, and the MJO. *Journal of Climate*, *28*(3), 1148–1165. https://doi.org/10.1175/jcli-d-14-00398.1

Sabeerali, C. T., Ajayamohan, R. S., Giannakis, D., & Majda, A. J. (2017). Extraction and prediction of indices for monsoon intraseasonal oscillations: An approach based on nonlinear Laplacian spectral analysis. *Climate Dynamics*, *49*(9–10), 3031–3050. https://doi.org/10.1007/s00382-016-3491-y

Schreck, C. J., III, Shi, L., Kossin, J. P., & Bates, J. J. (2013). Identifying the MJO, equatorial waves, and their impacts using 32 years of HIRS upper-tropospheric water vapor. *Journal of Climate*, *26*(4), 1418–1431. https://doi.org/10.1175/jcli-d-12-00034.1

Székely, E., Giannakis, D., & Majda, A. J. (2016). Extraction and predictability of coherent intraseasonal signals in infrared brightness temperature data. *Climate Dynamics*, *46*(5–6), 1473–1502. https://doi.org/10.1007/s00382-015-2658-2

Tung, W.-W., Giannakis, D., & Majda, A. J. (2014). Symmetric and antisymmetric convection signals in the Madden–Julian oscillation. Part I: Basic modes in infrared brightness temperature. *Journal of the Atmospheric Sciences*, *71*(9), 3302–3326. https://doi.org/10.1175/JAS-D-13-0122.1

Waliser, D., Sperber, K., Hendon, H., Kim, D., Maloney, E., Wheeler, M., et al. (2009). MJO simulation diagnostics. *Journal of Climate*, *22*, 3006–3030. https://doi.org/10.1175/2008JCLI2731.1

Wang, X., Slawinska, J., & Giannakis, D. (2020). Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports*, *10*(1), 2636. https://doi.org/10.1038/s41598-020-59128-7

Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, *132*(8), 1917–1932. https://doi.org/10.1175/1520-0493(2004)132<1917:aarmmi>2.0.co;2

Zhang, C. (2005). Madden-Julian Oscillation. *Reviews of Geophysics*, *43*(2), RG2003. https://doi.org/10.1029/2004RG000158