Geopotential-based Multivariate MJO Index: Extending RMM-like Indices to Pre-Satellite Era

Jeremy Cheuk-Hin Leung (chleung@pku.edu.cn)
Guangzhou Institute of Tropical and Marine Meteorology

Weihong Qian
Peking University

Peiqun Zhang
CMA: China Meteorological Administration

Banglin Zhang
Guangzhou Institute of Tropical and Marine Meteorology

Research Article

Keywords: Madden-Julian Oscillation, MJO index, GMM index, upper-level zonal geopotential gradient, pre-satellite era

Posted Date: July 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-708454/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Climate Dynamics on January 30th, 2022. See the published version at https://doi.org/10.1007/s00382-022-06142-2.
Geopotential-based Multivariate MJO Index: Extending RMM-like Indices to Pre-Satellite Era

Jeremy Cheuk-Hin LEUNG\(^1,2\), Weihong QIAN\(^1,2\,*\), Peiqun Zhang\(^3\), Banglin ZHANG\(^1,4,5\,*\)

\(^1\)Guangzhou Institute of Tropical and Marine Meteorology/Guangdong Provincial Key Laboratory of Regional Numerical Weather Prediction, China Meteorological Administration, China

\(^2\)Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, China

\(^3\)Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, China

\(^4\)Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), China

\(^5\)College of Atmospheric Sciences, Lanzhou University, China

Submitted to *Climate Dynamics* (June 2021)

Corresponding author: Banglin Zhang (zhangbl@gd121.cn), Weihong Qian (qianwh@pku.edu.cn)
Abstract

Model simulations suggest that Madden-Julian Oscillation (MJO) activity changes under the anthropogenic climate change background. However, satellite observations, which provide information of MJO convection activity, are not available before the 1970s, hindering research on the historical long-term variability of MJO. This study aims at extending the data length of MJO indices that include both MJO circulation and convection features, such as the widely used Real-Multivariate MJO (RMM) index, to the pre-satellite era. This paper introduces a new MJO index construction method, of which the outgoing longwave radiation (OLR) input is derived from upper-level geopotential, and names it as the Geopotential-based Multivariate MJO (GMM) index. The GMM index is derived over 1902–2008 and compared with the filtered version of RMM (FMM) index during 1981–2008. The GMM index is shown to (1) have the same climatological properties as the FMM index, (2) be statistically highly correlated to the FMM index, and (3) be able to indicate MJO activities and its convection features in the pre-satellite era. The overall bivariate correlation between FMM index and GMM index based on ERA-20C ranges from 0.959 to 0.968 in different phases. Evaluation results confirm the validity of the proposed MJO index construction method, which could capture MJO convection activity in the pre-satellite era and can be applied to all MJO indices that require input of OLR. This study provides an alternative way that overcomes the difficulty of historical MJO studies, and will be beneficial to our understanding of the long-term change of MJO.

Keywords

Madden-Julian Oscillation · MJO index · GMM index · upper-level zonal geopotential gradient · pre-satellite era
1. Introduction

The Madden-Julian Oscillation (MJO; Madden and Julian 1971, 1972), named after Madden and Julian, is an important tropical atmospheric system often treated as a bridge linking weather and climate (Zhang 2013). The MJO, despite being named as an oscillation, acts more like an eastward-moving atmospheric pulse with convection clusters 5000–10,000 km spatial scale (Nakazawa 1988; Lau et al. 2012) and notable zonal wind anomalies along the equator. Lower-tropospheric equatorial westerly and easterly anomalies are observed to the west and east of the MJO deep convection center, respectively. The low-level westerly and easterly anomalies are, respectively, the Rossby and Kelvin wave responses to the heat source over the deep convection center (Gill 1980; Wang and Rui 1990; Hsu and Li 2012; Wang and Chen 2017). The upper-level zonal wind also shows similar Rossby-Kelvin wave responses.

Given the dominant planetary-scale circulation and convection features of the MJO, MJO signals are usually extracted from outgoing longwave radiation (OLR) and tropospheric zonal wind data. Numerous methods and indices were proposed to derive MJO signals from observed atmospheric data. Among all, the Real-Multivariate MJO (RMM) index (Wheeler and Hendon 2004, hereafter WH04), developed based on OLR, zonal wind at 200-hPa (u200) and 850-hPa (u850), is the most widely used MJO index among all. Constructed from the two leading principal components (PCs) of combined empirical orthogonal function (CEOF), the RMM index could well visualize each MJO event’s propagation and intensity in a two-dimensional Cartesian phase diagram, in terms of RMM phase and RMM intensity. Although some research pointed out the limitations and shortages of the RMM index (e.g. Roundy and Schreck III 2009; Ventrice et al. 2011; Straub 2013; Wolding and Maloney 2015; Liu et al. 2016), the simplicity and convenience of the RMM index makes it the most commonly used MJO index at the present time.

A limitation of the RMM index is its short data record. Similar to most MJO indices, the RMM index is derived from OLR, a satellite-based data that captures MJO’s convection features (Liebmann and Smith 1996), thus it is difficult to construct the RMM index in the pre-satellite era when satellite data is absent. In other words, this hinders scientists from studying the observed long-term variability of MJO before the 1970s by using the RMM index. In order to explore MJO’s long-term change and interdecadal variation, Oliver and Thompson (2012) reconstructed the RMM index over the period 1905–2008 by regressing tropical surface pressures onto WH04’s RMM index. The MJO-related convection information in the RMM index was assumed to be included in the regression model, although no
convection data was directly input into the regression model. Their reconstructed index only accounts for 69% of the RMM index.

Other different ways were tried to extract MJO signals in the pre-satellite era. A simple yet common method is examining the intraseasonal variability of lower- and upper-tropospheric circulation (e.g. zonal wind) over the MJO-active area. For example, Slingo et al. (1999) showed an increasing trend of interannual MJO activity since 1958, by analyzing the variances of bandpass (20–100 days) filtered u200 and u850 averaged between 10°S and 10°N. Chen et al. (2017) also defined an MJO intensity index in a similar way (amplitude of u850 averaged over 10°S–10°N, 120–160°E) in his study of interannual and interdecadal MJO variability during 1861–2010. Another method reflecting MJO activity is to compute the two leading PCs of MJO circulation responses, such as u200, u850, velocity potential, stream function or combinations of them (Slingo et al. 1999; Jones and Carvalho 2006; Ventrice et al. 2013). A widely known example is the Velocity Potential MJO (VPM) index, which captures MJO signals from u200, u850 and 200hPa velocity potential (Ventrice et al. 2013). Although the VPM index was originally not designed to solve the data length problem of the RMM index, the VPM index does not require observations of OLR and is thus available in the pre-satellite era. Among the above, however, most methods stated above could hardly consider both the convection and circulation features of MJO in their studies due to the absence of OLR observations.

Since the MJO plays an important role in extreme weather climate events, it is important to understand how the MJO activity changes under the anthropogenic climate change background. However, due to the lack of satellite observations, studies on historical MJO variability could hardly capture the MJO’s convection activity; and, studies about the impacts of climate change on MJO are mostly based on sophisticated modelling techniques, which might not provide perfect representations of the climate system (Henderson et al. 2017; Maloney et al. 2019). Thus, aiming at solving the above difficulties, this study introduces a new method, based on recent findings about the close relationship between atmospheric geopotential field and MJO deep convection (Leung and Qian 2017, see Section 2), to construct an extended MJO index that (1) takes both MJO’s deep convection and circulation features into account, and (2) covering both the satellite and pre-satellite eras. Leung and Qian (2017) discovered that the zonal gradient of equatorial 150-hPa geopotential (z150) anomaly (hereafter 150-hPa $\nabla z'$) is able to indicate MJO deep convection centers, which is consistent with the similarity of the upper tropospheric circulation pattern in the MJO to a free Kelvin wave (Roundy 2020). As mentioned above, OLR
observation data is only available since 1974, thus one has to obtain a longer time series of OLR before extending the MJO index. The discovery of 150-hPa $\nabla_z \zeta'$ provides a new option to reproduce the MJO-related OLR and the MJO index before 1974, from pre-satellite era z150 data. More detail about the 150-hPa $\nabla_z \zeta'$ is discussed in Section 3 after the data introduction in Section 2. The procedure of constructing the new MJO index is also introduced in Section 3. The validity of the new proposed method and MJO index are evaluated in Section 4. Section 5 examines the extended MJO index’s ability to indicate MJO’s activity in the pre-satellite era. Conclusions and discussion are given in Section 6.

2. Data

In the following sections, the extended MJO index was built from the European Centre for Medium-Range Weather Forecasts (ECMWF) Atmospheric Reanalysis of the 20th Century (ERA-20C) product (ECMWF; Poli et al. 2016). The ERA-20C provides 4D-Var analysis of surface and upper-air data from 1900–2010 by assimilating surface pressure and marine winds observations only. Daily-mean z150, u200 and u850 data with a horizontal resolution of 1°×1° were used to derive the extended MJO index. Since satellite and upper-air observations are not assimilated in the ERA-20C product, the reanalysis product may not provide the best estimate of the atmospheric system. Thus, z150, u200 and u850 data from the ECMWF Re-analysis Interim (ERA-I) product (Dee et al. 2011) was also used to evaluate the validity of the proposed method of calculating the extended MJO index. The ERA-I and ERA-20C data provides 6-hourly values only, and were converted into daily-mean values by taking averages.

Daily-mean OLR observation was obtained from the National Oceanic and Atmospheric Administration (NOAA) gridded interpolated OLR data archive (Liebmann and Smith 1996). The OLR data is based on polar-orbiting satellite observation and covers globally with a spatial resolution of 2.5°×2.5° from 1974 to 2019.

In order to verify the extended MJO index’s ability to indicate MJO activity in the pre-satellite data, historical rainfall observations are used in the following analyses. The historical rainfall data is obtained from the Global Historical Climatology Network – Daily (GHCND) archive (Durre et al. 2008, 2010; Menne et al. 2012), which contains daily precipitation records from over 100,000 stations in 180 countries and territories up to 175 years. In this study, rainfall records of stations over Central North Australia (0°–15°S, 130°–135°E) and Northeast Australia (0°–15°S, 140°–145°E) that cover the period
of 1902–2008 are selected.

3. Construction of the Geopotential-based Multivariate MJO (GMM) index

3.1 150-hPa $\nabla_z \zeta'$ – a proxy of OLR measurement

Since OLR measurements are based on satellite observations, a good proxy for OLR is needed for constructing an extended MJO index that contains MJO’s convection information. Leung and Qian (2017), by examining the observed geopotential structures of all MJO events during 1979–2013, discovered that 150-hPa $\nabla_z \zeta'$ (i.e., zonal gradient of equatorial $z_{150}$ anomaly) is in phase with MJO’s convection activity. Namely, 150-hPa $\nabla_z \zeta'$ always shows positive above the MJO convection center, and vice versa.

The relationship between 150-hPa $\nabla_z \zeta'$ and MJO convection is an outcome of MJO’s Kelvin wave response. The MJO has a first baroclinic structure. In the lower troposphere, an equatorial low and high are respectively located to the east and west of the MJO convection center and a pair of off-equatorial low in its two flanks; in the upper troposphere, the equatorial low and high are respectively to the west and east of the convection center and a pair of off-equatorial highs is in its two flanks (Adames and Wallace 2014). The MJO convection center is located between the upper-level high and low anomaly to its east and west respectively, and this is why 150-hPa $\nabla_z \zeta'$ is in phase with MJO’s convection activity. Through numerical simulation experiments based on the ideal theoretical MJO framework by Wang and Chen (2017), it was shown that the upper-level positive $\nabla_z \zeta'$ center always appears right above the MJO convection center once the MJO Rossby-Kelvin wave packet starts to develop (Leung 2019). Leung and Qian (2017), using four sets of reanalysis products (including ERA-I, ERA40, NCEP1 and NCEP2), also showed the strong correlation between 150-hPa $\nabla_z \zeta'$ and OLR anomaly after applying a wavenumber-frequency filter (space-time bandpass filter, Wheeler and Kiladis 1999; Kiladis et al. 2006) of 0–10 wavenumber and 15–100 day period (hereafter MJO-filter). Among the four reanalysis products, the 150-hPa $\nabla_z \zeta'$ derived from the ERA-I dataset has the highest correlation coefficients with OLR anomaly.
3.2 Procedure of constructing the GMM index

Given the high correlation of 150-hPa $\nabla_z z'$ with OLR as well as the long record of ERA-20C reanalysis, one could estimate the OLR by 150-hPa $\nabla_z z'$ in the pre-satellite era. This could compensate for the missing OLR data due to the absence of satellite observation. Using the estimated OLR and zonal wind data from ERA-20C reanalysis, an extended MJO index that takes both MJO-related circulation and convection features into account can be derived based on procedures similar to that in WH04. In order to distinguish with the currently existing MJO indices, the MJO index introduced in this paper will be referred to as “Geopotential-based Multivariate MJO Index” (GMM index). The detailed procedure of constructing the GMM index is as follow:

(1) Data pre-processing

Daily-mean z150, u200 and u850 data from reanalysis products and OLR data were used in the following steps. In order to remove unrelated seasonal, interannual and interdecadal variability etc., the mean and first three harmonics of the annual cycle of each variable are first removed. Then, a wavenumber-frequency filter (space-time bandpass filter, Wheeler and Kiladis 1999; Kiladis et al. 2006) of 0–10 wavenumber and 15–100 day period (MJO filter) is applied to extract the intraseasonal, eastward-propagating MJO-related anomalies (hereafter, anomalies are referred to as MJO-filtered anomalies unless specified). The resulting u850, u200, OLR and z150 anomalies were assumed to be mostly containing MJO information and used to construct the GMM index below. The MJO filter may cause weakening effects at both ends of time series, and hence the data of the first 2 years and last 2 years of each dataset is excluded in the following analyses.

(2) Calculation of 150-hPa $\nabla_z z'$

After obtaining the MJO-filtered anomalies, the 150-hPa $\nabla_z z'$ (unit: gpm m$^{-1}$), namely the zonal gradient of z150 anomaly, is derived by taking the zonal spatial derivative of MJO-filtered equatorial z150 anomaly: $\nabla_z z' = \frac{\partial z'}{\partial X}$, where $z'$ and $X$ are the z150 anomaly...
averaged over 10°S–10°N and the zonal distance respectively.

(3) Derivation of OLR from 150-hPa $\nabla_{z} \zeta'$

The OLR is derived from 150-hPa $\nabla_{z} \zeta'$ based on linear regression. Leung and Qian (2017) verified that the linear relationship between 10°S–10°N 150-hPa $\nabla_{z} \zeta'$ and 15°S–15°N OLR holds not only in the satellite era, but also during 1974–1977 using the National Centers for Environmental Prediction (NCEP) Reanalysis 1 (NCEP R1) (Kalnay et al. 1996) and 40-year ECMWF Re-Analysis (ERA-40) (Uppala et al. 2005) data. However, this relationship was not confirmed with ERA-20C. Here, we show in Fig. 1 that the linear relationship between 15°S–15°N OLR anomaly and 10°S–10°N 150-hPa $\nabla_{z} \zeta'$ derived from ERA-20C (Figs. 1c, 1d) does hold over MJO-active regions as that derived from ERA-I (Figs. 1a, 1b). Despite the fact that 150-hPa $\nabla_{z} \zeta'$ derived from ERA-20C has comparatively lower correlation coefficients (-0.823 and -0.775 in Figs. 1c and 1d compared to -0.893 and -0.872 in Figs. 1a and 1b) with OLR anomaly, 150-hPa $\nabla_{z} \zeta'$ is still a valid choice for deriving OLR in the pre-satellite era. In the following context, OLR anomaly is referred to OLR anomaly averaged over 15°S–15°N and 150-hPa $\nabla_{z} \zeta'$ is referred to that averaged over 10°S–10°N, unless specified. The different latitude domains of OLR anomaly and 150-hPa $\nabla_{z} \zeta'$ are selected based on the results in Leung and Qian (2017).

It was also reported that 150-hPa $\nabla_{z} \zeta'$ may have 1- to 2-day time lag with OLR anomaly over different MJO stages (Leung and Qian 2017). Considering MJO’s eastward propagating property, the time lag between 150-hPa $\nabla_{z} \zeta'$ and OLR implies that there may be zonal displacements between the two variables. Thus, before applying linear regression, we first calculate the correlation coefficients between OLR anomaly of each longitude ($\lambda_{OLR}$) and 150-hPa $\nabla_{z} \zeta'$ of 15 degrees longitude nearby, and then determine the longitude ($\lambda_{z}$) at which 150-hPa $\nabla_{z} \zeta'$ has the largest correlation coefficient with OLR anomaly for $\lambda_{OLR}$. 

7
Considering the seasonal variation of MJO activity, the calculation $\lambda_z$ for each $\lambda_{\text{OLR}}$ is repeated for each month, thus $\lambda_z$ is a matrix with dimension of length of $\lambda_{\text{OLR}}$ times number of months ($144 \times 5$). Since this study only covers the extended winter period, only five months (NDJFM) are included. Then, OLR anomaly at each $\lambda_{\text{OLR}}$ is linearly regressed on 150-hPa $\nabla_x z'$ at $\lambda_z$ (i.e. largest correlation), obtaining the regression coefficient $m(\lambda_{\text{OLR}}, \lambda_z, \text{month})$ and intercept term $c(\lambda_{\text{OLR}}, \lambda_z, \text{month})$ (Eq. 1):

$$O_L R(\lambda_{\text{OLR}}) = m(\lambda_{\text{OLR}}, \lambda_z, \text{month}) \cdot \nabla_x z'(\lambda_z) + c(\lambda_{\text{OLR}}, \lambda_z, \text{month}) \quad (\text{Eq. 1})$$

By plugging the regression coefficient $m(\lambda_{\text{OLR}}, \lambda_z, \text{month})$ and intercept term $c(\lambda_{\text{OLR}}, \lambda_z, \text{month})$ into Eq. 1, the global equatorial daily OLR anomaly during 1902–2008 can be obtained from the ERA-20C 150-hPa $\nabla_x z'$. Noted that MJO convection activity is associated with negative OLR anomalies and positive 150-hPa $\nabla_x z'$ (Section 3.1), which means their correlation coefficients should be theoretically negative, thus the above-mentioned “largest correlation coefficient between OLR and 150-hPa $\nabla_x z'$” refers to correlation coefficient that is closest to -1. Fig. 1 is an example illustrating the linear relationship between $O_L R(\lambda_{\text{OLR}})$ and its corresponding 150-hPa $\nabla_x z'(\lambda_z)$ during November of 1981–2008.

Fig. 2 plots the maximum correlation coefficient of OLR anomaly at each longitude ($O_L R(\lambda_{\text{OLR}})$) with its corresponding 150-hPa $\nabla_x z'(\lambda_z)$ during extended winter of 1981–2008. The correlation between the two variables is larger between 60°E and 180°E (MJO-active regions, correlation coefficients $\geq -0.7$), and relatively smaller in the Western Hemisphere, due to the weaker MJO convection activity. The correlation coefficients of OLR anomaly with 150-hPa $\nabla_x z'$ derived from ERA-20C (Fig. 2b) are slightly weaker than that with 150-hPa $\nabla_x z'$ derived from ERA-I (Fig. 2a). This suggests the poorer data quality of the ERA-20C reanalysis, mostly due to the lack of satellite data input, in reflecting the MJO’s atmospheric status. Nevertheless, the close relationship between OLR and 150-hPa $\nabla_x z'$ still holds in ERA-20C data with their correlation coefficients ranging between -0.7 and -0.9 over the MJO-active region (statistically significant at the 99.9% confidence level), implying that the ERA-20C 150-hPa $\nabla_x z'$ could well reflect MJO’s convection feature and it is reasonable
to derive OLR anomaly from ERA-20C products. The statistical significance of correlation coefficients here and throughout the paper is tested using Student’s t-test, with the degree of freedom estimated following Wang and Li (2017) considering the reduction of degree of freedom due to filtering.

Fig. 3 is an example demonstrating the derived OLR based on ERA-I and ERA-20C 150-hPa $\nabla_x z'$ during the 2006–2007 extended winter following the above procedures. According to the observed OLR anomaly (NOAA gridded interpolated OLR), there were two MJO events observed during the 2006–2007 extended winter, respectively initiated in December 2006 and February 2007 (marked by arrows in Fig. 3a). The two MJO events are also indicated in the Hovmöller diagram of 150-hPa $\nabla_x z'$ derived from ERA-I (Fig. 3d) and ERA-20C (Fig. 3e), featured with positive 150-hPa $\nabla_x z'$ propagating signals (arrows in Figs. 3d and 3e). By applying the linear regression method introduced above (Eq. 1), the OLR anomalies derived from ERA-I and ERA-20C 150-hPa $\nabla_x z'$ are plotted in Figs. 3b and 3c, respectively, and both of them do reflect the two MJO events’ signals. Although the derived OLR anomalies are apparently weaker than observations from 180°E to 60°E (Western Hemisphere and Africa) because of the weaker correlation between 150-hPa $\nabla_x z'$ and OLR anomalies (Fig. 2), this error does not affect much when constructing the GMM index (shown later in Fig. 4). Another important note is the overall smaller amplitude of OLR anomalies derived from ERA-20C (note the different colorbars in Fig. 3). This is due to the overall relatively smaller correlation coefficients of OLR with 150-hPa $\nabla_x z'$ derived from ERA-20C (Fig. 2b), compared to that with ERA-I 150-hPa $\nabla_x z'$ (Fig. 2a). Yet this does not affect the GMM index calculation either (shown later in Fig. 4).

(4) Construction of GMM index

After deriving OLR anomalies from 150-hPa $\nabla_x z'$, the remaining steps are similar to that documented in previous studies. Following WH04’s method of constructing RMM index, the CEOF analysis is applied on the 15°S–15°N averaged MJO-filtered u200, u850 anomalies and OLR derived from 150-hPa $\nabla_x z'$. The resulting PCs (PC1 and PC2) of the first two leading modes (EOF1 and EOF2) are then used to define the phase and intensity of MJO index on a two-dimensional Cartesian phase diagram. The fundamental difference of the MJO index
presented here is that the OLR input of CEOF is not the primary satellite-observed OLR data but derived from atmospheric geopotential field, which is available with longer data record.

Fig. 4 presents the patterns of first two leading EOF modes by applying the CEOF analysis on three combination sets of MJO-filtered variables during extended winter of 1981–2008:

NOAA gridded interpolated OLR + ERA-I u200 + ERA-I u850 (Set 1, Fig. 4a), OLR derived from ERA-I 150-hPa $\nabla_x z'$ + ERA-I u200 + ERA-I u850 (Set 2, Fig. 4b), and OLR derived from ERA-20C 150-hPa $\nabla_x z'$ + ERA-20C u200 + ERA-20C u850 (Set 3, Fig. 4c). Set 1 is the same as the RMM index except the CEOF input variables are MJO-filtered anomalies, and is referred to as the reference that represents the “correct” MJO index. Set 2 is similar to Set 1 besides the OLR input is derived from ERA-I 150-hPa $\nabla_x z'$ (GMM-ERAI index), i.e. the GMM index derived from ERA-I. Set 3 is the GMM index derived from ERA-20C, with everything being the same as Set 2 except that all input data is based on ERA-20C, i.e. the GMM index derived from ERA-20C (GMM-ERA20C index).

The first two leading EOF modes of all 3 sets explains about 50% of total variance and show the major MJO patterns consistent with that mentioned in WH04: EOF1 shows enhanced convection activities over Maritime Continent and suppressed convection over Africa and Western Hemisphere while EOF2 has enhanced and suppressed convection respectively over Pacific Ocean and Indian Ocean, respectively (Fig. 4). Only small differences exist between the EOFs based on observed and derived OLR: both EOF1 and EOF2 of Set 2 and Set 3 slightly underestimate the MJO convection activities over the Western Hemisphere and Africa (from 180°E to 60°E) (black line in Fig. 4b); the region of anomalous OLR of Set 1 and Set 2 are slightly narrower than that of Set 3. Since the OLR input is the only source of error of Set 2, compared to Set 1, we could attribute the above difference to the error of OLR derivation from 150-hPa $\nabla_x z'$, namely the weaker correlation between OLR and 150-hPa $\nabla_x z'$ over the Western Hemisphere and Africa. Meanwhile, although the OLR anomalies derived from ERA-20C 150-hPa $\nabla_x z'$ are overall smaller than observations (Fig. 3c), the CEOF result of Set 3 is basically the same as that of Set 2. Unlike Set 2, the sources of error of Set 3 could be due to both the relatively worse data quality of ERA-20C and the error of OLR derivation from 150-hPa $\nabla_x z'$. In spite of some differences in Set 2 and Set 3 compared with Set 1, the first two leading EOF modes of Set 2 and Set 3 do capture the major MJO convection and circulation
features. The correlation coefficients of the resulted EOF1 and EOF2 OLR components between Set 1 and Set 2 respectively reach 0.962 and 0.922; and that between Set 1 and Set 3 respectively reach 0.929 and 0.892 (significant at the 99.9% confidence level). These results imply that the small inconsistencies in the CEOF results do not produce large errors in the final GMM index (more details shown in Section 4).

The above procedures of calculating the GMM index are basically the same as that introduced in WH04 except that (1) the OLR input is derived from 150-hPa $\nabla_x z'$ and (2) all variables have been MJO-filtered. The remarkable similarity of the above CEOF results (Figs. 4b and 4c) to that in WH04 confirms the feasibility of constructing an MJO index that includes MJO convection features without the use of raw OLR observations.

4. Validity of the GMM index and sources of errors

In this section, we will assess the validity of the GMM index by (1) examining the climatological properties of GMM index, (2) statistically evaluating the GMM index and (3) the univariate indices based on the derived OLR. As mentioned above, the GMM index in this study is derived from MJO-filtered anomalies, thus the RMM index that is based on MJO-filtered anomalies (hereafter FMM index) is referred to as a standard reference for the following evaluation. The FMM index is constructed from the CEOF results of MJO-filtered ERA-I u200, ERA-I u850 and NOAA gridded interpolated OLR anomalies, which is equivalent to Set 1 in Section 3.2 (Fig. 4a). The GMM indices built from Set 2 and Set 3 are respectively defined as the GMM-ERAI index and GMM-ERA20C index, among which the latter covers both the pre-satellite and modern satellite era. Considering the weakening effect of the MJO filter at both ends of time series, the data of the first 2 years and last 2 years of each dataset is neglected in the following analyses. i.e. the FMM, GMM-ERAI indices cover the extended winter during 1981–2008 while the GMM-ERA20C index covers 1902–2008.

4.1 Climatological properties of the GMM index

An advantage of GMM index, as if RMM and FMM indices, is its capability of reflecting MJO’s intensity and dividing MJO’s life cycle into 8 phases. By combining the PCs of CEOF results (Fig. 4), one could visualize the FMM and GMM indices on a two-dimensional Cartesian phase diagram. Fig. 5 plots all the data points of GMM-ERA20C index of the extended winter and each month during 1902–
2008, showing the climatology of GMM-ERA20C index. Following Lafleur et al. (2015), the GMM
index of each day is classified into four categories: (1) inactive (IA, GMM<1.0), (2) active (A,
1.0≤GMM<1.5), (3) very active (VA, 1.5≤GMM<2.5) and (4) extremely active (EA, GMM≥2.5) MJO
days. Among the 107-year period (151 days × 107 years = 16157 days), inactive and active MJO days
accounts for 36% and 24% days of the whole period while very active and extremely active MJO days
occupied 35% and 5% of the whole period, respectively. This is consistent with the climatology of RMM
index discussed in Lafleur et al. (2015).

By comparing the composite spatial distribution of OLR and z150 anomalies of different MJO
phases based on the three MJO indices, the GMM-ERAI and GMM-ERA20C indices are shown to
reproduce the same MJO features as that indicated by the FMM index. Fig. 6 shows the composited
NOAA interpolated OLR and ERA-I z150 of each MJO phase during the extended winter of 1981–2008.
The MJO phases of each column are defined based on the three MJO indices and are referred to as FMM
phase, GMM-ERAI phase and GMM-ERA20C phase, respectively. The spatial distributions of both
OLR and z150 anomalies, especially their center locations and intensities, of GMM-ERAI (Fig. 6b) and
GMM-ERA20C (Fig. 6c) indices are consistent with that of FMM index (Fig. 6a). In Phase 1 of all the
three indices, negative OLR anomalies occupy Africa and Western Indian Ocean and enhanced
convection is observed over Maritime Continent and Western Pacific. The convection cluster propagates
eastward. The MJO Rossby-Kelvin wave packet becomes clear in the upper tropospheric geopotential
field in Phase 3 when the convection center reaches Eastern Indian Ocean. The convection center is
located between the Kelvin trough to the west and the Kelvin ridge to the east, where 150-hPa \( \nabla \cdot z' \) is
maximized; meanwhile, the positive OLR anomalies over the Pacific are associated with negative 150-
hPa \( \nabla \cdot z' \). The negative OLR anomaly center moves to Central Pacific and weakens in Phases 7 and 8,
corresponding to the dissipation of MJO. The pattern correlations of OLR and z150 composites based on
the FMM index with those based on the GMM-ERAI and GMM-ERA20C indices respectively range
from 0.96–0.99 and 0.95–0.98 in difference MJO phases. The above result is consistent with that in
WH04 and demonstrates that both GMM-ERAI and GMM-ERA20C indices are able to reproduce the
same MJO circulation and convection features as the FMM and RMM indices do.

4.2 Statistical errors of the GMM index in the satellite era

The main purpose of building an MJO index is to visualize the intensities and positions of MJO

activities. Thus, in addition to the climatological properties of GMM index, it is more important to validate the ability of GMM index to reflect the correct intensities and phases of MJO events. We have shown from OLR observations that two MJO events occurred during the 2006–2007 extended winter (Fig. 3). Here, we compare the performances of GMM-ERA and GMM-ERA20C indices with the FMM index (Fig. 7).

As shown in Fig. 3a, MJO convection of the first MJO event initiated in mid-December, sustained for about 1 month and dissipated in mid-January; the second MJO event appeared over Africa in early-February, propagated at a slower speed and finally disappeared in late-March. The FMM index is able to reproduce the two MJO events mentioned above (Fig. 7a). The FMM index stays weak from November to early-December and its amplitude attains 1 in Phase 1 on 15 December 2006, which corresponds to the first MJO event. The FMM index reaches its maximum amplitude at the boundary between Phase 4 and Phase 5 on 3 January 2007, which is consistent with the observed OLR anomaly center over 130°E (Fig. 3a), and then dissipates in late-January in Phase 7. Then, the FMM index once again attains the active MJO level on 7 February 2007 in Phase 8, implying the start of the second MJO event. The FMM index reaches maximum on 21 February in Phase 2, then weakens to inactive MJO level on 6 March, and strengthens to active MJO-level again on 20 March. According to the FMM index, there are indeed two separate MJO events during February–March 2007, which is inconsistent with the observed OLR anomalies that show one MJO event only. This could be explained from the zonal wind anomalies which show two separate eastward propagating upper-tropospheric easterly and lower-tropospheric westerly systems (Fig. 8), implicating two MJO events, during the same period. The RMM index and VPM index both indicate two MJO events during February–March 2007 (Figures omitted). The reason why Fig. 3a shows only one MJO event during February–March 2007 is probably due to that the MJO-filtered anomalies combine the two successive MJO events into one.

As shown in Figs. 7b and 7c, the GMM-ERA and GMM-ERA20C indices are both able to indicate the two MJO events during the 2006–2007 extended winter. Results show that both the GMM-ERA and GMM-ERA20C indices have the ability to reflect MJO activities. However, compared to the FMM index, the GMM index in general overestimates the MJO strengths. Specifically, the maximum FMM intensity during the 2006–2007 extended winter is 2.48, but the maximum intensities of GMM-ERA and GMM-ERA20C over the same period are respectively 2.73 and 2.78. The GMM-ERA and GMM-ERA20C indices overall respectively underestimates and overestimates MJO intensities where FMM>=1 by 0.03
(2.0%) and 0.15 (8.6%) during 1981–2008. In addition, errors of the GMM index are in general more obvious form Phase 6 to Phase 8, for example, the second MJO event initiated in early-February 2007 in Phase 8 according to the FMM index but the GMM index indicates that the event started in late-January in Phase 7. And the errors of GMM-ERA20C index is slightly larger than that of the GMM-ERAI index.

Although both the GMM-ERAI and GMM-ERA20C indices are generally able to reproduce the three MJO events during the 2006–2007 extended winter (Fig. 7) and show the same climatological properties (Figs. 5 and 6) as the FMM index, there are undoubtedly errors existing in the GMM index compared to the FMM index. A more objective way to evaluate the accuracy of GMM index is calculating its statistical errors with the FMM index. Following the way of MJO numerical simulation evaluation (Lin et al. 2008; Rashid et al. 2011), we compare the bivariate correlation ($BCORR$, Eq. 2) and bivariate root-mean square error ($BRMSE$, Eq. 3) of GMM index with FMM index of different phases and amplitudes in the period of 1981–2008.

$$BCORR = \frac{\sum_{t=1}^{N} [PC_{1,FMM}(t)PC_{1,GMM}(t)+PC_{2,FMM}(t)PC_{2,GMM}(t)]}{\sqrt{\sum_{t=1}^{N} [PC_{1,FMM}^2(t)+PC_{2,FMM}^2(t)] \sum_{t=1}^{N} [PC_{1,GMM}^2(t)+PC_{2,GMM}^2(t)]}}$$  \hspace{1cm} (Eq. 2)

$$BRMSE = \sqrt{\frac{\sum_{t=1}^{N} ([PC_{1,FMM}(t)-PC_{1,GMM}(t)]^2+[PC_{2,FMM}(t)-PC_{2,GMM}(t)]^2)}{N}}$$  \hspace{1cm} (Eq. 3)

where $PC_{1,FMM}(t)$ and $PC_{2,FMM}(t)$ are the first two leading PCs of the FMM index, $PC_{1,GMM}(t)$ and $PC_{2,GMM}(t)$ are the first two leading PCs of the GMM index, $t$ and $N$ denote time and total number of samples. The larger $BCORR$ is, the more similar the two indices are; on the contrary, a greater value of $BRMSE$ implies larger error of the GMM index.

The statistical errors of the GMM-ERAI index are first examined. As mentioned above, the only difference between the GMM-ERAI index and the FMM index is the OLR input when applying CEOF analyses: the FMM index is based on NOAA OLR observations while the GMM-ERAI index is based on OLR values derived from ERA-I 150-hPa $\nabla x z'$. Table 1 gives the $BCORR$ and $BRMSE$ between the GMM-ERAI and FMM indices. On the whole, if considering all MJO phases and intensities, the GMM-ERAI index is highly similar to the FMM index, with their $BCORR$=0.988 and $BRMSE$=0.225. The $BCORR$ of GMM-ERAI index of all MJO phases on inactive (FMM<1), active (1≤FMM<1.5), very active (1.5≤FMM<2.5) and extremely active (FMM≥2.5) MJO days are 0.968, 0.985, 0.990 and 0.995, respectively; and the corresponding $BRMSE$ are respectively 0.179, 0.212, 0.276 and 0.305.

Kiladis et al. (2014) have compared the similarities among the different MJO indices, and the
of RMM index with OLR-based MJO Index (OMI), filtered MJO OLR (FMO) and VPM indices during winter (DJF) are 0.75, 0.74 and 0.92, respectively. One should bear in mind that every MJO index has its own advantages and disadvantages. The high BCORR (and small BRMSE) between the GMM-ERAI and FMM indices does not mean that the GMM-ERAI index is a better indicator of MJO activities compared to other indices, but implies that applying geopotential-derived OLR anomalies as an input in the MJO index calculation only produces very small differences on the resulted index. Comparing the overall index error (FMM>0) in each phase, the largest errors of GMM-ERAI index are observed in Phases 6–8, where the BCORR are respectively 0.985, 0.986 and the BRMSE are respectively 0.259, 0.251, 0.239; while the smallest errors are observed in Phases 2–4, where the BCORR (BRMSE) are respectively 0.991, 0.990, 0.991 and the BRMSE are respectively 0.194, 0.217, 0.208. Given that the only possible error source of the GMM-ERAI index comes from the derivation of OLR anomalies, the relatively large index errors in Phases 6–8 is likely due to the larger OLR derivation error over the Central Pacific to Western Hemisphere (Figs. 2a and 3b). The EOF1 patterns of the GMM-ERAI and FMM indices (Fig. 4a v.s. 4b) also show obvious differences from the Central Pacific to Western Hemisphere. Meanwhile, the smaller differences over the Indian Ocean and Maritime Continent explain the higher evaluation score in Phases 2–4.

Similar evaluation results are obtained for different categories (inactive, active, very active and extremely active MJO) of MJO days. For active, very active and extremely active MJO days, the GMM-ERAI index has relatively larger errors in Phases 1, 6–8 and smaller errors in Phases 2–5, which can be explained by the lower precision of OLR derivation over the Pacific and Western Hemisphere. Meanwhile, larger errors of the GMM-ERAI index are found in Phases 1, 3 and 8 for inactive MJO days, which is a bit different from the above results, but it is meaningless to discuss phases before MJO is developed. In short, the above evaluation result shows that the GMM-ERAI index is highly similar to the FMM index, confirming the validity of constructing GMM index (or RMM-like index) based on OLR values derived from 150-hPa \( \nabla_x z' \).

We further evaluate the performance of GMM-ERA20C index, which can be extended to the early 20th century but may have larger errors due to the poorer data quality of ERA-20C reanalysis product. As done in above, the differences between GMM-ERA20C and FMM indices are examined by calculating their BCORR and BRMSE (Table 2). Comparing the whole series including all MJO intensities (FMM>0) and phases, the overall BCORR and BRMSE are respectively 0.964 and 0.434,
which is about approximately double that between GMM-ERA1 and FMM indices. The $BCORR$ of GMM-ERA20C index with FMM index is as high as that between VPM and RMM indices (0.92, Kiladis et al. 2014), implying the high similarity between the GMM-ERA20C and FMM index. The $BCORR$ of GMM-ERA20C index of all MJO phases on inactive ($FMM<1$), active ($1 \leq FMM<1.5$), very active ($1.5 \leq FMM<2.5$) and extremely active ($FMM \geq 2.5$) MJO days are 0.873, 0.953, 0.977 and 0.986, respectively; the $BRMSE$ are respectively 0.391, 0.421, 0.479 and 0.549. In general, the error of GMM-ERA20C index is greater than that of GMM-ERA1 index, but is still sufficiently high (small) to conclude that the GMM-ERA20C index could well indicate MJO activities as the FMM index.

Unlike the GMM-ERA1 index, the largest errors of GMM-ERA20C index do not center around the Pacific and Western Hemisphere. As shown in Table 2, the lowest $BCORR$ and largest $BRMSE$ of GMM-ERA20C index are distributed in different MJO phases for different MJO intensities, without noticeable patterns. If considering all MJO intensities, the smallest $BCORR$ are observed in Phases 1, 5 and 7 while the largest $BRMSE$ in Phases 3, 4 and 5. As mentioned above, the errors of GMM-ERA20C index come from both the derivation error of OLR values from z150 data and the relatively lower accuracies of ERA-20C u200, u850 and z150 data. While the former mostly contributes to the errors in Phases 1 and 6–8, the latter could affect all MJO phases. Thus, the random distribution of large GMM-ERA20C index errors in different MJO phases suggests that the precision of 20th century reanalysis data contributes more to the index error.

4.3 Statistical errors of univariate MJO indices based on derived OLR

Previous studies reported that the calculation of RMM index is more sensitive to zonal wind variation (Straub 2013; Liu et al. 2016), which suggests that the validity of geopotential-derived OLR as a substitute of satellite OLR observation could not be concluded solely from the high similarity between the GMM and FMM indices. Straub (2013) found that the bivariate correlations between the full RMM index and an RMM index constructed with OLR removed (i.e., u850 and u200 only) is 0.99 during 1979–2010; and, the full RMM is also highly correlated with the univariate version of the RMM with u200 (correlation coefficient = 0.91) or u850 (correlation coefficient = 0.90). These suggest that OLR added little value to the RMM index beyond what was already determined by the zonal winds.

Hence, in order to justify the use of derived OLR as a component in the MJO index calculation, we further evaluate the accuracy of univariate OLR index based on the OLR derived from 150-hPa $\nabla_x z'$ in
this subsection. Similar to the above analyses, univariate OLR indices are constructed by taking the two
leading modes of univariate EOF on filtered OLR anomalies. The univariate MJO index based on NOAA
interpolated OLR anomalies is considered as a standard reference and is compared with univariate MJO
indices based on OLR anomalies derived from ERA-I 150-hPa $\nabla_x z'$ and ERA-20C 150-hPa $\nabla_x z'$, respectively.

Results show that the EOF modes of OLR derived from both ERA-I and ERA-20C 150-hPa $\nabla_x z'$
are also highly similar to that of observed OLR. The correlation of the two leading EOF modes between
NOAA OLR and ERA-I $\nabla_x z'$ derived OLR are respectively 0.872 and 0.908 (significant at the 99.9%
confidence level) during 1981–2008; while that between NOAA OLR and ERA-20C $\nabla_x z'$ derived OLR
are 0.946 and 0.954, respectively (significant at the 99.9% confidence level). These indicate that the
univariate EOF results of the derived OLR anomalies show consistent MJO convection features with that
of observed OLR. In addition, the time series of time series of univariate OLR indices based on 150-hPa
$\nabla_x z'$ is also highly correlated with that based on NOAA interpolated OLR. The overall $BCORR$ of the
univariate OLR index based on ERA-I 150-hPa $\nabla_x z'$ is 0.829, ranging from 0.729–0.891 for different
phases; that based on ERA-20C 150-hPa $\nabla_x z'$ is 0.847, ranging from 0.825–0.866 for different MJO
phases. The $BCORR$ of the univariate OLR index based on ERA-I 150-hPa $\nabla_x z'$ of all MJO phases on
inactive, active, very active and extremely active MJO days are 0.682, 0.815, 0.847 and 0.887,
respectively; and, the $BCORR$ of the univariate OLR index based on ERA-20C 150-hPa $\nabla_x z'$ of MJO
phases on inactive, active, very active and extremely active MJO days are 0.604, 0.798, 0.896 and 0.927,
respectively (all significant at the 99.9% confidence level). The high correlation coefficients of the EOF
modes and time series of univariate OLR index further justify the ability of geopotential-derived OLR to
capture MJO’s convection features.

5 Reconstructing MJO events in the pre-satellite era

It has been verified in Section 4 that the proposed GMM-ERA20C index is highly similar to the
FMM index even though OLR observation is not used in the index calculation. In other words, one could
reconstruct RMM-like indices (i.e. MJO indices that include both MJO convection and circulation
features) without using observed OLR data. This breaks through previous limitation that MJO indices
are not able to indicate MJO convection activities in the pre-satellite era. In this section, we extend the
GMM-ERA20C index back to the pre-satellite era by applying the same method, and explore MJO events
that occurred in the early 20th century.

Fig. 9 plots the time series of the GMM-ERA20C index intensity during 1902–2008. If an MJO event is defined when the GMM (or FMM) index intensity is greater than 1 for more than 15 consecutive days, there were in total 58 winter MJO events in the satellite era (1981–2008; 2.15 events per year) and 167 events in the pre-satellite era (1902–1978; 2.20 events per year) according to the constructed GMM-ERA20C index. The number of MJO events estimated by the FMM, GMM-ERAI and GMM-ERA20C indices are the same in the satellite era (1981–2008). The strongest MJO event in the pre-satellite era occurred in the 1905–1906 winter, with its maximum intensity reaching 3.41 on 19 February 1906. And, the GMM-ERA20C index intensity shows a small but statistically significant strengthening trend (+0.03 per decade, significant at the 99.9% confidence level) during 1902–2008, suggesting the overall MJO activities have become stronger. Since the main goal of this study is to introduce the method of constructing GMM index and give a comprehensive evaluation on the new index, the long-term trend of pre-satellite MJO activities is not explored in detail here.

The MJO activity during the extended winter of 1905–1906 is examined based on the GMM-ERA20C index as an example (Fig. 10a). The year was selected because the strongest MJO day was recorded on 19 February 1906. According to the GMM-ERA20C index, there were two MJO events with three convective episodes during the extended winter of 1905–1906. (1) The first event was developed before November 1905, and had already reached Phase 5 on 1 November 1905. The MJO maintained its strength in the coming three weeks, traveled through Phases 6–1, and finally dissipated on 22 November 1905. (2) The second MJO event is an example of successive MJO events. The MJO event primarily developed on the boundary of Phase 1 and Phase 2 on 29 December 1905. It then intensified rapidly and the GMM-ERA20C intensity reached above 2 in Phase 3. While the GMM-ERA20C index temporally weakened in Phase 6, it then strengthened again in early February 1906, which indicates the start of another convective episode. It was a strong successive MJO event straight after the second event. The third MJO convective episode reached the very-active MJO category (1.5≤GMM<2.5) very soon after its initiation and reached its maximum as an extremely active MJO (GMM=3.41) in between Phase 1 and Phase 2 on 19 February 1906. It then gradually weakened and dissipated in Phase 6 on 15 March 1906.

The above MJO events indicated by the GMM-ERA20C index is clearly shown in the Hovmöller diagrams of 150-hPa $V_z z'$, u200, u850 anomalies and OLR anomalies derived from z150 (Figs. 10b, c,
d). The 150-hPa $\nabla_x z'$ and OLR anomalies clearly indicate the eastward propagations and strengths of the three MJO convective episodes (see red arrows), where the zonal wind anomalies at the end of January circumnavigate the western hemisphere and connect the anomalous convective signal. However, one might notice that the strongest OLR anomalies do not necessarily occur on the strongest days of GMM-ERA20C index, e.g. the strongest convection of the third MJO event appeared in late February (Fig. 10c) instead of 18 February (Fig. 10a). This is because the GMM index, which is constructed in the same way as RMM index, is more easily influenced by the MJO circulation activities (i.e. u200 and u850) than by convection activities (i.e. OLR) (Liu et al. 2016).

In order to verify the validity of GMM-ERA20C index in the pre-satellite era, we compare the GMM-ERA20C index with the daily historical rainfall observations in Australia. MJO activity is a major factor affecting the rainfall amount over Australia. Concretely, MJO tends to induce positive precipitation anomalies over North Australia during MJO Phases 4–7 (Wheeler and Hendon 2004; Wheeler et al. 2009). Fig. 11 compares the historical rainfall amount observed by stations over the Central North Australia (stations over $0^\circ$–$15^\circ$S and $130^\circ$–$135^\circ$E, including Beatrice Hill, Darwin Botanic Gardens, Elsey, Katherine Council and Pine Creek) and Northeast Australia (stations over $0^\circ$–$15^\circ$S and $140^\circ$–$145^\circ$E, including Coen Post Office, Moreton Telegraph Station and Musgrave) regions (Table 3) in different MJO phases defined by the GMM-ERA20C index. It shows that the overall rainfall, including its mean, median and maximum values, over the Central Northern Australia region peak in MJO Phases 4–5 and that over Northeast Australia in Phases 5–6 during 1902–2008 (Figs. 11a & b). The result is consistent with the previous findings (Wheeler and Hendon 2004; Wheeler et al. 2009), except that the median is slightly higher in Phase 4 than 6 in Fig. 12d. The same result is also true from observations in the pre-satellite era (1902–1978, Figs. 11c & d). The above results imply that the GMM-ERA20C index is able to indicate MJO’s activity, including its convection feature and its impacts, even when satellite data is not available.

6. Conclusions and Discussion

Aiming at extending the data length of historical MJO indices that requires input of convection data, this study introduces a new method to construct RMM-like indices (i.e. MJO indices that include both convection and circulation information) by estimating MJO convection activities in the pre-satellite era based on upper-tropospheric geopotential anomaly. Based on the intrinsic relationship between MJO
convection and upper-level geopotential (i.e. MJO convection center is located in between the upper-
level high and low anomalies) (Adames and Wallace 2014; Leung and Qian 2017), the proposed method
(1) first derives MJO-related OLR anomalies by a linear regression model of OLR and 150-hPa $\nabla_x z'$ and then (2) applies the derived OLR anomalies to the calculation procedures of any MJO indices (such
as the CEOF calculation of RMM index by WH04).

As an example, the proposed method is applied to extend a filtered version of the RMM index
(FMM index), which is based on observed $u_{200}$, $u_{850}$ and OLR input. By replacing the OLR input with
ORL anomalies derived from the ERA-I and ERA-20C reanalyzed $z_{150}$, GMM-ERA1 and GMM-
ERA20C indices are respectively constructed and compared with the original FMM index. The validity
of GMM-ERA1 and GMM-ERA20C indices are evaluated in three aspects: (1) the climatological
properties of the index, (2) statistical errors of the index in the satellite era and (3) the index’s ability to
indicate MJO activities in the pre-satellite era.

(1) The overall evaluation results show that the GMM index is highly similar to the FMM index
in terms of its climatological characteristics and ability to indicate MJO events. Both the GMM-
ERA1 and GMM-ERA20C indices are shown to have the same composite spatial distribution
of OLR and $z_{150}$ anomalies as the FMM index. According to the GMM-ERA20C index data
during 1902–2008, the inactive, active, very active and extremely active MJO days account for
36%, 24%, 35% and 5% of the whole period, being consistent with previous studies (e.g.
Lafleur et al. 2015).

(2) The overall statistical error of the GMM index is notably small. The correlation coefficients of
the EOF1 and EOF2 OLR components between FMM and GMM-ERA1 indices respectively
reach 0.962 and 0.922; and that between FMM and GMM-ERA20C are 0.929 and 0.892. The
overall $BCORR$ of the time series of the GMM-ERA1 and GMM-ERA20C indices are 0.988
and 0.964, respectively, compared to the FMM index; and the $BRMSE$ are respectively 0.225
and 0.434. The $BCORR$ of GMM-ERA1 and GMM-ERA20C index ranges from 0.985–0.991
and 0.959–0.968 for different MJO phases. The high $BCORR$ and low $BRMSE$ confirm that the
proposed method in the paper could reproduce an MJO index highly similar to the FMM index.

(3) The validity of using geopotential-derived OLR as an indicator of MJO convection is further
examined by evaluating the univariate OLR index based on the derived OLR anomalies. The
overall $BCORR$ of the univariate OLR index derived from ERA-I and ERA-20C 150-hPa $\nabla_x z'$
are respectively 0.829 and 0.847, compared to that derived from observed OLR; the corresponding BCORR ranges from 0.729–0.891 and 0.825–0.866 for different MJO phases. These results confirm the ability of geopotential-derived OLR to indicate MJO convection.

(4) Comparison between the GMM-ERA20C index and historical rainfall records (1902–2008) in Australia shows that MJO tends to induce more precipitation over Central North Australia (around Darwin, stations over 0°–15°S and 130°–135°E) and Northeast Australia (around Cape York Peninsula, stations over 0°–15°S and 140°–145°E) regions during MJO Phases 4–5 and 5–6, respectively. The result is consistent with previous studies (Wheeler and Hendon 2004; Wheeler et al. 2009) and thus verifies the ability of the GMM-ERA20C index to indicate MJO activity in both the modern satellite era and pre-satellite era.

In Section 4.2, statistical evaluation results show high BCORR and low BRMSE between the GMM and FMM indices. Their BCORR is higher than that of RMM index with OMI, FMO and VPM indices, whose values are 0.75, 0.74 and 0.92, respectively (Kiladis et al. 2014). A reason for the high similarity between GMM and FMM indices is that filtered anomalies are used in the above analyses, which exclude information of high-frequency variability. The high correlation does not mean that the GMM index is a better indicator of MJO activities compared to other indices, but suggests the validity and high accuracy of replacing observed OLR anomalies with OLR derived from 150-hPa $\nabla x z'$ when constructing RMM-like indices. The error of the GMM-ERA20C index could come from the derivation of OLR anomalies from 150-hPa $\nabla x z'$ and the data quality of ERA-20C reanalysis product. Our results show that the latter factor may produce larger errors than the former factor. Nevertheless, the small error of GMM-ERA20C index demonstrates the feasibility of constructing reliable RMM-like indices over the pre-satellite era.

According to the long-term trend of GMM-ERA20C index, the overall MJO activities have become more active during 1902–2008, in terms of both the index intensity and number of MJO events. Since the main target of this study is introducing a new method of constructing RMM-like indices, the long-term trend of pre-satellite MJO activities is not given in detail here and will be discussed in our future work.

The long-term behavior of the MJO in the pre-satellite era is an important topic and deserves more attention in scientific literature. Meanwhile, the short record of OLR observation limits research on the historical long-term variability of MJO. The above evidence confirms the validity of the GMM index and the proposed MJO index construction method. The proposed method can be applied to any MJO indices that require input of OLR data, such as the OMI and FMO indices, by simply replacing the OLR
data with that derived from 150-hPa $\nabla_x z'$. Making use of the proposed method, one could extend the records of MJO indices, especially that involve MJO convection information, to the pre-satellite era and study the long-term trend and inter-decadal variability of MJO activity which can hardly be done before due to the lack of satellite observations. This study provides an alternative way overcoming the difficulty of historical MJO studies, and will be beneficial to our understanding of long-term variability of MJO.

Declarations

Funding. This work is funded and supported by the Guangdong Basic and Applied Basic Research Foundation (2020A1515110275), the Guangdong Province Introduction of Innovative R&D Team Project (2019ZT08G669), National Natural Science Foundation of China (41775067), Science and Technology Planning Project of Guangdong Province, China (2018B020208004), Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), National Key Research and Development Program of China (2018YFC1505906) and State Oceanic Administration International Cooperation Program (GASI-IPOVAI-03).

Conflict of interest. The author declares that there is no competing interest.

Availability of data and material. The interpolated OLR data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/ in any documents or publications using these data. The ERA Interim and ERA-20C reanalysis data can be access from the ECMWF data archive at https://apps.ecmwf.int/datasets/data/interim-full-daily/ and https://apps.ecmwf.int/datasets/data/era20c-daily/, respectively. The historical rainfall archive of GHCND is available at https://www.ncdc.noaa.gov/ghcnd-data-access.

References

Adames ÁF, Wallace JM (2014) Three-Dimensional Structure and Evolution of the MJO and Its Relation to the Mean Flow. J Atmos Sci 71:2007–2026. https://doi.org/10.1175/JAS-D-13-0254.1

Chen X, Li C, Tan Y (2017) Further inquiry into characteristics of MJO in boreal winter. Int J Climatol 37:4451–4462. https://doi.org/10.1002/joc.5098
Dee DP, Uppala SM, Simmons AJ, et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597. https://doi.org/10.1002/qj.828

Durre I, Menne MJ, Gleason BE, et al (2010) Comprehensive Automated Quality Assurance of Daily Surface Observations. J Appl Meteor Climatol 49:1615–1633. https://doi.org/10.1175/2010JAMC2375.1

Durre I, Menne MJ, Vose RS (2008) Strategies for Evaluating Quality Assurance Procedures. J Appl Meteor Climatol 47:1785–1791. https://doi.org/10.1175/2007JAMC1706.1

European Centre for Medium-Range Weather Forecasts ERA-20C Project (ECMWF Atmospheric Reanalysis of the 20th Century). Res Data Arch Natl Cent Atmos Res Comput Inf Syst Lab. https://doi.org/10.5065/D6VQ30QG

Gill AE (1980) Some simple solutions for heat induced tropical circulation. Q. J. R. Meteorol. Soc. 106:447–462

Henderson SA, Maloney ED, Son S-W (2017) Madden–Julian Oscillation Pacific Teleconnections: The Impact of the Basic State and MJO Representation in General Circulation Models. J Clim 30:4567–4587. https://doi.org/10.1175/JCLI-D-16-0789.1

Hsu P, Li T (2012) Role of the Boundary Layer Moisture Asymmetry in Causing the Eastward Propagation of the Madden–Julian Oscillation*. J Clim 25:4914–4931. https://doi.org/10.1175/JCLI-D-11-00310.1

Jones C, Carvalho LM V (2006) Changes in the activity of the Madden-Julian oscillation during 1958-2004. J Clim 19:6353–6370. https://doi.org/10.1175/JCLI3972.1

Kalnay E, Kanamitsu M, Kistler R, et al (1996) The NCEP/NCAR 40-Year Reanalysis Project. Bull
Kiladis GN, Dias J, Straub KH, et al (2014) A Comparison of OLR and Circulation-Based Indices for Tracking the MJO. Mon Weather Rev 142:1697–1715. https://doi.org/10.1175/MWR-D-13-00301.1

Kiladis GN, Thorncroft CD, Hall NMJ (2006) Three-Dimensional Structure and Dynamics of African Easterly Waves. Part I: Observations. J Atmos Sci 63:2212–2230. https://doi.org/10.1175/JAS3741.1

Lafleur DM, Barrett BS, Henderson GR (2015) Some climatological aspects of the Madden-Julian Oscillation (MJO). J Clim 28:6039–6053. https://doi.org/10.1175/JCLI-D-14-00744.1

Lau WKM, Waliser DE, Wang B (2012) Theories - Intraseasonal Variability in the Atmosphere-Ocean Climate System. In: Lau WK-M, Waliser DE (eds). Springer Berlin Heidelberg, Berlin, Heidelberg, pp 335–398

Leung JC-H (2019) Extraction of Madden-Julian Oscillation Convection Signals from Geopotential Anomalies. Peking University

Leung JC-H, Qian W (2017) Monitoring the Madden–Julian oscillation with geopotential height. Clim Dyn 49:1981–2006. https://doi.org/10.1007/s00382-016-3431-x

Liebmann B, Smith CA (1996) Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset. Bull Am Meteorol Soc 77:1275–1277

Lin H, Brunet G, Derome J (2008) Forecast Skill of the Madden–Julian Oscillation in Two Canadian Atmospheric Models. Mon Weather Rev 136:4130–4149.

https://doi.org/10.1175/2008MWR2459.1
Liu P, Zhang Q, Zhang C, et al (2016) A Revised Real-Time Multivariate MJO Index. Mon Weather Rev 144:627–642. https://doi.org/10.1175/MWR-D-15-0237.1

Madden R a., Julian PR (1972) Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period. J. Atmos. Sci. 29:1109–1123

Madden RA, Julian PR (1971) Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. J. Atmos. Sci. 28:702–708

Maloney ED, Adames ÁF, Bui HX (2019) Madden–Julian oscillation changes under anthropogenic warming. Nat Clim Chang 9:26–33. https://doi.org/10.1038/s41558-018-0331-6

Menne MJ, Durre I, Vose RS, et al (2012) An Overview of the Global Historical Climatology Network-Daily Database. J Atmos Ocean Technol 29:897–910. https://doi.org/10.1175/JTECH-D-11-00103.1

Nakazawa T (1988) Tropical Super Clusters within Intraseasonal Variations over the Western Pacific. J Meteorol Soc Japan Ser II 66:823–839. https://doi.org/10.2151/jmsj1965.66.6_823

Oliver ECJ, Thompson KR (2012) A Reconstruction of Madden–Julian Oscillation Variability from 1905 to 2008. J Clim 25:1996–2019. https://doi.org/10.1175/JCLI-D-11-00154.1

Poli P, Hersbach H, Dee DP, et al (2016) ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. J Clim 29:4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1

Rashid HA, Hendon HH, Wheeler MC, Alves O (2011) Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. Clim Dyn 36:649–661. https://doi.org/10.1007/s00382-010-0754-x

Roundy PE (2020) Interpretation of the spectrum of eastward-moving tropical convective anomalies. Q J R Meteorol Soc 146:795–806. https://doi.org/https://doi.org/10.1002/qj.3709
Roundy PE, Schreck III CJ (2009) A combined wave-number--frequency and time-extended EOF approach for tracking the progress of modes of large-scale organized tropical convection. Q J R Meteorol Soc 135:161–173. https://doi.org/https://doi.org/10.1002/qj.356

Slingo JM, Rowell DP, Sperber KR, Nortley F (1999) On the predictability of the interannual behaviour of the Madden-Julian oscillation and its relationship with el Niño. Q J R Meteorol Soc 125:583–609. https://doi.org/https://doi.org/10.1002/qj.49712555411

Straub KH (2013) MJO Initiation in the Real-Time Multivariate MJO Index. J Clim 26:1130–1151. https://doi.org/10.1175/JCLI-D-12-00074.1

Uppala SM, Källberg PW, Simmons AJ, et al (2005) The ERA-40 re-analysis. Q J R Meteorol Soc 131:2961–3012. https://doi.org/https://doi.org/10.1256/qj.04.176

Ventrice MJ, Thornicroft CD, Roundy PE (2011) The Madden–Julian Oscillation’s Influence on African Easterly Waves and Downstream Tropical Cyclogenesis. Mon Weather Rev 139:2704–2722. https://doi.org/10.1175/MWR-D-10-05028.1

Ventrice MJ, Wheeler MC, Hendon HH, et al (2013) A Modified Multivariate Madden–Julian Oscillation Index Using Velocity Potential. Mon Weather Rev 141:4197–4210. https://doi.org/10.1175/MWR-D-12-00327.1

Wang B, Chen G (2017) A general theoretical framework for understanding essential dynamics of Madden–Julian oscillation. Clim Dyn 49:2309. https://doi.org/10.1007/s00382-016-3448-1

Wang B, Rui H (1990) Dynamics of the Coupled Moist Kelvin–Rossby Wave on an Equatorial β Plane. J. Atmos. Sci. 47:397–413

Wang L, Li T (2017) Roles of convective heating and boundary-layer moisture asymmetry in slowing down the convectively coupled Kelvin waves. Clim Dyn 48:2453–2469.
Wheeler M, Kiladis GN (1999) Convectively Coupled Equatorial Waves: Analysis of Clouds and Temperature in the Wavenumber–Frequency Domain. J Atmos Sci 56:374–399. https://doi.org/10.1175/1520-0469(1999)056<0374:CCEWAO>2.0.CO;2

Wheeler MC, Hendon HH (2004) An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. Mon Weather Rev 132:1917–1932. https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2

Wheeler MC, Hendon HH, Cleland S, et al (2009) Impacts of the Madden–Julian Oscillation on Australian Rainfall and Circulation. J Clim 22:1482–1498. https://doi.org/10.1175/2008JCLI2595.1

Wolding BO, Maloney ED (2015) Objective Diagnostics and the Madden–Julian Oscillation. Part I: Methodology. J Clim 28:4127–4140. https://doi.org/10.1175/JCLI-D-14-00688.1

Zhang C (2013) Madden–Julian Oscillation: Bridging Weather and Climate. Bull Am Meteorol Soc 94:1849–1870. https://doi.org/10.1175/BAMS-D-12-00026.1
Figures

(a) 150-hPa $\nabla \times z(91.5^\circ E)$ vs OLR($90^\circ E$)
max corr = -0.893 (99.9%) (ERA-I)

(b) 150-hPa $\nabla \times z(111^\circ E)$ vs OLR($115^\circ E$)
max corr = -0.872 (99.9%) (ERA-I)

(c) 150-hPa $\nabla \times z(83.25^\circ E)$ vs OLR($90^\circ E$)
max corr = -0.823 (99.9%) (ERA-20C)

(d) 150-hPa $\nabla \times z(103.5^\circ E)$ vs OLR($115^\circ E$)
max corr = -0.775 (99.9%) (ERA-20C)

Figure 1

Scatter plots of (a) $\nabla \times z(90^\circ E)$ (unit: W m$^{-2}$) and (b) $\nabla \times z(115^\circ E)$ (unit: W m$^{-2}$) against their corresponding most correlated 150-hPa $\nabla \times z$ (unit: 10-6 gpm m$^{-1}$) during November of 1981–2008 based on ERA-I. (c) and (d) are the same as (a) and (b) except against 150-hPa $\nabla \times z$ (unit: 10-6 gpm m$^{-1}$) based on ERA-20C. The red dashed line is the regression line of each case. Noted that the longitudes of 150-hPa $\nabla \times z$ are
different for different  and datasets. Correlation coefficients of all panels are shown in the figure title and are all statistically significant at the 99.9% confidence level.

**Figure 2**

The maximum correlation coefficient of OLR anomaly at each longitude ( ) with its corresponding 150-hPa ' ( ) during extended winter (NDJFM) of 1981–2008. OLR anomaly are based on NOAA gridded interpolated OLR data. 150-hPa ' ( ) are derived from (a) ERA-I and (b) ERA-20C reanalysis data, respectively. Black dots denote correlation coefficients reaching the 99.9% confidence level. The numbers on the y axis denote each month of the extended winter (NDJFM).
Figure 3

Hovmöller diagrams of OLR obtained from (a) NOAA interpolated OLR product (shading, 4 W m⁻² 32 interval), as well as that derived from (b) ERA-I (shading, 4 W m⁻² interval) and (c) ERA-20C 150-hPa (shading, 3 W m⁻² interval) during the 2006–2007 extended winter. (d) and (e) are the Hovmöller diagrams of 150-hPa (shading, 2×10⁻⁶ gpm m⁻¹ interval) derived from (d) ERA-I and (e) ERA-20C. Arrows denote the two MJO events during the 2006–2007 extended winter.
The 15°S–15°N OLR (black solid line), u200 (blue dotted line) and u850 (red dashed line) components of the two leading modes of CEOF based on three combination sets of MJO-filtered variables: (a) NOAA gridded interpolated OLR + ERA-I u200 + ERA-I u850 (Set 1, FMM index), (b) OLR derived from ERA-I 150-hPa + ERA-I u200 + ERA-I u850 (Set 2, GMM-ERA1 index) and (c) OLR derived from ERA-20C 150-hPa + ERA-20C u200 + ERA20C u850 (Set 3, GMM-ERA20C index). CEOF analyses were performed over extended winter during 1981–2
Figure 5

GMM-ERA20C index of (a) the extended winter, (b) November, (c) December, (d) January, (e) February and (f) March during 1902–2008. Grey, green, blue and red dots respectively denote inactive (IA, GMM<1.0), active (A, 1.0≤GMM<1.5), very active (VA, 1.5≤GMM<2.5) and extremely active (EA, GMM≥2.5) MJO days, following Lafleur et al. (2015).
Figure 6

Composites of NOAA interpolated OLR (shading, 5 W m\(^{-2}\) 52 interval) and ERA-I z150 (contour, 4 gpm interval) anomalies of different MJO phases defined based on (a) FMM, (b) GMM-ERAi and (c) GMM-ERA20C indices during 1981–2008. Figure titles of panels (b) and (c) give the pattern correlations of OLR (corr_olr) and z150 (corr_z), compared with that of panel (a). All correlation coefficients are statistically significant at the 99.9% confidence level.
Figure 7

The (a) FMM, (b) GMM-ERAi and (c) GMM-ERA20C indices during the 2006–2007 extended winter. Orange, red, blue, light purple, deep purple dots denote index points respectively in November, December, January, February and March.

Figure 8

Hovmöller diagrams of 15°S–15°E u850 (shading, 0.4 m/s interval) and u200 (contour, 2 m/s interval) anomalies obtained from (a) ERA-I and (b) ERA-20C during the 2006–2007 extended winter.
Figure 9

Daily time series of GMM-ERA20C intensity during (a) 1902–1928, (b) 1929–1954, (c) 1955–1980 and that of FMM, GMM-ERAI and GMM-ERA20C intensities during (d) 1981–2008. Blue, red and green lines denote series of GMM-ERA20C, GMM-ERAI and FMM indices, respectively.
Figure 10

(a) GMM-ERA20C index during the 1905–1906 extended winter. Orange, red, blue, light purple, deep purple dots denote index points respectively in November, December, January, February and March. (b), (c) and (d) are the same as Figs. 3e, 3c and 8b except during the 1905–1906 extended winter. Arrows denote the three MJO convection episodes during the 1905–1906 extended winter.
Boxplots of historical rainfall observations (unit: mm/day) of different GMM-ERA20C phases during 1902–2008 over (a) Central North Australia (0°–15°S, 130°–135°E) and (b) Northeast Australia (0°–15°S, 140°–145°E). (c) and (d) are the same as (a) and (b) except during 1902–1978. Noted that only rainfall data on active MJO days (GMM-ERA20C intensity > 1) is included in the calculation. Orange solid lines and blue dotted lines denote the median and mean values of rainfall in each GMM-ERA20C phase, respectively. The black dotted line is the average rainfall over all GMM-ERA20C phases.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.jpg
- Table2.jpg
- Table3.jpg