Uncertainty Quantification of the Madden–Julian Oscillation with Gaussian Processes

Haoyuan Chen[1] **Emil Constantinescu**[2] **Vishwas Rao**[2] **Cristiana Stan**[3]
[1]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843
[2]Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439
[3]Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, VA 22030
chenhaoyuan2018@tamu.edu
{emconsta,vhebbur}@anl.gov
cstan@gmu.edu

Abstract

The Madden–Julian Oscillation (MJO) is an influential climate phenomenon that plays a vital role in modulating global weather patterns. In spite of the improvement in MJO predictions made by machine learning algorithms, such as neural networks, most of them cannot provide the uncertainty levels in the MJO forecasts directly. To address this problem, we develop a nonparametric strategy based on Gaussian process (GP) models. We calibrate GPs using empirical correlations. Furthermore, we propose a posteriori covariance correction that extends the probabilistic coverage by more than three weeks.

1 Introduction

The Madden–Julian Oscillation (MJO), first discovered by [1], is the dominant mode of intraseasonal variability (20–100 days) of the tropical atmosphere. The MJO is described as one of the engines of the climate system [2] powering the atmospheric circulation in the tropics and extratropics. The atmospheric teleconnections to the extratropics represent one of the major sources of predictability on the subseasonal-to-seasonal (S2S) timescale [3]. S2S forecasts have been identified to be more broadly useful to climate decision makers than are climate projections [4]. The MJO forms in the equatorial Indian Ocean and propagates eastward along the equator alternating between phases of active and suppressed rainfall. In 2004, Wheeler and Hendon [5] introduced the Real-time Multivariate MJO (RMM) index for describing the MJO characteristics including its magnitude and phase. The RMM index consists of a pair (in quadrature) of principal component (PC) time series known as RMM1 and RMM2 ($\text{RMM} = \sqrt{\text{RMM1}^2 + \text{RMM2}^2}$). RMM1 and RMM2 are the first two PCs of combined outgoing longwave radiation and zonal winds in the lower (850 hPa) and upper (200 hPa) troposphere averaged between 15S and 15N.

The MJO can alter weather patterns over the Earth, influence the El Niño–Southern Oscillation events, and affect the tropics and mid-latitudes, such as monsoons, rainfalls, extreme heat events, and tropical cyclogenesis. The accurate forecast of MJO can assist in agriculture, disaster response, and economic planning. In recent decades, many efforts have been made to improve MJO forecasting, but most methods rely on dynamical models [6, 7, 8] or artificial neural networks (ANNs) [9, 10, 11, 12, 13, 14] that have limited ability in quantifying the uncertainty of predictions. To fill this gap, we introduce a probabilistic model for predicting the amplitude and phase of the MJO.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

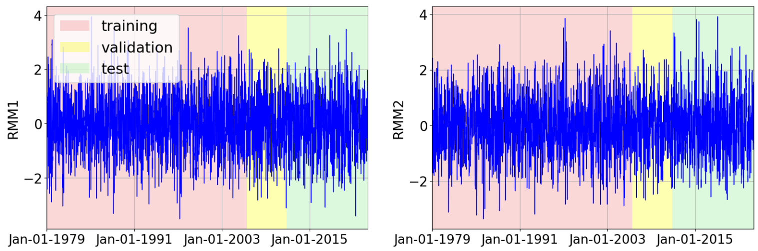

Figure 1: Time series of RMMs (Jan-01-1979 to Dec-31-2022).

In this paper we employ *Gaussian process* (GP) models with empirical correlations to predict and quantify the uncertainty of the RMMs. In addition to quantifying the uncertainties, we demonstrate that our nonparametric model has a better prediction skill than the ANN models for the first five forecast lead days. The results of our experiment are also available on the project website `http://gp-mjo.github.io/`.

2 Data

The daily MJO RMM index dataset [1] used in the study is provided by the Bureau of Meteorology. RMM1 and RMM2 values are available from June 1, 1974, to the most recent date. Because of the missing values before 1979, we select the dataset from January 1, 1979, to December 31, 2022, for our study (Figure 1). The dataset is divided into three subsets: the training set is used to determine the parameters of the prediction and corresponding variance, the validation set is used to obtain the corrected variance with increasing lags, and the test set is used to verify the results.

3 Method

We defer the background material on forecasting and GPs to Appendix A.

Model We denote the values of RMM1 and RMM2 on the tth day by $z_t^{(1)}$ and $z_t^{(2)}$, respectively. Suppose $[z_t^{(1)}, z_t^{(2)}]^\top$ is a bivariate GP defined as in (9). Our training data $\mathcal{D}^{(1:2)} = \{\mathbf{X}^{(1:2)}, \mathbf{y}^{(1:2)}\}$ is of size n with lag L and is denoted by

$$\mathbf{X}^{(1:2)} := \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} \in \mathbb{R}^{2L \times n}, \quad \mathbf{y}^{(1:2)} := \begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix} \in \mathbb{R}^{2 \times n}, \tag{1}$$

where $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, $\mathbf{y}^{(1)}$, $\mathbf{y}^{(2)}$ are the matrices and vectors defined in equation (13). The objective is to model the probability distribution of $[z_{t^*}^{(1)}, z_{t^*}^{(2)}]^\top$ over the next τ days conditioned on the past L days. Inspired by the cross- and auto-correlations of the RMMs shown in Figure 2, we first use the cubic spline interpolation to simulate $\mathrm{Cov}[\mathbf{y}^{(1:2)}, \mathbf{X}^{(1:2)}]$ and $\mathrm{Cov}[\mathbf{X}^{(1:2)}, \mathbf{X}^{(1:2)}]$. After obtaining the empirical mean $\mathbb{E}[\mathbf{y}^{(1:2)}]$ and covariances $\mathrm{Cov}[\mathbf{y}^{(1:2)}, \mathbf{X}^{(1:2)}]$, $\mathrm{Cov}[\mathbf{X}^{(1:2)}, \mathbf{X}^{(1:2)}]$ from the training data $\mathcal{D}^{(1:2)}$, we can apply them to equations (11) and (12) to obtain the prediction $\boldsymbol{\mu}_{t^*}$ and corresponding covariance $\mathbf{K}_{t^*}$ at time t^*. Next, we add the prediction $[\hat{z}_{t^*}^{(1)}, \hat{z}_{t^*}^{(2)}]^\top$ to the predictors used for the next step to obtain the predictive distribution of $[\hat{z}_{t^*+1}^{(1)}, \hat{z}_{t^*+1}^{(2)}]^\top$ by using equations (11) and (12) again. We repeat the steps until we reach the time $t^* - 1 + \tau$ (see Figure 6 and Appendix A.2 for more details).

Correction to the predictive variance Note that the covariance $\mathbf{K}_{t^*}$ in equation (12) is related only to the value of lag L, which is one in our study, and is unrelated to the lead time τ or the predictor values. However, as we predict for longer lead times, the predictive variance should increase because observations are replaced by predicted values. To correct the covariance, for each lead time we use a validation set of size n_v with lag L to compute the averaged bias between the posterior mean and true

[1] `http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt`

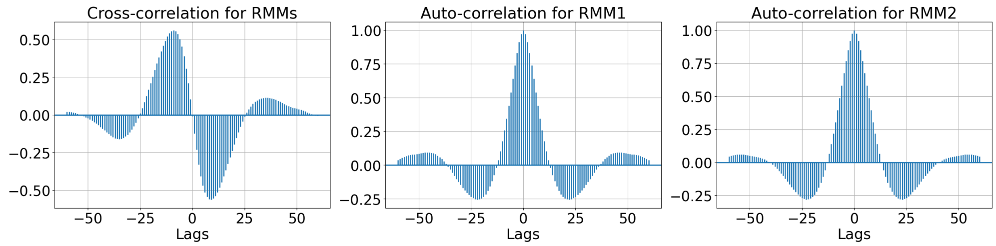

Figure 2: Cross-correlations and auto-correlations of RMMs with maximum lag = 60 days.

observations. Hence, the corrected variance $\tilde{V}_*^{(j)}(\tau)$ is given by

$$\begin{aligned} \tilde{V}_*^{(j)}(\tau) :&= \mathrm{Var}[z_{t^*}^{(j)}(\tau)] \approx \mathrm{Var}[\hat{z}_{t^*}^{(j)}(\tau)] + \mathrm{Bias}\left(\hat{z}_{t^*}^{(j)}(\tau), z_{t^*}^{(j)}(\tau)\right)^2, \\ &\approx \mathbf{K}_{t^*}[j,j] + \frac{1}{n_v} \sum_{t=1}^{n_v} \left(\hat{z}_t^{(j)}(\tau) - z_t^{(j)}(\tau)\right)^2, \end{aligned} \tag{2}$$

where $\hat{z}_{t^*}^{(j)}(\tau)$ is the predicted value for lead time τ obtained by the above iteration, $z_{t^*}^{(j)}(\tau)$ is the corresponding true observation, and $\mathbf{K}_{t^*}[j,j]$ is the $[j,j]$th entry of the covariance matrix $\mathbf{K}_{t^*}$, $j = 1, 2$. Then we scale the $\mathbf{K}_{t^*}$ to the corrected covariance $\tilde{\mathbf{K}}_{t^*}(\tau)$ for lead time τ in (14) by using the variances $\{\tilde{V}_*^{(j)}(\tau)\}_{j=1}^2$.

Confidence region The $(1-\alpha)$ confidence region of the p-variate normal distribution is a hyperellipsoid bounded by $\chi_p^2(\alpha)$, the chi-square distribution with p degrees of freedom at the level α [15]. Lemmas A.1 and A.2 aid in constructing a confidence region for the prediction $[\hat{z}_{t^*}^{(1)}(\tau), \hat{z}_{t^*}^{(2)}(\tau)]^\top$ at lead time τ, where $[\hat{z}_{t^*}^{(1)}(\tau), \hat{z}_{t^*}^{(2)}(\tau)]^\top \sim \mathcal{N}(\boldsymbol{\mu}_{t^*}, \tilde{\mathbf{K}}_{t^*}(\tau))$ after updating the covariance.

4 Experiments and results

Prediction skill The performance of the model is measured by the bivariate correlation coefficient (COR) and root mean squared error (RMSE) defined in equations (15) and (16). We also analyze the phase error E_ϕ and the amplitude error E_A of RMMs defined in equations (17) and (18) of our model. The evaluation is conducted for two values of the lag, $L = 40, 60$, size of the training set $n = 10000$, size of the validation set $n_v = 2000$, number of predictions for computing the errors $n_p = 365 \times 5 = 1825$, and forecast lead time $\tau = 1, 2, \ldots, 60$.

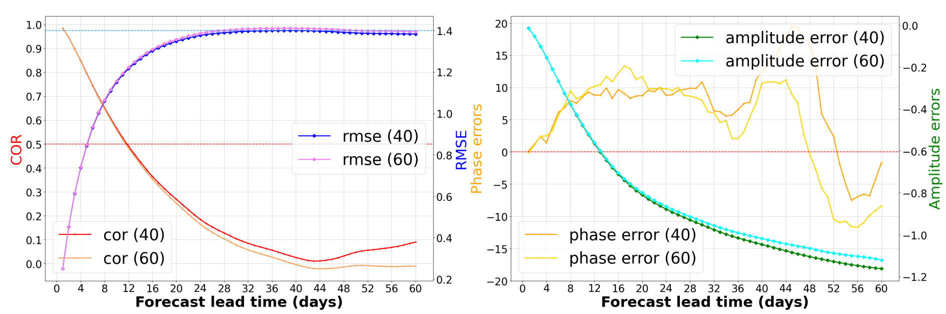

Figure 3: Prediction skill quantifiers and errors of the model. *Left:* COR and RMSE over 1,825 predictions with lag $L = 40$, 60, respectively. *Right:* Phase error (degrees) and amplitude error over 1,825 predictions with lag $L = 40$, 60, respectively.

The results of the prediction skill and errors are shown in Figure 3. The values COR $= 0.5$ and RMSE $= 1.4$ are the commonly used skill thresholds for a climatological forecast [16]. In this figure we see that our model has a prediction skill of 12 days for both lag $L = 40$ and $L = 60$ with threshold COR $= 0.5$. Regarding the RMSE, the prediction skill is longer than 60 days for $L = 40$ with threshold RMSE $= 1.4$, while the RMSE for lag $L = 60$ crosses the threshold value on day 28 and fluctuates

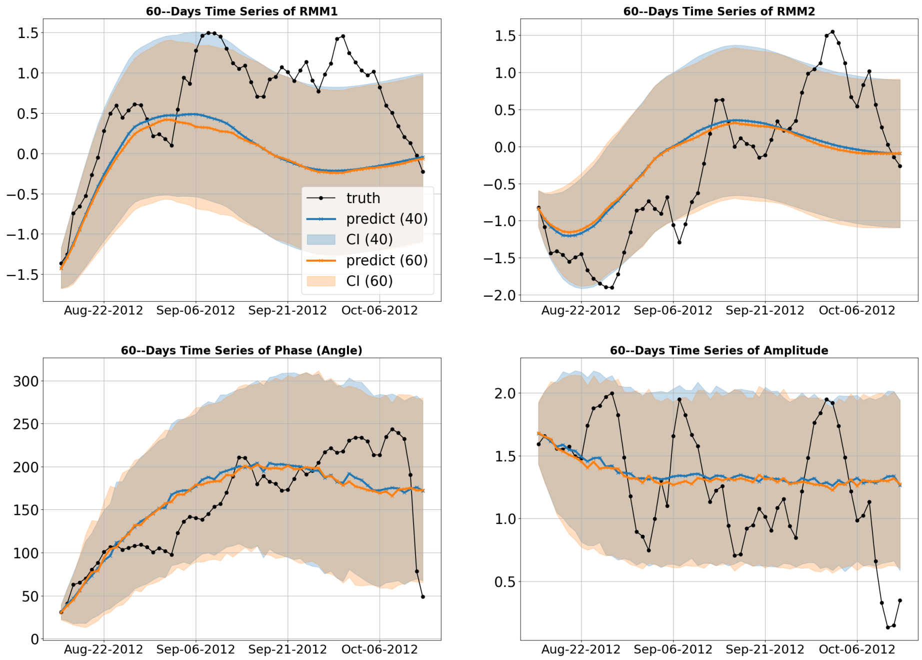

Figure 4: 60-day time series of MJO for Aug–15–2012 to Oct–13–2012 for lag $L = 40, 60$. We denote observations (truth) by black dots; denote predictions for lag $L = 40$ and $L = 60$ by blue cross and orange cross, respectively; and denote $\pm\sigma$ confidence interval by shading. *Top left*: RMM1. *Top right*: RMM2. *Bottom left*: Phase (angle in the degrees $0° - 360°$). *Bottom right*: Amplitude.

around RMSE $= 1.4$ throughout the following days. The fast decline of COR is due to the fact that we use the empirical correlations from historical RMMs of large size in our model. Specifically, when the forecast lead time increases, the predicted RMMs will become smaller and smoother due to the empirical correlations over a long period of time, giving rise to the smaller variations of RMMs than the true observations and therefore a lower COR. The small value of the predicted RMMs also accounts for the tiny changes in RMSE after day 24 of the forecast lead time. As for the phase error (the angle of RMMs in degrees), we observe that most phase errors are positive, indicating a faster propagation relative to the observations. For the amplitude errors, we note that all of them are negative. Because of the smaller values of the predicted RMMs with forecast lead time increasing, the amplitude is underestimated, resulting in negative amplitude errors.

Uncertainty quantification Here we pick one sample (Aug–15–2012 to Oct–13–2012) out of $n_p = 1825$ predictions with $\tau = 60$ forecast lead days to present the uncertainty quantification of the predicted MJO in Figure 4. This example shows that predictions capture the general trend seen in observations. The $\pm\sigma$ confidence intervals grow as the forecast lead time increases and cover most of the range of observations. To obtain the complete picture of MJO prediction, we summarize results in Figure 4 and an MJO phase diagram as shown in Figure 8; the animated plots are on the project website `https://gp-mjo.github.io/`. See Appendix B.2 for more details.

5 Conclusion

In this work we develop a probabilistic model to predict the MJO and quantify the prediction uncertainty by using GPs with empirical correlations. Our model demonstrates good performance on the daily RMM index dataset from January 1, 1979, to December 31, 2022. The model is highly effective and stable because it does not require optimizing the hyperparameters and is insensitive to lags of the predictors. Future work will focus on the effects of seasonal factors in our model.

Acknowledgments and Disclosure of Funding

Argonne National Laboratory's work was supported by the U.S. Department of Energy, Office of Science under contract DE-AC02-06CH11357.

References

[1] Roland A Madden and Paul R Julian. Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *Journal of Atmospheric Sciences*, 28(5):702–708, 1971.

[2] Cristiana Stan, Cheng Zheng, Edmund Kar-Man Chang, Daniela IV Domeisen, Chaim I Garfinkel, Andrea M Jenney, Hyemi Kim, Young-Kwon Lim, Hai Lin, Andrew Robertson, et al. Advances in the prediction of MJO teleconnections in the S2S forecast systems. *Bulletin of the American Meteorological Society*, 103(6):E1426–E1447, 2022.

[3] Cristiana Stan, David M Straus, Jorgen S Frederiksen, Hai Lin, Eric D Maloney, and Courtney Schumacher. Review of tropical-extratropical teleconnections on intraseasonal time scales. *Reviews of Geophysics*, 55(4):902–937, 2017.

[4] Catherine Vaughan, Lawrence Buja, Andrew Kruczkiewicz, and Lisa Goddard. Identifying research priorities to advance climate services. *Climate Services*, 4:65–74, 2016.

[5] Matthew C Wheeler and Harry H Hendon. An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly weather review*, 132(8):1917–1932, 2004.

[6] Frédéric Vitart. Madden–Julian Oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, 143(706):2210–2220, 2017.

[7] Yuna Lim, Seok-Woo Son, and Daehyun Kim. MJO prediction skill of the subseasonal-to-seasonal prediction models. *Journal of Climate*, 31(10):4075–4094, 2018.

[8] Hyemi Kim, Matthew A Janiga, and Kathy Pegion. MJO propagation processes and mean biases in the SubX and S2S reforecasts1. *Journal of Geophysical Research: Atmospheres*, 124(16):9314–9331, 2019.

[9] Benjamin A Toms, Karthik Kashinath, Da Yang, et al. Testing the reliability of interpretable neural networks in geoscience using the Madden-Julian Oscillation. *arXiv preprint arXiv:1902.04621*, 2019.

[10] Panini Dasgupta, Abirlal Metya, CV Naidu, Manmeet Singh, and MK Roxy. Exploring the long-term changes in the Madden Julian Oscillation using machine learning. *Scientific Reports*, 10(1):18567, 2020.

[11] H Kim, YG Ham, YS Joo, and SW Son. Deep learning for bias correction of MJO prediction. *Nature Communications*, 12(1):3087, 2021.

[12] Riccardo Silini, Marcelo Barreiro, and Cristina Masoller. Machine learning prediction of the Madden–Julian Oscillation. *npj Climate and Atmospheric Science*, 4(1):57, 2021.

[13] Riccardo Silini, Sebastian Lerch, Nikolaos Mastrantonas, Holger Kantz, Marcelo Barreiro, and Cristina Masoller. Improving the prediction of the Madden–Julian Oscillation of the ECMWF model by post-processing. *Earth System Dynamics*, 13(3):1157–1165, 2022.

[14] Antoine Delaunay and Hannah M Christensen. Interpretable deep learning for probabilistic MJO prediction. *Geophysical Research Letters*, 49(16):e2022GL098566, 2022.

[15] Richard Arnold Johnson, Dean W Wichern, et al. *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 2002.

[16] Harun A Rashid, Harry H Hendon, Matthew C Wheeler, and Oscar Alves. Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. *Climate Dynamics*, 36:649–661, 2011.

[17] Jan Dutton. What is the Madden–Julian Oscillation. `https://www.worldclimateservice.com/2021/09/20/madden-julian-oscillation/`, 2021.

[18] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. *Advances in Neural Information Processing Systems*, 31, 2018.

[19] Yuyang Wang, Alex Smola, Danielle Maddix, Jan Gasthaus, Dean Foster, and Tim Januschowski. Deep factors for forecasting. In *International conference on machine learning*, pages 6607–6617. PMLR, 2019.

[20] Andreas S Weigend. *Time series prediction: forecasting the future and understanding the past*. Routledge, 2018.

[21] Haibin Cheng, Pang-Ning Tan, Jing Gao, and Jerry Scripps. Multistep-ahead time series prediction. In *Advances in Knowledge Discovery and Data Mining: 10th Pacific-Asia Conference, PAKDD 2006, Singapore, April 9-12, 2006. Proceedings 10*, pages 765–774. Springer, 2006.

[22] Antti Sorjamaa, Jin Hao, Nima Reyhani, Yongnan Ji, and Amaury Lendasse. Methodology for long-term prediction of time series. *Neurocomputing*, 70(16-18):2861–2869, 2007.

[23] Larry R Medsker and LC Jain. Recurrent neural networks. *Design and Applications*, 5(64-67):2, 2001.

[24] Inés M Galván and Pedro Isasi. Multi-step learning rule for recurrent neural models: an application to time series forecasting. *Neural Processing Letters*, 13:115–133, 2001.

[25] Liu Yunpeng, Hou Di, Bao Junpeng, and Qi Yong. Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network. In *2017 14th web information systems and applications conference (WISA)*, pages 305–310. IEEE, 2017.

[26] Lawrence Rabiner and Biinghwang Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1):4–16, 1986.

[27] Alessandro Rossi and Giampiero M Gallo. Volatility estimation via hidden Markov models. *Journal of Empirical Finance*, 13(2):203–230, 2006.

[28] Adriana Horelu, Catalin Leordeanu, Elena Apostol, Dan Huru, Mariana Mocanu, and Valentin Cristea. Forecasting techniques for time series from sensor data. In *2015 17th international symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 261–264. IEEE, 2015.

[29] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT Press Cambridge, MA, 2006.

A Background

A.1 MJO

The strength and position of the MJO are usually defined by two indices known as RMM1 and RMM2. They are derived as the dominant PCs of the multivariate EOF analysis of the satellite measurement of outgoing longwave radiation and the strength of the east-west component of the wind field at 850 mb and 200 mb [17]. The position and strength of the MJO are usually presented as a point on a Wheeler–Hendon diagram (Figure 5); within the center circle, the MJO is considered inactive/weak. The diagram is split into 8 phases, which represent different regions of the globe.

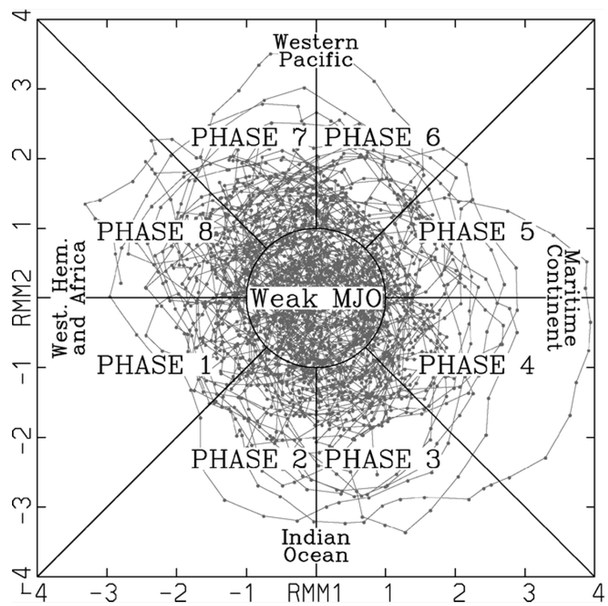

Figure 5: (RMM1, RMM2) phase space points for all available days in DJF season from 1974 to 2003 [5].

A.2 Forecasting

Probabilistic forecasting problem In the general probabilistic forecasting problem [18, 19], we usually denote M univariate time series by $\{z_{1:T_j}^{(j)}\}_{j=1}^{M}$, where $z_{1:T_j}^{(j)} :\ = (z_1^{(j)}, z_2^{(j)}, \ldots, z_{T_j}^{(j)})$ is the jth time series and $z_t^{(j)}$ is the value of the jth time series at time t, $1 \le t \le T_j$. Our goal is to model the distribution of $z_{T_j+1:T_j+\tau}^{(j)}$ at the next τ time conditioned on the past:

$$p(\, z_{T_j+1:T_j+\tau}^{(j)} \Big| \, z_{1:T_j}^{(j)}; \Theta), \tag{3}$$

where Θ is the set of the learnable parameters shared by all M time series.

For the MJO data, $z_t^{(1)}$ represents the RMM1 index on day t, $z_t^{(2)}$ represents the RMM2 index on day t, and $M = 2$. We set the same lag L for both RMM1 and RMM2, namely, $T_1 = T_2 = L$. The goal of this work is to obtain the predictive distribution of the vector $[z_t^{(1)}, z_t^{(2)}]^\top$ at the next τ time conditioned on the past L days:

$$p\Bigg(\begin{bmatrix} z_{L+1:L+\tau}^{(1)} \\ z_{L+1:L+\tau}^{(2)} \end{bmatrix} \Bigg| \begin{bmatrix} z_{1:L}^{(1)} \\ z_{1:L}^{(2)} \end{bmatrix}; \Theta \Bigg). \tag{4}$$

Multistep time series forecasting The objective of multistep time series forecasting [20, 21, 22] is to predict M-variate time series at the next τ time $\{z_{T_j+1:T_j+\tau}^{(j)}\}_{j=1}^{M}$ given $\{z_{1:T_j}^{(j)}\}_{j=1}^{M}$, where $\tau > 1$.

A multistep prediction is typically carried out using the iterative method. In this technique, the values computed for each step ahead are sent to the next step as inputs. The iterative method can be written in the autoregressive model as follows:

$$\begin{bmatrix} z_t^{(1)} \\ \vdots \\ z_t^{(M)} \end{bmatrix} = \begin{bmatrix} f_1(z_{t-T_1:t-1}^{(1)}) \\ \vdots \\ f_M(z_{t-T_M:t-1}^{(M)}) \end{bmatrix}, \tag{5}$$

where $f_1, \dots, f_M$ are random functions. After the learning process, the predicted values at the next τ time are given by

$$\hat{z}_{t+\tau-1}^{(j)} = \begin{cases} f_j(z_{t-T_j:t-1}^{(j)}) & \text{if } \tau = 1 \\ f_j(z_{t-T_j-1+\tau:t-1}^{(j)}, \hat{z}_{t:t-2+\tau}^{(j)}) & \text{if } \tau = 2, \dots, T_j \\ f_j(\hat{z}_{t-T_j-1+\tau:t-2+\tau}^{(j)}) & \text{if } \tau = T_j + 1, \dots, \end{cases} \tag{6}$$

where $j = 1, \dots, M$, $\hat{z}_t^{(j)}$ is the predicted value of the jth sequence of time series at time t. Figure 6 illustrates the case where $M = 2$, $T_1 = T_2 = L$ for the iterated method. The iterated method has also been applied to many classical machine learning models such as *recurring neural networks* [23, 24, 25], *hidden Markov models* [26, 27, 28], and so on.

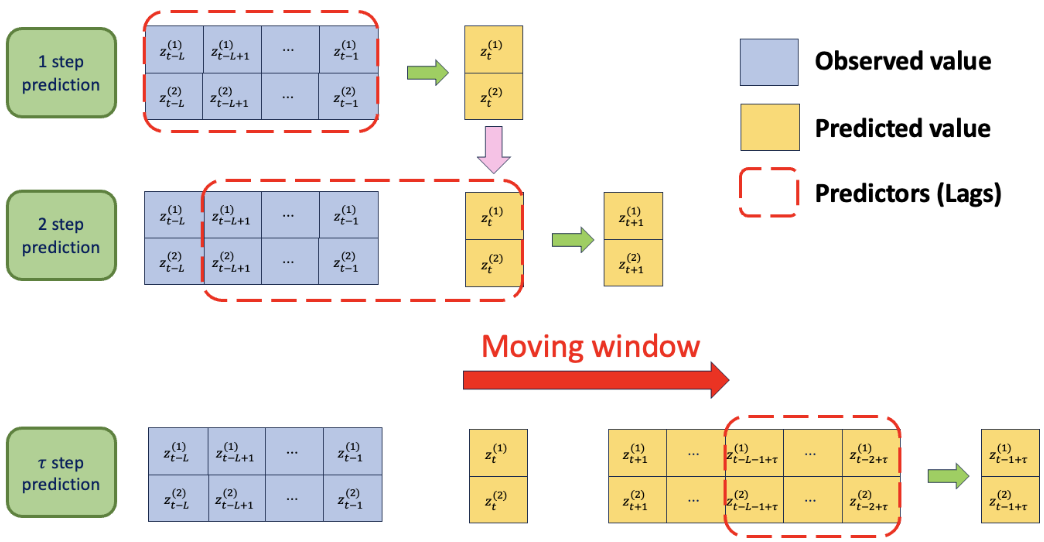

Figure 6: Iterated method for the multistep time series forecasting for two outputs with lag = L, lead time = τ ($\tau > L$).

A.3 Gaussian processes

GPs A Gaussian process (GP) [29] is a collection of random variables such that every finite number of which have a multivariate normal distribution. A GP is defined by a mean function $\mu(\cdot)$ and a covariance function $K(\cdot, \cdot)$ and is denoted by $\mathcal{GP}(\mu(\cdot), K(\cdot, \cdot))$.

GP regression Given a dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ comprising the inputs $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ (where $\mathbf{x}_i \in \mathbb{R}^d$) and the corresponding observations $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ (where $y_i \in \mathbb{R}$), suppose $y_i = f(\mathbf{x}_i)$; here $f\colon \mathbb{R}^d \to \mathbb{R}$ is a random function. Gaussian process regression assumes that the unknown function is a prior GP, denoted as $f(\cdot) \sim \mathcal{GP}(\mu(\cdot), K(\cdot, \cdot))$. Then the posterior distribution at a set of test points $\mathbf{X}^* = \{\mathbf{x}_i^*\}_{i=1}^m$ (where $\mathbf{x}_i^* \in \mathbb{R}^d$) has the following form:

$$p(f(\mathbf{X}^*) | \mathcal{D}) = \mathcal{N}(\mathbb{E}[f(\mathbf{X}^*)|\mathcal{D}], \mathrm{Cov}[f(\mathbf{X}^*)|\mathcal{D}]), \tag{7}$$

$$\mathbb{E}[f(\mathbf{X}^*)|\mathcal{D}] = \mu(\mathbf{X}^*) + K(\mathbf{X}^*, \mathbf{X})\big[K(\mathbf{X}, \mathbf{X})\big]^{-1}(\mathbf{y} - \mu(\mathbf{X}))\,, \tag{8a}$$

$$\mathrm{Cov}[f(\mathbf{X}^*)|\mathcal{D}] = K(\mathbf{X}^*, \mathbf{X}^*) - K(\mathbf{X}^*, \mathbf{X})\big[K(\mathbf{X}, \mathbf{X})\big]^{-1}K(\mathbf{X}, \mathbf{X}^*). \tag{8b}$$

A.4 GP for the MJO

Empirical GPs for the bivariate time series Here we denote the bivariate time series of RMMs by $\{z_t^{(j)}\}_{t=1}^T$, $j = 1, 2$, where T is the length of the time series. Suppose

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix} \sim \mathcal{GP}\Big(\mu\big(\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix}\big), K\big(\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix}, \begin{bmatrix} z_{t'}^{(1)} \\ z_{t'}^{(2)} \end{bmatrix}\big)\Big). \tag{9}$$

Then the predictive distribution of $[z_t^{(1)}, z_t^{(2)}]^\top$ at time t^* for the lag L is

$$p\Big(\begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix} \Big| \begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big) = \mathcal{N}(\boldsymbol{\mu}_{t^*}, \mathbf{K}_{t^*}), \tag{10}$$

and $\boldsymbol{\mu}_{t^*} \in \mathbb{R}^{2\times 1}$, $\mathbf{K}_{t^*} \in \mathbb{R}^{2\times 2}$ can be obtained by the following equations (11) and (12):

$$\begin{aligned}
\boldsymbol{\mu}_{t^*} &= \mathbb{E}\Big[\begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix} \Big| \begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big] \\
&= \mathbb{E}\Big[\begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix}\Big] + \mathrm{Cov}\Big[\begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix}, \begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big] \\
&\quad \mathrm{Cov}\Big[\begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}, \begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big]^{-1} \left(\begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix} - \mathbb{E}\Big[\begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big]\right) \\
&\approx \mathbb{E}\Big[\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix}\Big] + \mathrm{Cov}\Big[\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix}, \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}\Big] \\
&\quad \mathrm{Cov}\Big[\begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}, \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}\Big]^{-1} \left(\begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix} - \mathbb{E}\Big[\begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}\Big]\right),
\end{aligned} \tag{11}$$

$$\begin{aligned}
\mathbf{K}_{t^*} &= \mathrm{Cov}\Big[\begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix} \Big| \begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big] \\
&= \mathrm{Cov}\Big[\begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix}, \begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix}\Big] - \mathrm{Cov}\Big[\begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix}, \begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big] \\
&\quad \mathrm{Cov}\Big[\begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}, \begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}\Big]^{-1} \mathrm{Cov}\Big[\begin{bmatrix} z_{t^*-L:t^*-1}^{(1)} \\ z_{t^*-L:t^*-1}^{(2)} \end{bmatrix}, \begin{bmatrix} z_{t^*}^{(1)} \\ z_{t^*}^{(2)} \end{bmatrix}\Big] \\
&\approx \mathrm{Cov}\Big[\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix}, \begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix}\Big] - \mathrm{Cov}\Big[\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix}, \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}\Big] \\
&\quad \mathrm{Cov}\Big[\begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}, \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}\Big]^{-1} \mathrm{Cov}\Big[\begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}, \begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix}\Big],
\end{aligned} \tag{12}$$

where

$$\mathbf{X}^{(j)} = \begin{bmatrix} z_{1:L}^{(j)} \\ z_{2:L+1}^{(j)} \\ \vdots \\ z_{n:L+n-1}^{(j)} \end{bmatrix}^\top \in \mathbb{R}^{L\times n}, \quad \mathbf{y}^{(j)} = \begin{bmatrix} z_{L+1}^{(j)} \\ z_{L+2}^{(j)} \\ \vdots \\ z_{L+n}^{(j)} \end{bmatrix}^\top \in \mathbb{R}^{1\times n}, \quad j = 1, 2. \tag{13}$$

In equations (11) and (12), we use the empirical mean and covariance of n training data with lag L to approximate the expectation of the target and the covariance of the target and predictors.

Covariance update The corrected covariance $\tilde{\mathbf{K}}_{t^*}(\tau)$ corresponds to the lead time τ and can be scaled via the following transformation:

$$\mathbf{K}_{t^*} = \begin{bmatrix} \mathbf{K}_{t^*}[1,1] & \mathbf{K}_{t^*}[1,2] \\ \mathbf{K}_{t^*}[2,1] & \mathbf{K}_{t^*}[2,2] \end{bmatrix} \longrightarrow \tilde{\mathbf{K}}_{t^*}(\tau) = \begin{bmatrix} \tilde{V}_*^{(1)}(\tau) & \frac{\mathbf{K}_{t^*}[1,2]\sqrt{\tilde{V}_*^{(1)}(\tau)}\sqrt{\tilde{V}_*^{(2)}(\tau)}}{\sqrt{\mathbf{K}_{t^*}[1,1]}\sqrt{\mathbf{K}_{t^*}[2,2]}} \\ \frac{\mathbf{K}_{t^*}[2,1]\sqrt{\tilde{V}_*^{(1)}(\tau)}\sqrt{\tilde{V}_*^{(2)}(\tau)}}{\sqrt{\mathbf{K}_{t^*}[1,1]}\sqrt{\mathbf{K}_{t^*}[2,2]}} & \tilde{V}_*^{(2)}(\tau), \end{bmatrix} \tag{14}$$

where $\tilde{V}_*^{(1)}(\tau)$ and $\tilde{V}_*^{(2)}(\tau)$ are defined in formula (2).

Confidence region To obtain the confidence region of the distribution $\mathcal{N}(\boldsymbol{\mu}_{t^*}, \tilde{\mathbf{K}}_{t^*}(\tau))$, we use Lemma A.1 and Lemma A.2 as follows.

Lemma A.1. (Result 4.7 in Section 4.2 in [15]) *Let $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a p-variate normal distribution with location $\boldsymbol{\mu}$ and known covaraince $\boldsymbol{\Sigma}$. Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then*

(a) $(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution with p degrees of freedom.

(b) The $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution assigns probability $1-\alpha$ to the solid hyperellipsoid $\{\mathbf{x} \colon (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \le \chi^2_p(\alpha)\}$, where $\chi^2_p(\alpha)$ denotes the upper (100α)-th percentile of the χ^2_p distribution.

Proof. See proof of Result 4.7 in Section 4.2 in [15]. □

Lemma A.2. ((4-7) in Section 4.2 in [15]) *The hyperellipsoids $\{\mathbf{x} : (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2\}$ are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\,\mathbf{e}_i$, where λ_i's, $\mathbf{e}_i$'s are the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}$, namely, $\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i \mathbf{e}_i$, $i = 1, 2, \ldots, p$.*

Proof. From Result 4.1 in Section 4.2 in [15] we know that if $\boldsymbol{\Sigma}$ is positive definite and $\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i \mathbf{e}_i$, then $\lambda_i > 0$ and $\boldsymbol{\Sigma}^{-1}\mathbf{e}_i = \frac{1}{\lambda_i}\mathbf{e}_i$. That is, $(\frac{1}{\lambda_i}, \mathbf{e}_i)$ is an eigenvalue-eigenvector pair for $\boldsymbol{\Sigma}^{-1}$. According to the definition of the hyperellipsoid in quadratic form, we can conclude that the hyperellipsoids $\{\mathbf{x} : (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2\}$ are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\,\mathbf{e}_i$. □

Algorithm The entire algorithm of our model is described in Algorithm 1, and a flow chart of the model is shown in Figure 7.

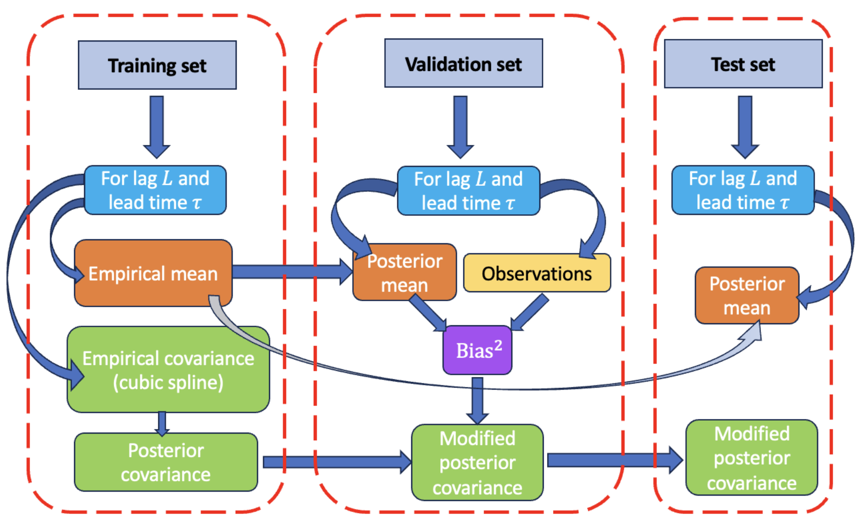

Figure 7: Diagram of the GP model for the MJO forecast.

Algorithm 1: GP model for the MJO forecast

Input: n : the size of training data
L : lag
t_v : start index for the predictions in validation set,
t_0 : start index for the predictions in test set,
τ : forecast lead time
$\{[z_t^{(1)}, z_t^{(2)}]\}_{t=1}^{L+n}$: training set
$\{[z_t^{(1)}, z_t^{(2)}]\}_{t=t_v}^{t_v+L+\tau+n_v-2}$: validation set
$\{[z_t^{(1)}, z_t^{(2)}]\}_{t=t_0-L}^{t_0-1}$: starting predictors in test set

Output: $\{\boldsymbol{\mu}_t\}_{t=t_0}^{t_0+\tau-1}$: predicted mean of $\{[\hat{z}_t^{(1)}, \hat{z}_t^{(2)}]\}_{t=t_0}^{t_0+\tau-1}$
$\{\tilde{\mathbf{K}}_{t_0}(t-t_0+1)\}_{t=t_0}^{t_0+\tau-1}$: predicted covariance of $\{[\hat{z}_t^{(1)}, \hat{z}_t^{(2)}]\}_{t=t_0}^{t_0+\tau-1}$

Construct the training dataset $\mathcal{D}^{(1:2)} = \{\mathbf{X}^{(1:2)}, \mathbf{y}^{(1:2)}\}$ by equations (1) and (13), $\mathbf{X}^{(1:2)} \in \mathbb{R}^{2L\times n}$, $\mathbf{y}^{(1:2)} \in \mathbb{R}^{2\times n}$

Compute $\mathbb{E}[\mathbf{y}^{(1:2)}]$

Obtain $\mathrm{Cov}\Big[\begin{bmatrix}\mathbf{X}^{(1:2)}\\ \mathbf{y}^{(1:2)}\end{bmatrix}, \begin{bmatrix}\mathbf{X}^{(1:2)}\\ \mathbf{y}^{(1:2)}\end{bmatrix}\Big] = \begin{bmatrix}\mathrm{Cov}[\mathbf{X}^{(1:2)}, \mathbf{X}^{(1:2)}] & \mathrm{Cov}[\mathbf{X}^{(1:2)}, \mathbf{y}^{(1:2)}]\\ \mathrm{Cov}[\mathbf{y}^{(1:2)}, \mathbf{X}^{(1:2)}] & \mathrm{Cov}[\mathbf{y}^{(1:2)}, \mathbf{y}^{(1:2)}]\end{bmatrix}$ by cubic spline interpolation

In validation set, obtain $\{\boldsymbol{\mu}_t, \mathbf{K}_t\}_{t=t_v+L+i-1}^{t_v+L+\tau+i-2}$ condition on $\{[z_t^{(1)}, z_t^{(2)}]\}_{t=t_v+t-1}^{t_v+L+i-2}$ for $i = 1, \ldots, n_v$ by equations (11) and (12); here $\mathbf{K}_t$ is equivalent for all t

In validation set, obtain modified covariances as a function of lead time $\{\tilde{\mathbf{K}}_{t_v}(t-t_v+1)\}_{t=t_v}^{t_v+\tau-1}$ by (2) and (14)

In test set, obtain $\{\boldsymbol{\mu}_t\}_{t=t_0}^{t_0+\tau-1}$ by equation (11)

In test set, apply the covariances obtained in validation set to the covariances in test set according to the corresponding lead time, $\tilde{\mathbf{K}}_{t_0}(l) \leftarrow \tilde{\mathbf{K}}_{t_v}(l)$, $l = 1, \ldots, \tau$

return $\boldsymbol{\mu}_t, \tilde{\mathbf{K}}_{t_0}(t-t_0+1)$, $t = t_0, \ldots, t_0+\tau-1$

B Experimental details

B.1 Evaluation metrics

We evaluate the performance of the model by computing the following metrics.

COR and RMSE Bivariate correlation coefficient (COR) and root mean squared error (RMSE) are defined as a function of the forecast lead time τ as follows:

$$\mathrm{COR}(\tau) = \frac{\sum_{t=1}^{n_p}\left(z_t^{(1)}\hat{z}_t^{(1)}(\tau) + z_t^{(2)}\hat{z}_t^{(2)}(\tau)\right)}{\sqrt{\sum_{t=1}^{n_p}\left(\left(z_t^{(1)}\right)^2 + \left(z_t^{(2)}\right)^2\right)}\sqrt{\sum_{t=1}^{n_p}\left(\left(\hat{z}_t^{(1)}(\tau)\right)^2 + \left(\hat{z}_t^{(2)}(\tau)\right)^2\right)}}, \tag{15}$$

$$\mathrm{RMSE}(\tau) = \sqrt{\frac{1}{n_p}\sum_{t=1}^{n_p}\left(\left(z_t^{(1)} - \hat{z}_t^{(1)}(\tau)\right)^2 + \left(z_t^{(2)} - \hat{z}_t^{(2)}(\tau)\right)^2\right)}, \tag{16}$$

where $z_t^{(1)}$ and $z_t^{(2)}$ are the observations of RMM1 and RMM2 on the tth day in the test set, $\hat{z}_t^{(1)}(\tau)$ and $\hat{z}_t^{(2)}(\tau)$ are the predictions of RMM1 and RMM2 on the tth day in the test set for the lead time of τ days, and n_p is the number of the predictions.

Phase error and amplitude error Phase error E_ϕ and amplitude error E_A are defined as a function of the forecast lead time τ as follows:

$$E_\phi(\tau) = \frac{1}{n_p}\sum_{t=1}^{n_p}\left(\hat{P}_t(\tau) - P_t\right), \tag{17}$$

$$E_A(\tau) = \frac{1}{n_p} \sum_{t=1}^{n_p} \big(\hat{A}_t(\tau) - A_t\big), \tag{18}$$

where P_t is the angle in degrees ($0^\circ - 360^\circ$) of the observation of RMMs $(z_t^{(1)}, z_t^{(2)})$ on the tth day in the test set, $\hat{P}_t(\tau)$ is the angle in degrees ($0^\circ - 360^\circ$) of the predictions of RMMs $(\hat{z}_t^{(1)}(\tau), \hat{z}_t^{(2)}(\tau))$ on the tth day in the test set for the lead time of τ days. A_t is the observation of RMM amplitude on the tth day in the test set, and $\hat{A}_t(\tau) :\, = \sqrt{\big(\hat{z}_t^{(1)}(\tau)\big)^2 + \big(\hat{z}_t^{(2)}(\tau)\big)^2}$ is the predicted amplitude on the tth day in the test set for the lead time of τ days.

B.2 Additional results

MJO phase diagram Figure 8 shows the MJO phase diagram for Aug–15–2012 to Oct–13–2012 of our model with 68.0% confidence region. It is obvious from the figure that almost all observations (black lines) are covered by the confidence region (colorful shadings), which shows the good performance of our model. Animated phase diagrams can also be found on the project website `https://gp-mjo.github.io/`, which show how the elliptical confidence region enlarges with time.

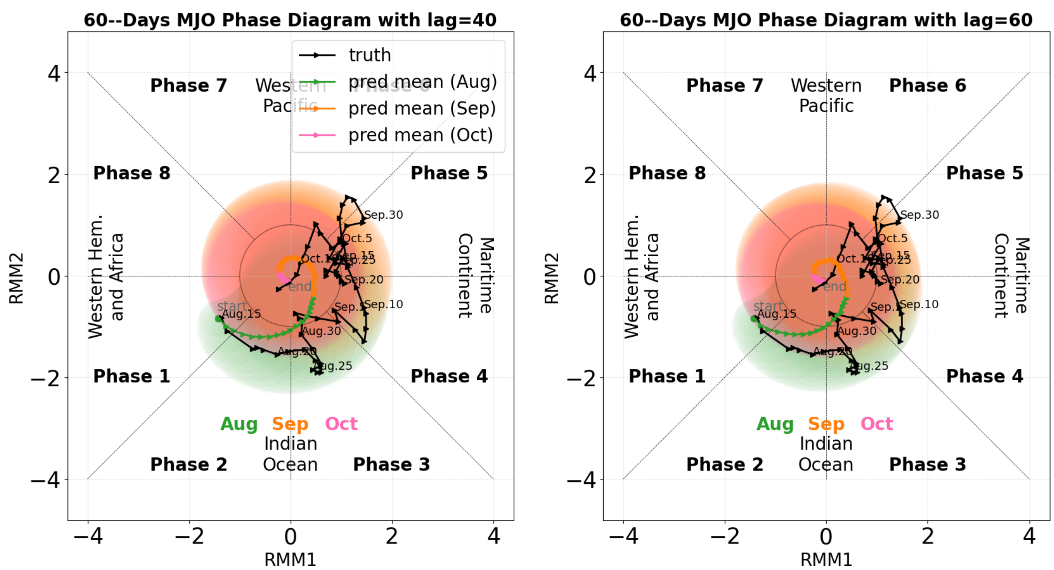

Figure 8: 60-days MJO phase diagram for Aug–15–2012 to Oct–13–2012 with lag $L = 40, 60$. Black lines are observations (truth). Green lines are predictions in August, and green shadings are 68.0% confidence region in August. Orange lines are predictions in September, and orange shadings are 68.0% confidence region in September. Pink lines are predictions in October, and pink shadings are 68.0% confidence region in October. *Left:* lag $L = 40$. *Right:* lag $L = 60$.