Wang Lu (Orcid ID: 0000-0002-4811-8847)

# Can reanalysis products with only surface variables assimilated capture MJO characteristics?

Jingxuan Cui[1,2], Lu Wang[1,2], Tim Li[3,1] and Bo Wu[2]

1.Key Laboratory of Meteorological Disaster, Ministry of Education (KLME) / Joint International Research Laboratory of Climate and Environmental Change (ILCEC) / Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing, China

2. LASG, Institute of Atmospheric Physics (IAP) Chinese Academy of Sciences (CAS), Beijing, China

3. International Pacific Research Center and Department of Atmospheric Sciences, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii 96822

Revised on July 16, 2019

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/joc.6270

Corresponding author: Dr. Lu Wang, Nanjing University of Information Science and Technology, Nanjing, China. Email: luwang@nuist.edu.cn



## Abstract

The Madden-Julian Oscillation (MJO), as a dominant mode of tropical intraseasonal oscillation, plays an important role in the variability of global weather and climate. However, current state of the art atmospheric circulation models have difficulty in reproducing observed MJO characteristics when forced by observed daily sea surface temperature alone. An important practical question is how much data a model needs in assimilation in order to reproduce real MJO events? By analyzing ERA-20C and NOAA-20CR reanalysis data, the authors tried to figure out whether a model could reproduce observed MJO events by assimilating the observed surface signal alone.

The phase propagation and vertical structure associated with MJO were compared between the reanalysis data and observations during 1979-2010. A total skill score considering both temporal correlation and spatial standard deviation were defined. The result showed that both ERA-20C and NOAA-20CR could reproduce the observed MJO characteristics very well, with the former superior to the latter, regardless of MJO intensity. Thus, a minimum requirement for an operational atmospheric model for MJO prediction is the assimilation of the observed surface signals.

**Key words:** ERA and NOAA 20C-reanalysis datasets, Madden-Julian Oscillation, data assimilation

## 1 Introduction

The Madden-Julian Oscillation (MJO) is the dominant mode of intraseasonal variability in the tropics (Madden and Julian, 1972; Zhang, 2005; see a recent book chapter review by Li and Hsu 2017). This intraseasonal oscillation with a period of 10-90 days and planetary zonal scale (Li and Zhou, 2009) is initiated over the western equatorial Indian Ocean (Zhao et al., 2013) and is characterized by large-scale convective system propagating slowly eastward along the equator. The MJO has great seasonality and modulates not only tropical but also mid-latitude circulation on various spatial and temporal scales (Li, 2014; Wang et al., 2013). For example, it can affect the intensity, onsets and breaks of the Indian and Australian monsoons (Yasunari, 1979; Li and Wang 2005; Wheeler and McBride, 2005; Hsu and Yang, 2016), the onsets of El Niño events (Kessler and Kleeman, 2000; Chen et al., 2016a, b; Liu et al., 2016), and tropical cyclone genesis (Liebmann et al., 1994; Li et al., 2018).

The vertical structure of anomalous vertical motion associated with MJO exhibits a westward tilting structure (Hendon and Salby, 1994; Wang and Li, 1994). While a first baroclinic mode vertical structure appears over the MJO convective center, a second baroclinic mode structure appears to the east and west of the deep convection (e.g., Majda and Biello, 2004; Mapes et al., 2006; Wang et al., 2017). The tilting is caused by the phase leading of boundary layer convergence (Li and Wang, 1994; Maloney and Hartmann, 1998; Liu and Wang, 2012,2017). In addition, anomalous moisture also exhibits a westward tilting structure with height (Hsu and Li, 2012). A moisture budget analysis indicated that this moisture phase leading results from advection of mean moisture by anomalous ascent motion associated with boundary layer convergence (Hsu and Li, 2012). The horizontal pattern of MJO low-level circulation exhibits a Kelvin-Rossby wave couplet (e.g., Hendon and Salby, 1994;

Wang and Li, 1994; Wang et al. 2018), with two cyclonic gyres residing at both sides of the equator to the west of the MJO convective center, and pronounced easterlies to the east of the convective center.

Numerous researchers have endeavored to obtain good simulations of the MJOs, but faced a big challenge (Slingo et al., 1996; Lin et al., 2006; Kim et al., 2009; Weaver et al., 2011; Hung et al., 2013; Jiang et al., 2015). Even a perfect atmospheric general circulation model (AGCM) could not reproduce the observed MJO events if it were forced by the observed daily sea surface temperature (SST) alone. This is because the MJO is in between synoptic time scale and seasonal time scale, and both the boundary layer condition and the initial condition are important to yield a real MJO (e.g., Ulate Medrano, 2015; Yang and Wang, 2018). To generate a long-term dataset comparable to the observed MJO, a data assimilation system is needed which could assimilate atmospheric observations into an AGCM frequently. The generated dataset is considered accurate enough to be applied in the MJO forecast experiments as the initial field. Previous studies have shown that the quality of the initial field plays a critical role in MJO forecast by dynamical models (e.g., Fu et al., 2013; Fu et al., 2015). With the diversifying of methods for observation, more and more observational data are available in data assimilation. However, it has been pointed out that the simulated results may not be improved with the increase of the amount of assimilated observations (Huntley et al., 2010). Moreover, the more fields to be assimilated would cost more computing resources and need more complicated assimilation methods. Therefore, a question with great value is how much data, at least, does a model need in assimilation to produce the real MJO events?

Reanalysis datasets, which are aimed to provide long-term datasets comparable to observations for studying the climate system variabilities, are generated by data

assimilation systems. Some atmospheric reanalysis datasets began to assimilate station-observations since 1950s, such as the 40-year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (Uppala et al., 2005), the Japanese 55-year Re-Analysis (JRA-55) (Kobayashi et al., 2015; Harada et al., 2016) developed by the Japan Meteorological Agency (JMA), and the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis-I (Kalnay et al., 1996), in which the station-observations include the surface pressure, temperature and humidity, wind snow depth and so on from land- and ship-stations. Since satellite observations could be obtained after 1979, updated data assimilation systems were developed and reanalysis datasets such as the Interim European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-Interim) (Dee et al., 2011), Japanese 25-year Re-Analysis (JRA-25) (Onogi et al., 2007) and NCEP/NCAR Reanalysis-II (Kanamitsu et al., 2002) were released. The new observations, including temperature-sensitive microwave radiances, total $O_3$ from the Total Ozone Mapping Spectrometer (TOMS) and $O_3$ from the Solar Backscattered Ultra Violet (SBUV), total column water vapor and surface wind-speed over ocean from the Special Sensor Microwave/Imager (SSM/I) and so on were employed. It has been recognized that ERA-Interim dataset is the best reanalysis dataset for representing the MJO phenomenon (e.g., Tian et al., 2010; Jiang et al., 2011; Kim et al., 2014) and therefore it is often used as the observation. In recent years, in order to study the climate variability for a longer period, the National Oceanic and Atmospheric Administration (NOAA) and the ECMWF released reanalysis datasets covering the whole $20^{th}$ century, which are named NOAA-20CR and ERA-20C respectively. Since the observations are few in early $20^{th}$ century, only the observed daily surface pressure field was assimilated to produce NOAA-20CR and surface marine winds was additionally assimilated in ERA-20C. It provides a

good opportunity to investigate the question that whether the atmospheric surface data is enough for assimilation to generate the observed MJO. To address this question, we may compare the MJO characteristics derived from observation and the two 20C-reanalysis datasets.

By now, few researchers have carefully evaluated the MJO performances in NOAA-20CR and ERA-20C. Gao et al. (2016) compared the evolution of sea surface latent heat flux anomalies associated with two MJO events during DYNAMO period, when an international field campaign took place over the Indian Ocean during late 2011 to early 2012, between NOAA-20CR and buoy-observation. It is found that the observational MJO propagation phase in one event can be captured in the reanalysis data, while it cannot be captured in the other. As the samples were too few, this conclusion is not enough to verify whether only assimilating atmospheric surface data can succeed or fail to generate the observed MJOs.

In this paper, we will evaluate the MJO characteristics, especially their phase propagation and vertical structure generated by the NOAA-20CR and ERA-20C. The results are expected to provide a reference for future assimilation experiments. The rest of this paper is organized as follows. In Section 2, data and methods are introduced. The comprehensive comparison of the MJO propagation phases and other characteristics of the MJO vertical structure in different type of MJO events derived from the two reanalysis products is described in Section 3. Finally, a brief summary and discussions are given in the last section.

## 2 Data and Method

This study employed reanalysis data of NOAA-20CR (Compo et al., 2011),

ERA-20C (Poli et al., 2016) and ERA-Interim during the period of 1979-2010, with variables such as zonal wind field, vertical velocity, temperature and specific humidity. The NOAA Interpolated Outgoing Longwave Radiation (OLR) data product is also used for the same period. All the datasets are archived on 2.5°×2.5° grid points, with a daily interval.

For ERA-Interim, the assimilated data includes regular observations such as surface pressure, 2-meter temperature, and 2-meter relative humidity from ships, drifting buoys, and land stations, and scatterometer ambiguous wind vectors, the satellite brightness temperature radiance observations, the satellite ozone retrievals and atmospheric motion state retrievals and observations collected by means of the GPS radio occultation (GPSRO) technique (Dee et al., 2011). For NOAA-20CR and ERA-20C, the station-observed surface pressure from the Locations of Stations in International Surface Pressure Databank (ISPD) version 3 is assimilated and additionally surface marine wind field from The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) version 2.5.1 is assimilated in ERA-20C, so we can roughly consider that the observational information contained in the two sets of 20C-reanalysis data is only from the surface of atmosphere. In the following evaluation, ERA-Interim data will be used as the true observation value and the base for events selection.

It is mentioned that the latest version of NOAA-20CR (version2c) is used in this study, which contains 56 reanalysis members and their ensemble product. The position of the MJO convection simulated by 56 members cannot be completely consistent at any given time, so the method of ensemble-mean may smooth out some MJO-related signals and affect the analysis of convection position. Therefore, instead of using the ensemble-mean product, we randomly select a member of the

NOAA-20CR V2c to make an analysis on the features of the MJO.

To evaluate the capability of 20C-reanalysis datasets in MJO simulation, the Lanczos band-pass filter (Duchon, 1979), multi-variate Empirical Orthogonal Function (EOF) analysis, the Real-time Multivariate MJO (RMM) index (Wheeler and Hendon, 2004) and correlation analysis are employed. The RMM index was developed to show the whole process of the MJO eastward propagation starting from the Western Indian Ocean and dying out in the Central Pacific (near the date line). It is based on a leading pair of EOFs obtained by applying a multi-variable-EOF (MV-EOF) analysis to 20-100-day bandpass filtered OLR and 850hPa and 200hPa zonal wind anomalies (U850 and U200) averaged over the equatorial region (15°N-15°S). This index can be used to describe the intensity and longitudinal position of the MJO at a given time.

Although the traditional RMM index was obtained by MV-EOF of three variables (i.e., OLR, U200 and U850), we find that the MV-EOF results of only two variables (i.e., U200 and U850) resemble the three-variable result very well (figures not shown). This is because the MV-EOF result is mostly explained by the circulation instead of the convection (Straub, 2013). Therefore, in this study we calculate RMM indices by MV-EOF of only U200 and U850. The observational RMM index is calculated by ERA-Interim data.

Since model simulations may have different performance in the strong and weak MJO signals, we investigate the results for the extremely strong MJO events and the ordinary MJO events respectively. The extremely strong MJO event is defined as when the observational RMM index derived from ERA-Interim dataset is greater than 2.5 standard deviation for 5 consecutive days, and the maximum value is greater than 3 standard deviation. The ordinary MJO event is defined as when the observational

RMM index is greater than 1 for continuous 5 days. Based on this criterion, 8 extremely strong MJO events and 98 ordinary MJO events were selected. All the selected events are the common parts satisfying the criterion derived from 2-variable and 3-variable RMM indices. The length of each event is taken as 62 days. The extremely strong MJO events are listed in Table 1.

To evaluate the MJO representation in the two 20C-reanalysis sets, in terms of the horizontal propagation and vertical structure, various metrics are employed in this study. It includes spatial correlation coefficients, standard deviation and root mean square error (RMSE). We also define skill scores, considering both pattern correlation coefficient and the spatial standard deviation, to measure the MJO convective intensity and the westward-tilting structure.

## 3 Results

### 3.1 Characteristics in strong MJO events

We first examined the phase propagation features of the 8 extremely strong MJO events. Figures 1 and 2 show the Hovmueller-diagrams of U850 and U200 averaged over 20°S-20°N for each of the extremely strong MJO events respectively. In the 8 MJO events, the zonal wind fields of the two sets of 20C-reanalysis look very similar to the observation. The correlation coefficient in U850 between NOAA-20CR and the observation is above 0.83, and the maximum is as high as 0.94. The correlations based on U200 are not as good as those of U850, but they are still higher than 0.7. The spatial correlations of U850 between ERA-20C and the observation are greater than 0.97, while those of U200 are above 0.87. Then, we define a propagation skill score (PSS) by the average of correlation coefficients obtained by U200 and U850, as

shown in Table 2. It is found that the averaged propagation skill of 8 events for ERA-20C is 0.96, while that for NOAA-20CR is 0.87.

To further investigate how accurate the two 20C-reanalysis data could reproduce the phase evolution of the MJOs, we compared the RMM indices directly. Figure 3 shows the propagation phase of each extremely strong MJO event in the two-dimensional phase space defined by RMM1 and RMM2 derived from ERA-20C and the observation. In each phase space, the anticlockwise line represents propagation process of the MJO, while the distance from the point on the line to original point represents the MJO intensity at that time. The line for ERA-20C is almost identical with the one for observation in any event. The results of NOAA-20CR are similar, and are not shown here. It can be seen intuitively that the 20C-reanalysis data has a good representation on the MJO propagation phase. In the whole process of convection development, the reanalysis and observational results are basically the same whether it is at the initial or ending position or during the propagation. Then, we quantitatively examined the correlation of the two-dimensional spatial phase in the eight events and the RMSE of the intensity between 20C-reanalysis and the observation, and the results are shown in Figure 4. As the figure shows, the correlation coefficients between ERA-20C and observation are high, with the minimum 0.67, which proves the good representation of the MJO propagation phase in statistics. The correlations of NOAA-20CR are not so high as those of ERA-20C, but are still higher than the 5% significant level. Significant spatial phase correlation and the small RMSE are seen in all events in ERA-20C and NOAA-20CR, indicating that both the 20C-reanalysis data are good at simulating the phase propagation and intensity of the extremely strong MJOs. The performance of ERA-20C is better than that of NOAA-20CR, which may be due to the additional surface marine winds assimilated in ERA-20C or due to the fact that ERA-20C and

ERA-Interim products were generated using similar atmospheric models and assimilation approaches (4-dimensional variational analysis).

Then, we examined how well do the two 20C-reanalysis sets reproduce the vertical structure of the observed MJOs, because it is of interest to know if a model could capture the MJO structures throughout the column by assimilating observed surface signals alone (i.e., only the surface pressure for the two 20C-reanalysis and additional surface marine winds for ERA-20C). The fields we assess below include water vapor anomalies, zonal wind anomalies, vertical velocity anomalies, divergence anomalies, equivalent potential temperature (EPT) anomalies, diabatic heating anomalies and available potential energy (APE) generation associated with the MJO convection. The last three metrics were based on the newly proposed dynamics-oriented diagnostics of the MJO by Wang et al. (2018). To compare the MJO vertical structures between the reanalysis and the observation, in the following we show the longitude-vertical cross-sections of each evaluated field averaged over 15°S-15°N composite for the 8 extremely strong MJO events, in which 0° in the x-axis denotes where the MJO convection maximum center is located.

Figures 5a and 5b display the composite results of zonal-vertical cross sections of specific humidity, zonal wind, divergence and vertical velocity anomalies for the observation, ERA-20C and NOAA-20CR. In observation, the maximum moisture center is at 700-500hPa, with a westward-tilting structure with height, accompanied by low-level convergence and high-level divergence. The maximum ascending motion center is at 300-400hPa, and a phase-leading horizontal convergence is shown in the mid- and low-level. As one can see, the gross features of the observed vertical structure of each field are well captured by the 20C-reanalysis. In order to quantitatively measure the performance of the 20C-reanalysis in reproducing the

vertical structures of the MJOs, we calculated the spatial correlation coefficients of each field between the observation and the 20C-reanalysis based on plots of the longitude-vertical cross sections (result shown in Table 3). It is found that the correlations of the water vapor anomalies, vertical velocity anomalies and divergence anomalies are all above 0.7, indicating that both 20C-reanalysis could well capture the gross features of the vertical structures of MJOs. Furthermore, ERA-20C has better performance than NOAA-20CR, with the former showing correlations more than 0.9 while the latter showing correlations less than 0.9.

Then, we compared the intensity of each field near the MJO convective center between 20C-reanalysis and the observation. As is seen, although the typical westward tilting structure with advanced low-level water vapor can be found in the two sets of 20C-reanalysis, the intensity obtained by 20C-reanalysis is weaker than the observation, and the convection develops slightly lower in NOAA-20CR. We defined the intensity score by the mean value of 700-500hPa field averaged on five longitudes to the east and west of the convective center. As shown in Table 4, the intensity scores of specific humidity for ERA-20C and NOAA-20CR are 0.76g/kg and 0.67g/kg respectively, which are less than the observed 1.03g/kg. But the intensity scores of vertical velocity for ERA-20C and NOAA-20CR are a bit more than the observed value. We also defined a score to measure the tilting structure in the observation and the 20C-reanalysis. It is calculated by the integral field at the low atmosphere (850-500hPa) between 5 to 40 longitudes at the east side of the convection (see results in Table 4). The results obtained from the water vapor and vertical velocity all show that the score is less in ERA-20C and NOAA-20CR than the observation, indicating that there is a deficiency in the representation of phase-leading structure in 20C-reanalysis.

The results illustrate that the atmospheric surface signals can restrain the vertical structure of the MJO to some extent, but the details are still inaccurate, such as the obvious less moisture associated with the MJO convection. Moreover, the higher scores in ERA-20C also demonstrate that the surface marine wind field is a useful variable in the assimilation to reproduce the real MJO rather the observed surface pressure alone.

Figures 6a and 6b display the zonal-vertical distributions of EPT anomalies, diabatic heating anomalies and APE generation associated with MJOs for the observation, ERA-20C and NOAA-20CR respectively. Wang et al. (2018) suggested that the distributions of these fields are the critical metrics for evaluating the fidelity of the simulated MJOs. In observation, the MJO EPT anomaly exhibits a rearward-tilted structure in the lower troposphere (below the 400hPa level) (Figure 6a), reflecting the gradual deepening of the moisture layer toward the west on the east side of the convective center. Such feature is well-reflected in ERA-20C while is not so clear in NOAA-20CR. It is salient that, both ERA-20C and NOAA-20CR exhibit a weaker EPT maximum overlaying the convective center, which exists at 500hPa in observation (but lower in NOAA-20CR). The MJO-related diabatic heating anomaly is calculated using the temperature budget equation. In observation, the diabatic heating anomaly has a lead in the lower troposphere, suggesting the existence of shallow and congestus clouds; meanwhile, it tends to expand westward from the deep convective center in the upper-tropospheric heating between 500 and 300hPa, suggesting that stratiform clouds may follow deep convection (Wang et al., 2018). This tilting structure can be captured by both the two 20C-reanalysis to some extent. But the leading structure on the east side are not so clear in the 20C-reanalysis and the heating is stronger in NOAA-20CR (according with the stronger simulated vertical motion) (Fig. 6b contour). The MJO APE generation, which is the energy source for

MJO development and is determined by the phase overlapping (covariance) of the diabatic heating and temperature anomalies, maximizes near the major convective center (about 450hPa-300hPa) and extends eastward in low level on the east side due to the positive temperature anomalies and heating occurring to the east (figure not shown). To quantify the performance of the 20C-reanalysis in reproducing the vertical distributions of these physical quantities, we calculated the spatial pattern correlations of each field based on the zonal-vertical cross sections, and the results are listed in Table 3.

In order to obtain an overall evaluation of the performance of the 20C-reanalysis in reproducing the MJO vertical structure, we define a skill score considering both the magnitude and spatial pattern in reference to the previous skill score formula (Taylor, 2001; Hirota et al., 2011; Chen et al., 2013). It is given by the following formula:

$$\text{skill score} = \frac{(1+R)^2}{(SDR+\frac{1}{SDR})^2} \quad (1)$$

where R denotes spatial pattern correlation coefficient between the observation and 20C-reanalysis and SDR is the ratio of spatial standard deviation of the 20C-reanalysis against that of the observation. Figure 7 presents the skill scores obtained from each field, including water vapor, vertical velocity, divergence, EPT, diabatic heating and APE generation. The total skill score (TSS), which is estimated by the average of the score for each variable, is 0.91 for ERA-20C and is 0.76 for NOAA-20CR. It indicates that ERA-20C performs better than NOAA-20CR in reproducing the zonal-vertical structure of extremely strong MJO events.

## 3.2 Characteristic in ordinary MJO events

Apart from the extremely strong events, we also examined the performance of the 20C-reanalysis in reproducing the horizontal propagation and vertical structure of the ordinary MJO events. Figure 8 presents scatter plots of PSSs of the MJO convection in extremely strong and ordinary events for ERA-20C and NOAA-20CR respectively. It shows higher skill scores in the extremely strong events (group1) than that in the ordinary events (group2), with the mean (median) of skill scores in the latter about 0.92 (0.93) for ERA-20C and 0.82 (0.83) for NOAA-20CR.

Figure 9 presents the TSSs, which is the average of skill scores calculated from the six assessed fields of MJO vertical structure, for the extremely strong and ordinary MJO events respectively. Interestingly, higher skill scores are found in the ordinary MJO events rather than the extremely strong events in both sets of 20C-reanalysis. It is possibly caused by the relatively poor representation of the MJO intensity, since the most fields we examined show weaker amplitude in the 20C-reanalysis than the observation. This error is magnified in the events with extremely strong convections while it is relatively small in the events with weak MJO convections. Overall, the results of ERA-20C are always better than that of NOAA-20CR.

From the discussion above, it is concluded that regardless of the horizontal propagation or the MJO vertical structure, the capability for MJO representation in ERA-20C is significantly stronger than that in NOAA-20CR. What's more, regardless of the type of event, the skill scores in ERA-20C are higher than that in NOAA-20CR. However, even if the representation ability is relatively weak in NOAA-20CR, the skill can still reach 0.82 (for ordinary events) for MJO propagation and 0.76 (for extremely strong events) for MJO vertical structure. Therefore, in general, the 20C-reanalysis that only assimilate the observed atmospheric surface signals has excellent ability for describing the MJO features and have a good performance in both

phase propagation and vertical structure.

## 4 Summary

Data assimilation is necessary to generate a long-term dataset with high-quality for the MJO study. But whether the atmospheric surface data is enough for assimilation to generate the MJOs close to reality is unknown. To address this question, this study analyzed two reanalysis data (i.e., NOAA-20CR and ERA-20C) which were produced by only assimilating daily observed surface pressure and surface marine winds and compared the results with the observation represented by ERA-Interim reanalysis, which were recognized as the best data to study the MJO. The details of phase propagations and vertical structures associated with the MJOs during the period of 1979 to 2010 were examined carefully.

Eight extremely strong MJO events were screened out after the RMM analysis for the observation. By comparing the horizontal propagation characteristics in these events in 20C-reanalysis with the observational results, it is found that the MJO propagation processes in both two 20C-reanalysis sets are described well. The PSS in ERA-20C is as high as 0.96, while it is 0.87 in NOAA-20CR, which indicates that the ability of ERA-20C is superior to NOAA-20CR in terms of horizontal propagation. The correlation in propagation phase and RMSE in convective intensity also proved that the 20C-reanalysis is good for the representation of the MJO phase and intensity.

By examining the vertical structures of moisture anomalies, zonal wind anomalies, vertical velocity anomalies, divergence anomalies, EPT anomalies, diabatic heating anomalies and APE generation associated with the MJOs, we found that the 20C-reanalysis datasets that only assimilate the observational atmospheric

surface signals (i.e., only the surface pressure for the two 20C-reanalysis and additional surface marine winds for ERA-20C) could depict the MJO vertical structure reasonably. For each field, the correlation of vertical distribution between the 20C-reanalysis and the observation is significantly positive, with a TSS of 0.91 in ERA-20C and 0.76 in NOAA-20CR. However, the representations in the convection intensity and the phase-leading structure in low level are not so good.

We also analyzed the MJO features in ordinary events. After comparing the skill scores of the two groups of events, it can be illustrated that the 20C-reanalysis is excellent for describing all types of the MJO processes. In terms of horizontal propagation, the performance in strong events is superior to that in ordinary events in 20C-reanalysis. In terms of vertical structure, the representation of ordinary events is better, possibly due to the overall weaker representation in the MJO intensity that results in relative greater error existing in the strong events.

Both sets of 20C data can characterize the MJO very well, although ERA-20C is superior to NOAA-20CR, regardless of horizontal propagation or vertical structure, the extremely strong events or ordinary events. This may be due to the fact that the additional surface marine wind field is also assimilated in ERA-20C while only observed surface pressure is in NOAA-20CR. Beyond that, ERA-20C and the verified data (i.e., ERA-Interim) were produced by the same organization using similar models may also be the possible reason.

The results from this study indicate that the observed surface signals are good in data assimilation to generate high-quality data for the MJO study. It serves as a reference for the future study. In future assimilation experiment related to the MJO the assimilation of atmospheric low-level signals may be given priority. Meanwhile, it is admitted that adding the upper-air information such as the one from radiosondes into

the assimilation system together with surface observations could further improve the quality of the data and reduce the variability biases (Hersbach et al. 2017).

One may note that a previous study (i.e., Gao et al. (2016)) has indicated that NOAA-20CR is not good at reproducing the MJO phase evolution, which seems contrary to our results. However, it should be mentioned that their evaluation was only based on a very limited period (i.e., winter of 2011, when an international field experiment took place over the central equatorial Indian Ocean) wth only two MJO events, while this study employed a much longer period from 1979 to 2010 with more than one hundred MJO events. Furthermore, we also examined the long-term variation (1900-2010) of MJO-related fields based on the two 20C-reanalysis. They both show an obvious increase trend of precipitation intensity with global warming while a much weaker trend of wind intensity (figures not shown). This result is consistent with the recently published paper by Maloney (2019), who analyzed the MJO change under global warming using CMIP5 experiments of RCP4.5.

MJO prediction based on dynamic model is a difficult task. This study suggests that atmospheric component of the dynamic model can be accurately initialized through assimilating observed surface signals. Comparing with assimilations of diverse observations with vertical structures, this approach may have following two advantages. Firstly, it can keep the preferred vertical dynamic/thermodynamic structures of the model and thus reduce initial shocks in forecasts. Secondly, it can greatly reduce the difficulty of the development of initialization system, as well as its computation cost.

**Acknowledgments.** This work was supported by National key R&D Program of China 2018YFC1505801 and 2015CB453200, NSFC grants 41705059 and 41875069, NOAA NA18OAR4310298, NSF AGS-1643297, and the Startup Foundation for Introducing Talent of NUIST. This is SOEST contribution number 12345, IPRC contribution number 1234, and ESMC contribution number 123.

## Reference

Chen L, Li T, Behera SK, Doi T. 2016a. Distinctive precursory air–sea signals between regular and super El Niños. *Advances in Atmospheric Sciences* 33: 996–1004. https://doi.org/10.1007/s00376-016-5250-8.

Chen L, Yu Y, Sun DZ. 2013. Cloud and Water Vapor Feedbacks to the El Niño Warming: Are They Still Biased in CMIP5 Models? *Journal of Climate* 26:4947-4961. https://doi.org/10.1175/jcli-d-12-00575.1

Chen L, Yu Y, Zheng W. 2016b. Improved ENSO simulation from climate system model FGOALS-g1.0 to FGOALS-g2. *Climate Dynamics* 47: 2617–2634. https://doi.org/10.1007/s00382-016-2988-8.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The twentieth century reanalysis project. *Quarterly Journal of Royal Meteorological Society* 137: 1–28. https://doi.org/10.1002/qj.776.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137: 553–597.

https://doi.org/10.1002/qj.828.

Duchon CE. 1979. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology* 18: 1016-1022. https://doi.org/10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2

Fu X, Lee JY, Hsu PC, Taniguchi H, Wang B, Wang W, Weaver S. 2013. Multi-model MJO forecasting during DYNAMO/CINDY period. *Climate Dynamics* 41: 1067–1081. https://doi.org/10.1007/s00382-013-1859-9.

Fu X, Wang W, Lee JY, Wang B, Kikuchi K, Xu J, Li J, Weaver S. 2015. Distinctive Roles of Air–Sea Coupling on Different MJO Events: A New Perspective Revealed from the DYNAMO/CINDY Field Campaign. *Monthly Weather Review* 143: 794–812. https://doi.org/10.1175/MWR-D-14-00221.1.

Gao Y, Hsu PC, Hsu HH. 2016. Assessments of surface latent heat flux associated with the Madden–Julian Oscillation in reanalyses. *Climate Dynamics* 47: 1755–1774. https://doi.org/10.1007/s00382-015-2931-4.

Harada Y, Kamahori H, Kobayashi C, Endo H, Kobayashi S, Ota Y, Onoda H, Onogi K, Miyaoka K, Takahashi K. 2016. The JRA-55 Reanalysis: Representation of Atmospheric Circulation and Climate Variability. *Journal of the Meteorological Society of Japan. Ser. II* 94: 269–302. https://doi.org/10.2151/jmsj.2016-015.

Hendon HH, Salby ML. 1994. The life cycle of the Madden-Julian oscillation. *Journal of the Atmospheric Sciences* 51: 2225–2237. https://doi.org/10.1175/1520-0469(1994)051<2225:TLCOTM>2.0.CO;2

Hersbach H, Brönnimann S, Haimberger L, Mayer M, Villiger L, Comeaux J, Simmons A, Dee D, Jourdain S, Peubey C, Poli P, Rayner N, Sterin AM, Stickler A, Valente MA, Worley SJ. 2017. The potential value of early (1939–1967)

upper-air data in atmospheric climate reanalysis. Quarterly Journal of the Royal Meteorological Society 143: 1197–1210. https://doi.org/10.1002/qj.3040

Hirota N, Takayabu YN, Watanabe M, Kimoto M. 2011. Precipitation Reproducibility over Tropical Oceans and Its Relationship to the Double ITCZ Problem in CMIP3 and MIROC5 Climate Models. *Journal of Climate* 24: 4859–4873. https://doi.org/10.1175/2011JCLI4156.1.

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

Hsu PC, Yang Y. 2016. Contribution of atmospheric internal processes to the interannual variability of the South Asian summer monsoon: INTERANNUAL VARIABILITY OF SOUTH ASIAN SUMMER MONSOON. *International Journal of Climatology* 36: 2917–2930. https://doi.org/10.1002/joc.4528.

Hung MP, Lin JL, Wang W, Kim D, Shinoda T, Weaver SJ. 2013. MJO and Convectively Coupled Equatorial Waves Simulated by CMIP5 Climate Models. *Journal of Climate* 26: 6185–6214. https://doi.org/10.1175/JCLI-D-12-00541.1.

Huntley HS, Hakim GJ. 2010. Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Climate Dynamics* 35: 995–1009. https://doi.org/10.1007/s00382-009-0714-5.

Jiang X, Waliser DE, Olson WS, Tao WK, L'Ecuyer TS, Li KF, Yung YL, Shige S, Lang S, Takayabu YN. 2011. Vertical Diabatic Heating Structure of the MJO: Intercomparison between Recent Reanalyses and TRMM Estimates. *Monthly Weather Review* 139: 3208–3223. https://doi.org/10.1175/2011MWR3636.1.

Jiang X, Waliser DE, Xavier PK, Petch J, Klingaman NP, Woolnough SJ, Guan B,

Bellon G, Crueger T, DeMott C, Hannay C, Lin H, Hu W, Kim D, Lappen C-L, Lu M-M, Ma H-Y, Miyakawa T, Ridout JA, Schubert SD, Scinocca J, Seo K-H, Shindo E, Song X, Stan C, Tseng WL, Wang W, Wu T, Wu X, Wyser K, Zhang GJ, Zhu H. 2015. Vertical structure and physical processes of the Madden-Julian oscillation: Exploring key model physics in climate simulations: KEY PHYSICS IN MODELING THE MJO. *Journal of Geophysical Research: Atmospheres* 120: 4718–4748. https://doi.org/10.1002/2014JD022375.

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* 77: 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu M, Ebisuzaki W, Yang SK, Hnilo JJ, Fiorino M, Potter GL. 2002. NCEP-DOE AMIP-II REANALYSIS (R-2). *Bulletin of the American Meteorological Society* 14: 1631-1644. https://doi.org/10.1175/BAMS-83-11-1631

Kessler WS, Kleeman R. 2000. Rectification of the Madden–Julian Oscillation into the ENSO Cycle. *Journal of Climate* 13: 3560–3575. https://doi.org/10.1175/1520-0442(2000)013<3560:ROTMJO>2.0.CO;2.

Kim D, Lee MI, Kim D, Schubert SD, Waliser DE, Tian B. 2014. Representation of tropical subseasonal variability of precipitation in global reanalyses. *Climate Dynamics* 43: 517–534. https://doi.org/10.1007/s00382-013-1890-x.

Kim D, Sperber K, Stern W, Waliser D, Kang IS, Maloney E, Wang W, Weickmann K,

Benedict J, Khairoutdinov M, Lee M-I, Neale R, Suarez M, Thayer-Calder K, Zhang G. 2009. Application of MJO Simulation Diagnostics to Climate Models. *Journal of Climate* 22: 6413–6436. https://doi.org/10.1175/2009JCLI3063.1.

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K. 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II* 93: 5–48. https://doi.org/10.2151/jmsj.2015-001.

Li T. 2014. Recent Advance in Understanding the Dynamics of the Madden-Julian Oscillation. *Journal of Meteorological Research* 28, 1-33. https://doi.org/10.1007/s13351-014-3087-6

Li T, Hsu PC (2017) Fundamentals of Tropical Climate Dynamics. Springer. ISBN 978-3-319-59595-3. https://doi.org/10.1007/978-3-319-59597-9

Li T, Wang B. 1994. A thermodynamic equilibrium climate model for monthly mean surface winds and precipitation over the tropical Pacific. *Journal of the Atmospheric Sciences* 51: 1372-1385. https://doi.org/10.1175/1520-0469(1994)051<1372:ATECMF>2.0.CO;2

Li T, Wang B. 2005. A review on the western North Pacific monsoon: synoptic-to-interannual variabilities. *Terrestrial, Atmospheric and Oceanic Sciences* 16, 285-314. https://doi.org/10.3319/tao.2005.16.2.285(a

Li T, Wang L, Peng M, Wang B, Zhang C, Lau W, Kuo HC. 2018. A Paper on the Tropical Intraseasonal Oscillation Published in 1963 in a Chinese Journal. *Bulletin of the American Meteorological Society* 99: 1765-1779. https://doi.org/10.1175/BAMS-D-17-0216.1

Li T, Zhou C. 2009. Planetary scale selection of the Madden-Julian Oscillation.

*Journal of the Atmospheric Sciences* 66: 2429-2443. https://doi.org/10.1175/2009JAS2968.1

Liebmann B, Hendon HH, Glick JD. 1994. The relationship between tropical cyclones of the western Pacific and Indian Oceans and the Madden-Julian oscillation. *Journal of the Meteorological Society of Japan* 72: 401-412. https://doi.org/10.2151/jmsj1965.72.3_401

Liebmann B, Smith CA. 1996. Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset. *Bulletin of the American Meteorological Society* 77: 1275-1277.

Lin JL, Kiladis GN, Mapes BE, Weickmann KM, Sperber KR, Lin W, Wheeler MC, Schubert SD, Del Genio A, Donner LJ, Emori S, Gueremy JF, Hourdin F, Rasch PJ, Roeckner E, Scinocca JF. 2006. Tropical Intraseasonal Variability in 14 IPCC AR4 Climate Models. Part I: Convective Signals. *Journal of Climate* 19: 2665–2690. https://doi.org/10.1175/JCLI3735.1.

Liu F, Wang B. 2012. A frictional skeleton model for the Madden-Julian oscillation. *Journal of the Atmospheric Sciences* 69: 2749–2758. https://doi.org/10.1175/JAS-D-12-020.1

Liu F, Li T, Wang H, Deng L, Zhang Y. 2016. Modulation of boreal summer intraseasonal oscillation over the western North Pacific by the ENSO. *Journal of Climate* 29: 7189-7201. https://doi.org/10.1175/JCLI-D-15-0831.1.

Liu F, Wang B. 2017. Effects of moisture feedback in the frictional coupled Kelvin-Rossby wave model and implication to the Madden-Julian oscillation. *Climate Dynamics* 48:513-522. https://doi.org/10.1007/s00382-016-3090-y

Madden RA, Julian PR.1972. Description of global-scale circulation cells in the

tropics with a 40-50 day period. *Journal of the Atmospheric Sciences* 29: 1109-1123. https://doi.org/10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2

Majda AJ, Biello JA. 2004. A multiscale model for tropical intraseasonal oscillations. *Proceedings of the National Academy of Sciences* 101: 4736–4741. https://doi.org/10.1073/pnas.0401034101.

Maloney ED, Hartmann DL. 1998. Frictional Moisture Convergence in a Composite Life Cycle of the Madden–Julian Oscillation. *Journal of Climate* 11: 2387–2403. https://doi.org/10.1175/1520-0442(1998)011<2387:FMCIAC>2.0.CO;2.

Mapes B, Tulich S, Lin J, Zuidema P. 2006. The mesoscale convection life cycle: Building block or prototype for large-scale tropical waves? *Dynamics of Atmospheres and Oceans* 42: 3–29. https://doi.org/10.1016/j.dynatmoce.2006.03.003.

Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N, Taira R. 2007. The JRA-25 Reanalysis. *Journal of the Meteorological Society of Japan. Ser. II* 85: 369–432. https://doi.org/10.2151/jmsj.85.369.

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DGH, Peubey C, Thépaut JN, Trémolet Y, Hólm EV, Bonavita M, Isaksen L, Fisher M. 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate* 29: 4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1.

Slingo JM, Sperber KR, Boyle JS, Gregory D, Madden R, Matthews A, Palmer TN. 1996. Intraseasonal oscillations in 15 atmospheric general circulation models:

results from an AMIP diagnostic subproject. *Climate Dynamics* 12: 325-357. https://doi.org/10.1007/BF00231106

Straub KH. 2013. MJO Initiation in the Real-Time Multivariate MJO Index. *Journal of Climate* 26: 1130–1151. https://doi.org/10.1175/JCLI-D-12-00074.1.

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* 106: 7183–7192. https://doi.org/10.1029/2000JD900719.

Tian B, Waliser DE, Fetzer EJ, Yung YL. 2010. Vertical Moist Thermodynamic Structure of the Madden–Julian Oscillation in Atmospheric Infrared Sounder Retrievals: An Update and a Comparison to ECMWF Interim Re-Analysis. *Monthly Weather Review* 138: 4576–4582. https://doi.org/10.1175/2010MWR3486.1.

Ulate Medrano AM. 2015. Limitations in the MJO Initiation Over the Indian Ocean in Limited Area Models. *Ph.D thesis*. University of Miami. https://scholarlyrepository.miami.edu/oa_dissertations/1368

Uppala SM, KÅllberg PW, Simmons AJ, Andrae U, Bechtold VDC, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Berg LVD, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, Mcnally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society* 131: 2961–3012. https://doi.org/10.1256/qj.04.176.

Wang B, Li T. 1994. Convective interaction with boundary-layer dynamics in the development of a tropical intraseasonal system. *Journal of the Atmospheric Sciences* 51: 1386-1400. https://doi.org/10.1175/1520-0469(1994)051<1386:CIWBLD>2.0.CO;2

Wang B, Lee SS. 2017. MJO propagation shaped by zonal asymmetric structures: Results from 24 GCM simulations. Journal of Climate 30: 7933-7952. https://doi.org/10.1175/JCLI-D-16-0873.1.

Wang B, Lee SS, Waliser DE, Zhang C, Sobel A, Maloney E, Li T, Jiang X, Ha KJ. 2018. Dynamics-oriented diagnostics for the Madden-Julian Oscillation. *Journal of Climate* 31: 3117-3135. https://doi.org/10.1175/JCLI-D-17-0332.1.

Wang L, Li T, Nasuno T. 2018. Impact of Rossby and Kelvin Wave Components on MJO Eastward Propagation. *Journal of Climate* 31: 6913-6931, https://doi.org/10.1175/JCLI-D-17-0749.1

Wang L, Li T, Maloney E, Wang B. 2017. Fundamental Causes of Propagating and Non-propagating MJOs in MJOTF/GASS models. *Journal of Climate* 30: 3743-3769. https://doi.org/10.1175/JCLI-D-16-0765.1.

Wang L, Li T, Zhou T, Rong X. 2013. Origin of the Intraseasonal Variability over the North Pacific in Boreal Summer. *Journal of Climate* 26: 1211-1229. https://doi.org/10.1175/JCLI-D-11-00704.1

Weaver SJ, Wang W, Chen M, Kumar A. 2011. Representation of MJO Variability in the NCEP Climate Forecast System. *Journal of Climate* 24: 4676–4694. https://doi.org/10.1175/2011JCLI4188.1.

Wheeler MC, Hendon HH. 2004. An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather*

*Review* 132: 1917–1932. https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Yang YM, Wang B. 2018. Improving MJO simulation by enhancing the interaction between boundary layer convergence and lower tropospheric heating. *Climate Dynamics*. https://doi.org/10.1007/s00382-018-4407-9.

Yasunari T. 1979. Cloudiness fluctuations associated with the Northern Hemisphere summer monsoon. *Journal of the Meteorological Society of Japan* 57: 227-242. https://doi.org/10.2151/jmsj1965.57.3_227

Zhao C, Li T, Zhou T. 2013. Precursor signals and processes associated with MJO initiation over the tropical Indian Ocean. *Journal of Climate* 26: 291-307. https://doi.org/10.1175/JCLI-D-12-00113.1

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

## Figure Captions

Fig. 1 Zonal propagation of the MJO convection measured by 850hPa zonal wind (upper) and 200hPa zonal wind (lower) field averaged over 20°S-20°N in 8 extremely strong events during the period of 1979-2010 in ERA-20C (shading) and observation (contour). Unit: m/s.

Fig. 2 Zonal propagation of the MJO convection measured by 850hPa zonal wind (upper) and 200hPa zonal wind (lower) field averaged over 20°S-20°N in 8 extremely strong events during the period of 1979-2010 in NOAA-20CR (shading) and observation (contour). Unit: m/s.

Fig. 3 Propagation phase lines in two-dimensional phase space defined by RMM1 and RMM2 in 8 extremely strong events for observation (green) and for ERA-20C (blue). Results for NOAA-20CR not shown.

Fig. 4 Correlations (green circles, with values in the left y axis) of propagation phase lines and root mean square errors (purple triangles, with values in the right y axis) of convective intensity in extremely strong events between observation-derived and ERA-20C-derived (upper) and NOAA-20CR-derived (lower), with dash-line representing 5% significant level.

Fig. 5 Composite for zonal-vertical cross section of (a) specific humidity (shading, unit: g/kg) and zonal wind field (contour, unit: m/s) and (b) divergence (shading, unit: $10^{-6}s^{-1}$) and vertical velocity (contour, unit: $10^{-2}$Pa/s) on the day with strongest convection (day0) of 8 MJO events in observation and 20C-reanalysis datasets. The solid (dash) line indicating positive (negative) value.

Fig. 6 Composite for zonal-vertical cross section of (a) equivalent potential temperature (unit: K) and (b) diabatic heating (shading, unit: K/day) and vertical velocity (contour, unit: $10^{-2}$Pa/s) on the day with strongest convection (day0) of 8 MJO events in observation and 20C-reanalysis datasets. The solid (dash) line

indicating positive (negative) value.

Fig. 7 Skill scores, calculated according to Eq. (1), for vertical distribution of specific humidity ("hus"), vertical velocity ("wap"), divergence ("div"), equivalent potential temperature ("EPT"), diabatic heating ("Q") and available potential energy generation ("APE") on the day with strongest convection (day0) of 8 MJO events for ERA-20C (solid-fill) and NOAA-20CR (grid-fill).

Fig. 8 Propagation skill scores (PSSs) of the MJO convection in extremely strong (Group1) and ordinary (Group2) events for ERA-20C (upper) and NOAA-20CR (lower), with yellow line representing the mean and green line representing the median.

Fig. 9 Skill scores for vertical structure (i.e., TSS) of the composite MJO convection in extremely strong (red) and ordinary (blue) events for ERA-20C (left) and NOAA-20CR (right).

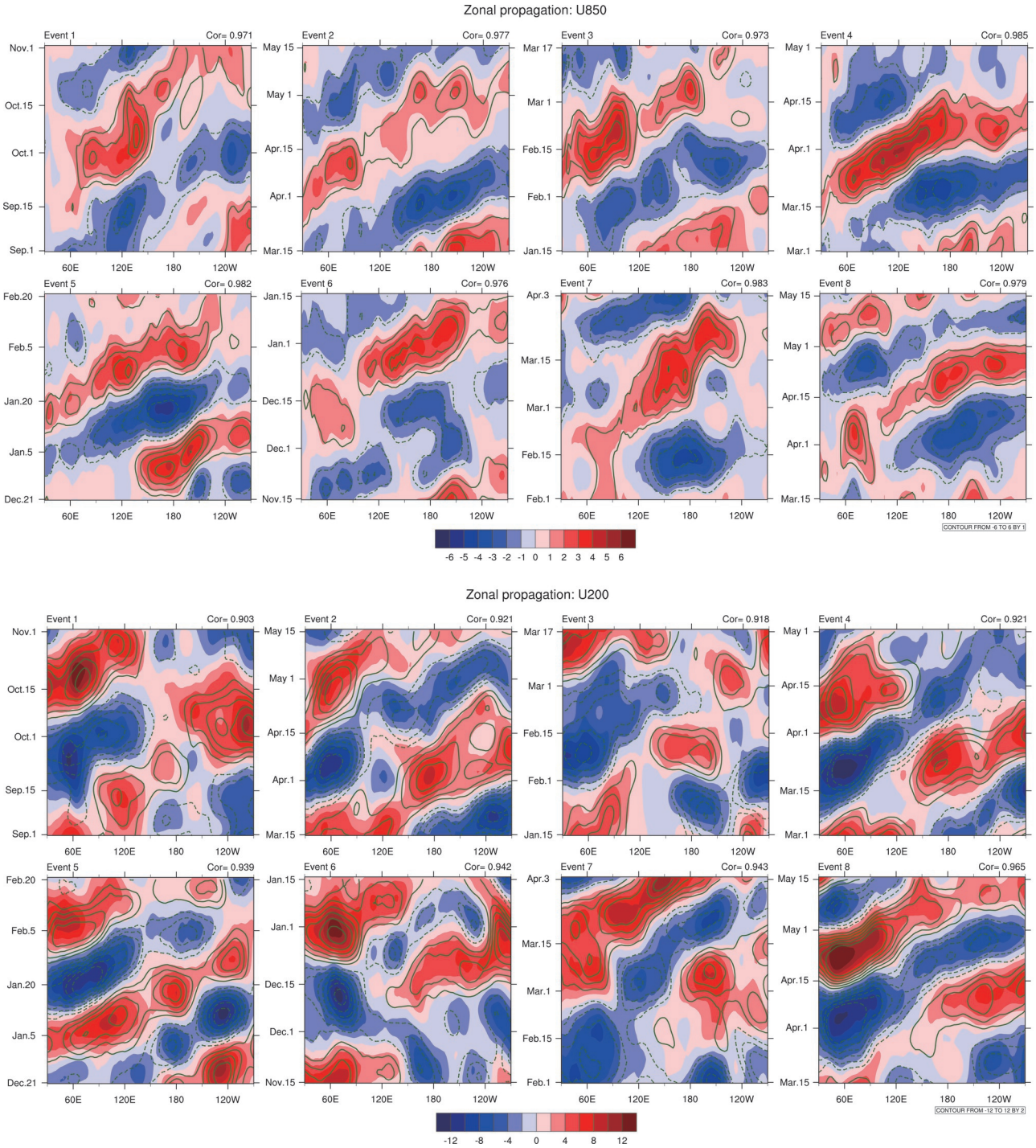
## Table Captions

Table 1 The periods for 8 extremely strong MJO events during 1979-2010.

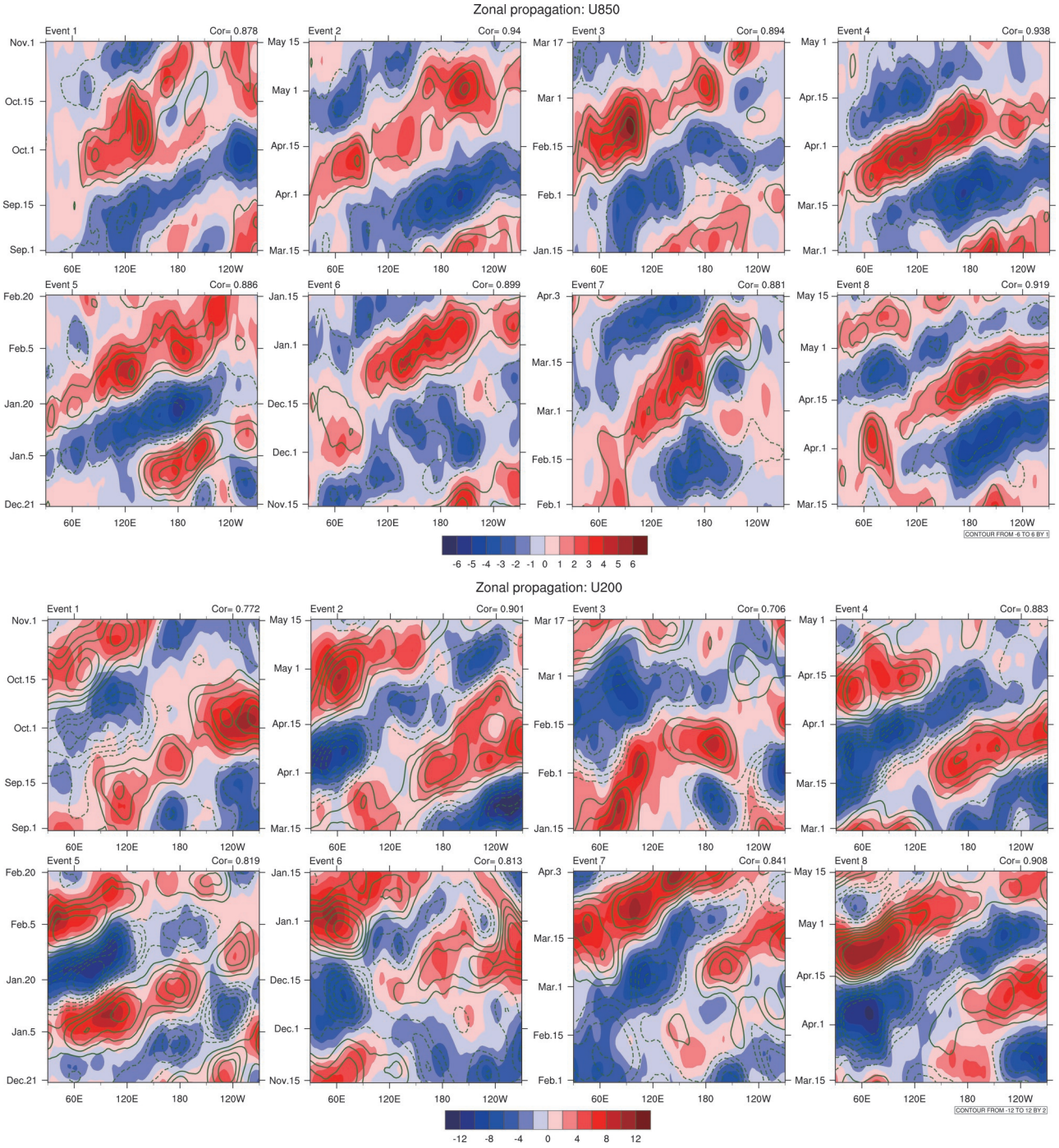
Table 2 Propagation skill scores (PSSs) of the MJO convection in extremely strong events.

Table 3 Spatial pattern correlation coefficients for vertical distribution of specific humidity ("hus"), vertical velocity ("wap"), divergence ("div"), equivalent potential temperature ("EPT"), diabatic heating ("Q") and available potential energy generation ("APE") on the day with strongest convection (day0) of 8 MJO events between observation and 20C-reanalysis datasets.

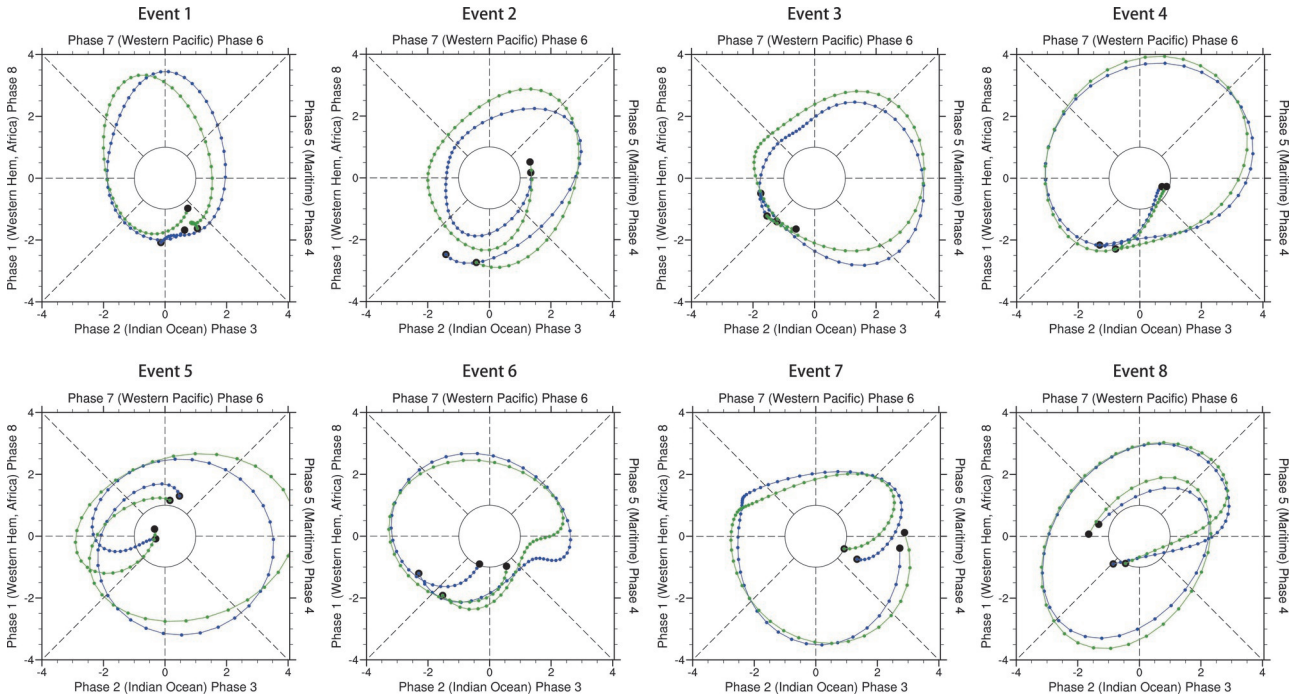
Table 4 Intensity scores of composite vertical structure and the leading structure ahead of convection in observation and 20C-reanalysis datasets (Q for specific humidity with unit of g/kg, W and DIV for vertical velocity and divergence with units of $10^{-2}$ Pa/s and $10^{-6}s^{-1}$, respectively).



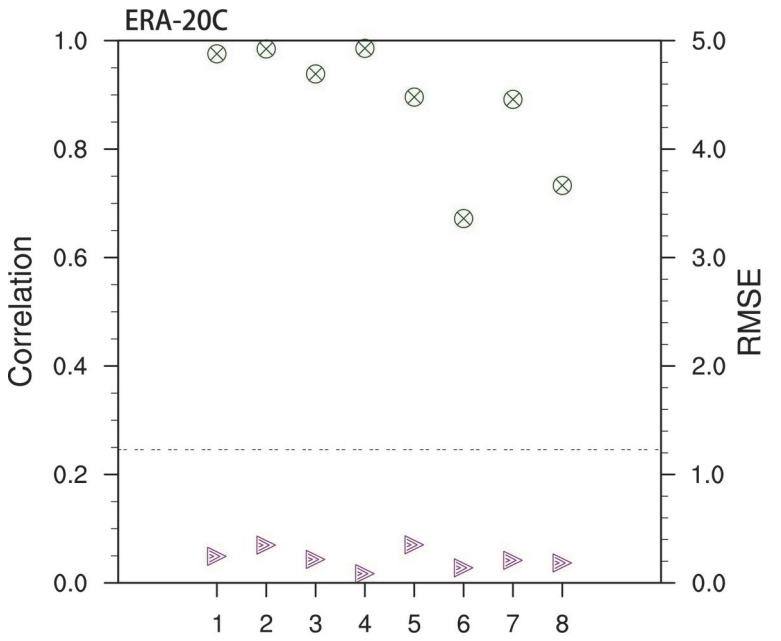

joc_6270_figure1.eps



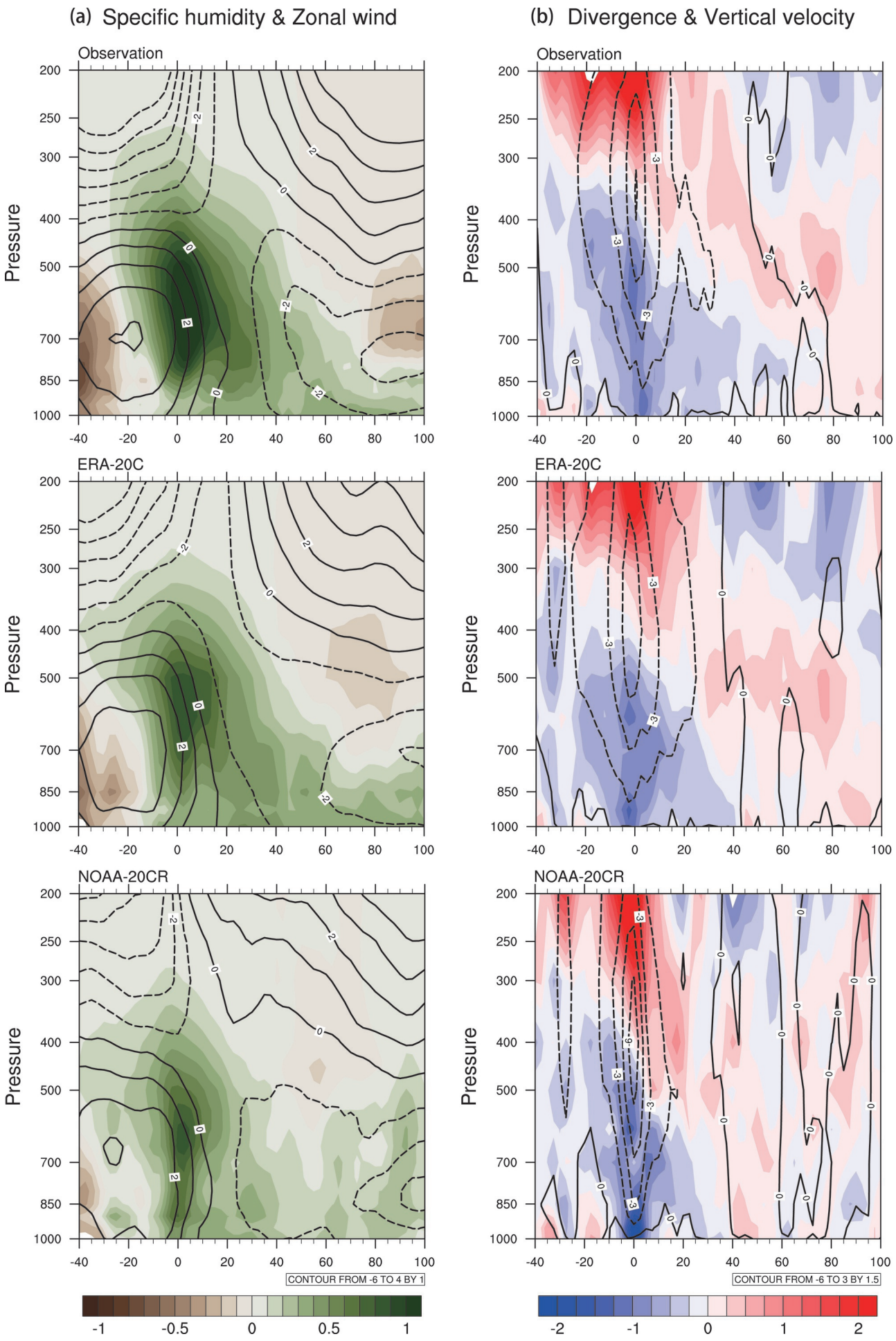

joc_6270_figure2.eps



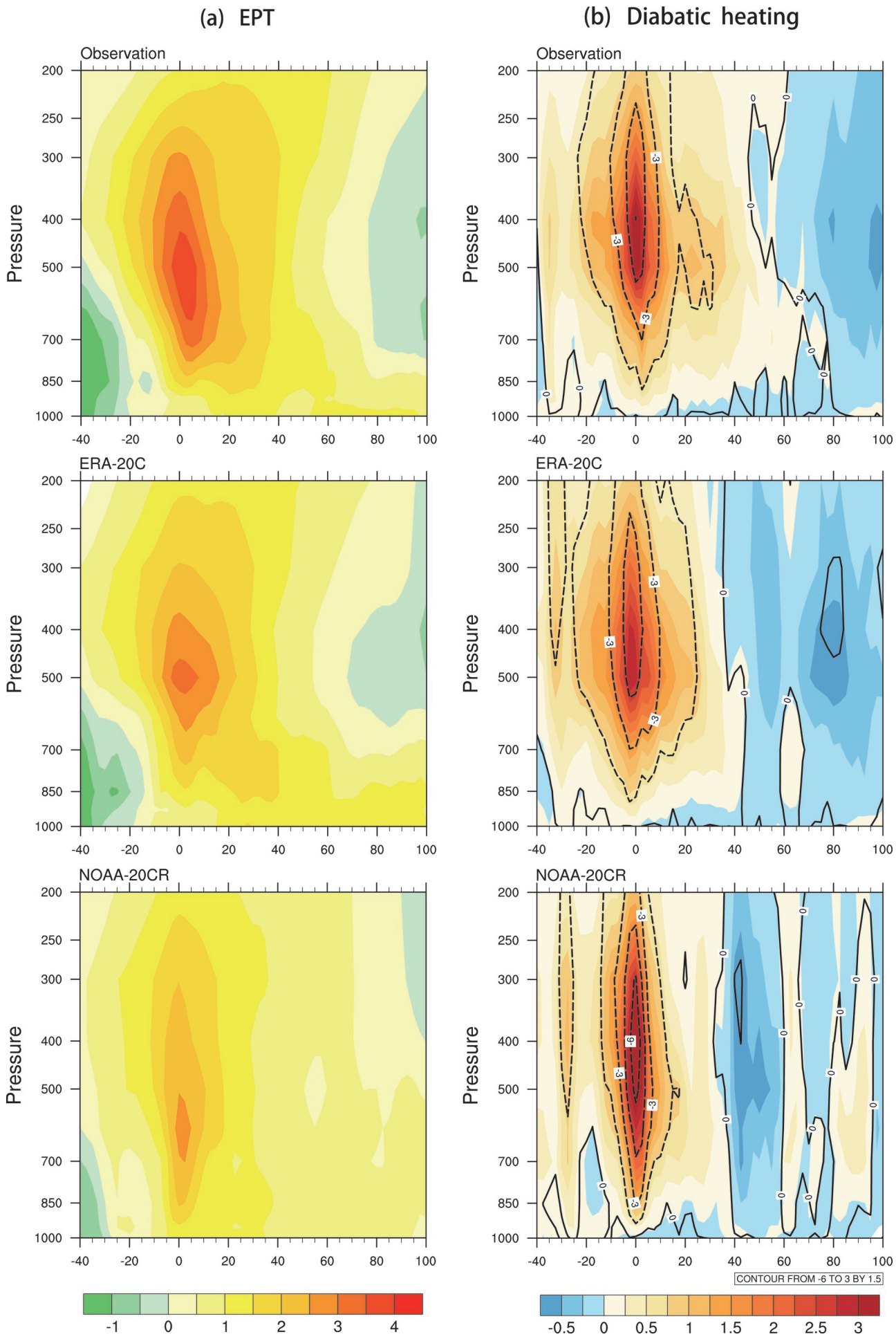

joc_6270_figure 3.eps



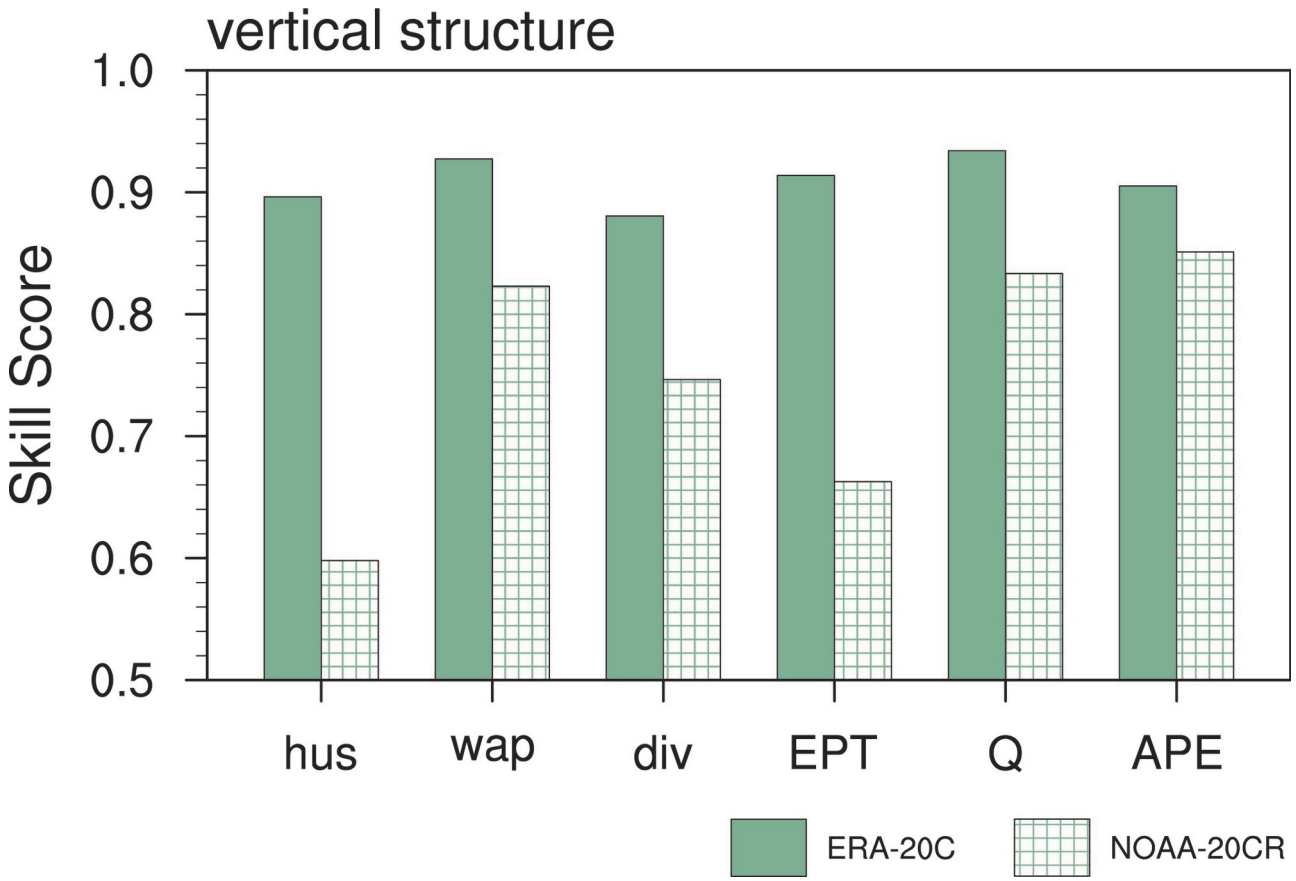

joc_6270_figure 4.eps



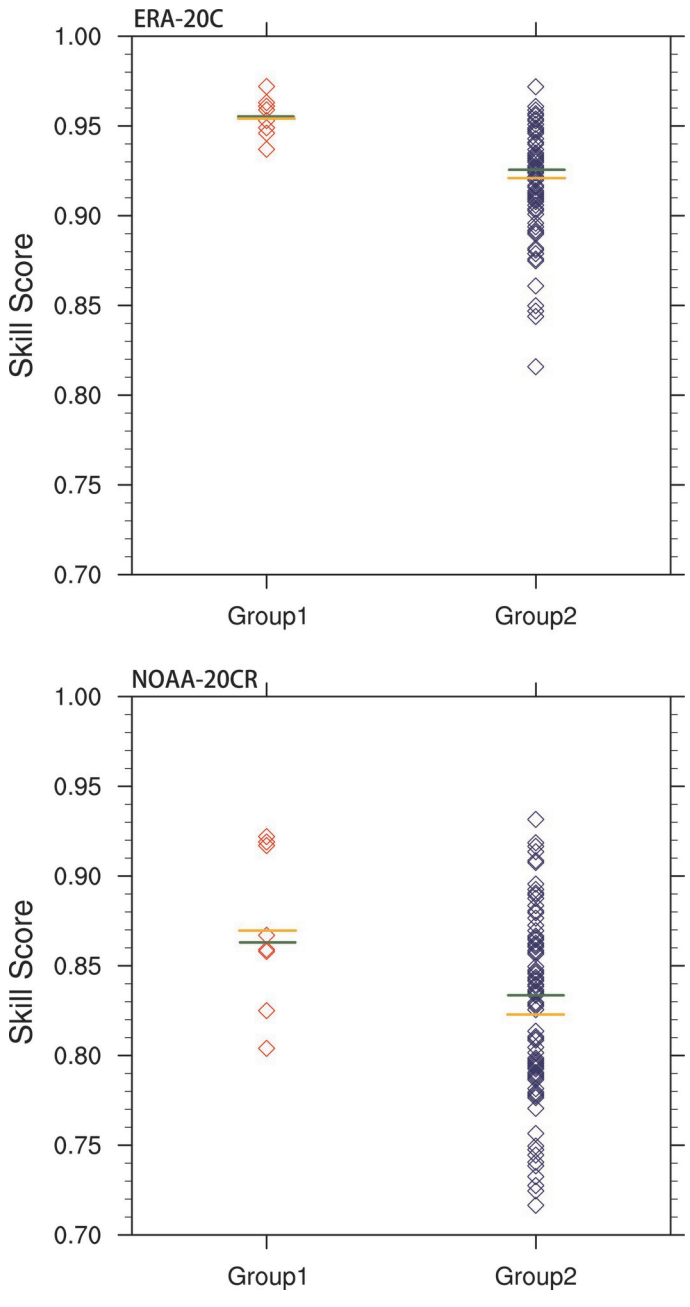

joc_6270_figure 5.eps



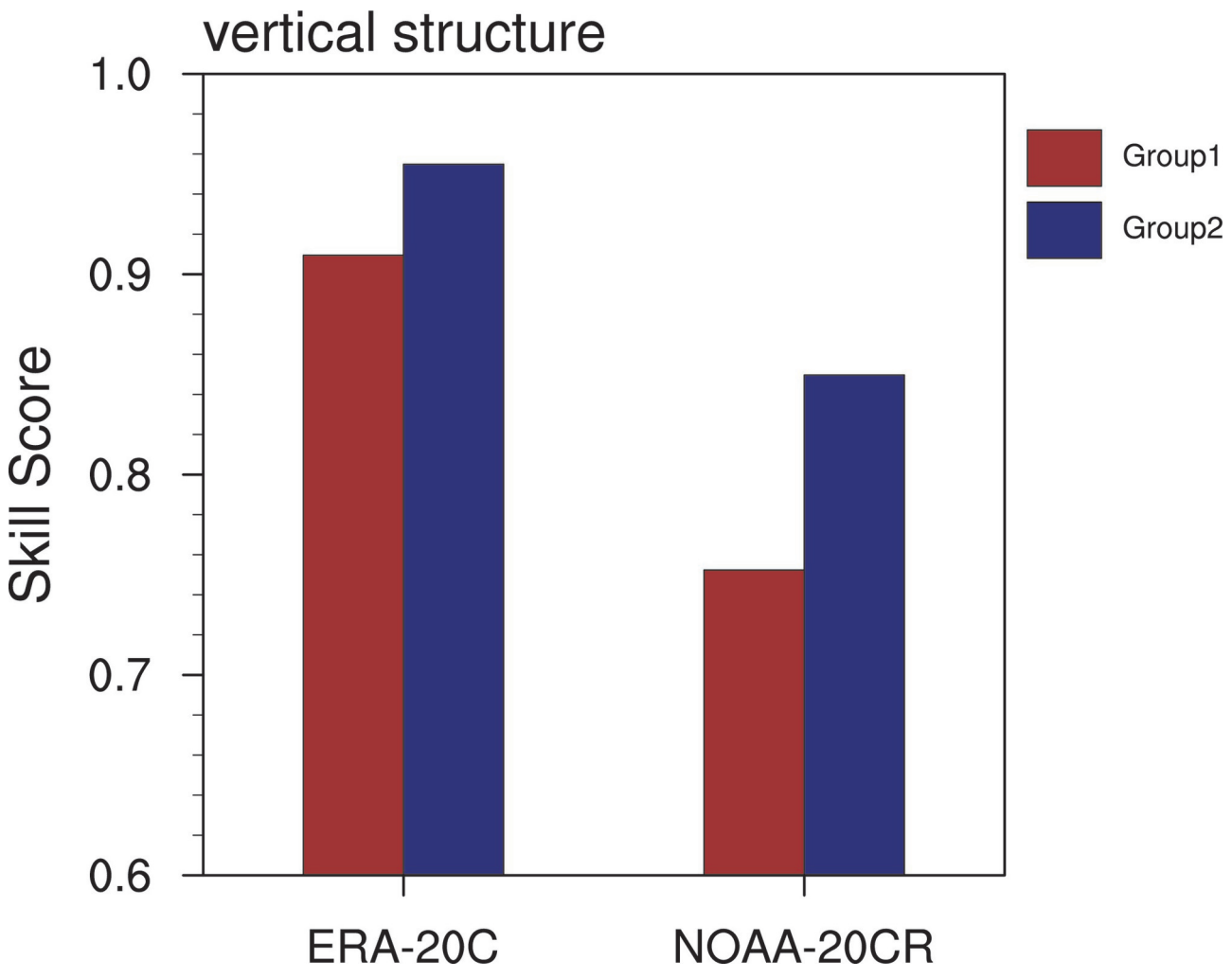

joc_6270_figure 6.eps




joc_6270_figure 7.eps




joc_6270_figure 8.eps

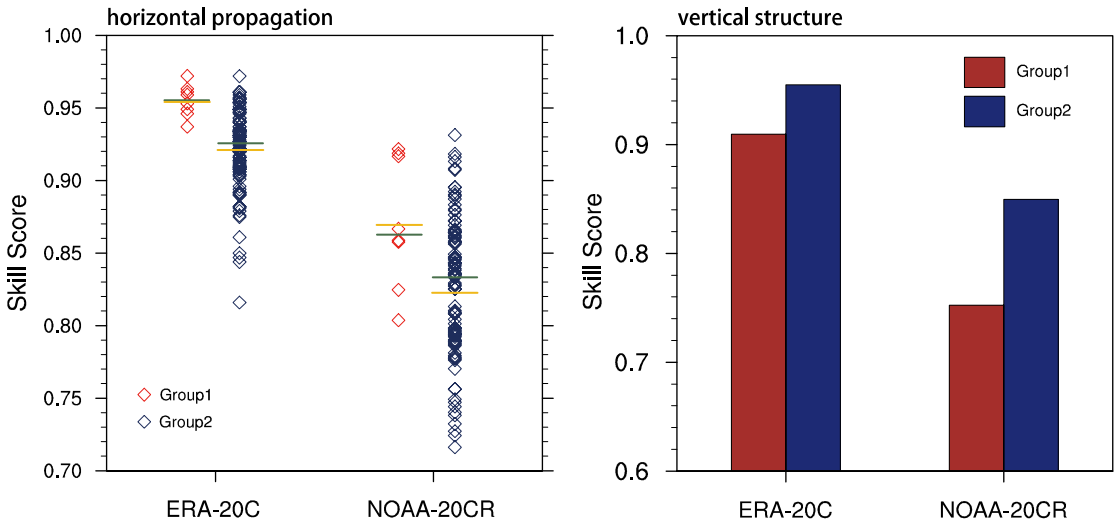



joc_6270_figure 9.eps

# Graphic abstract

Can reanalysis products with only surface variables assimilated capture MJO characteristics?

Jingxuan Cui, Lu Wang, Tim Li and Bo Wu




Both ERA-20C and NOAA-20CR with only the observed surface signal assimilated, could reproduce the observed MJO characteristics very well, with the former superior to the latter, regardless of MJO intensity.

| Event | Period |
|---|---|
| **1** | 19790901-19791101 |
| **2** | 19810315-19810515 |
| **3** | 19850115-19850317 |
| **4** | 19880301-19880501 |
| **5** | 19921221-19930220 |
| **6** | 19961115-19970115 |
| **7** | 19970201-19970403 |
| **8** | 20050315-20050515 |

**Table 1** The periods for 8 extremely strong MJO events during 1979-2010.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Avg |
|---|---|---|---|---|---|---|---|---|---|
| **ERA-20C** | 0.937 | 0.949 | 0.9455 | 0.953 | 0.9605 | 0.959 | 0.963 | 0.972 | **0.95** |
| **20CR** | 0.825 | 0.921 | 0.8 | 0.9105 | 0.8525 | 0.856 | 0.861 | 0.9135 | **0.87** |

**Table 2** Skill scores for horizontal propagation of the MJO convection in extremely strong events.

**Table 3** Spatial pattern correlation coefficients for vertical distribution of specific humidity ("hus"), vertical velocity ("wap"), divergence ("div"), equivalent potential temperature ("EPT"), diabatic heating ("Q") and available potential energy generation ("APE") on the day with strongest convection (day0) of 8 MJO events between observation and 20C-reanalysis datasets.

| | **ERA-20C** | **NOAA-20CR** |
|---|---|---|
| **hus** | 0.93 | 0.81 |
| **wap** | 0.94 | 0.81 |
| **Div** | 0.88 | 0.73 |
| **EPT** | 0.95 | 0.88 |
| **Q** | 0.94 | 0.83 |
| **APE** | 0.92 | 0.85 |

**Table 4** Intensity scores of composite vertical structure and the leading structure ahead of convection in observation and 20C-reanalysis datasets (Q for specific humidity with unit of g/kg, W and DIV for vertical velocity and divergence with units of $10^{-2}$ Pa/s and $10^{-6}s^{-1}$, respectively).

| | | **Observation** | **ERA-20C** | **NOAA-20CR** |
|---|---|---|---|---|
| Intensity score | Q | 1.03 | 0.76 | 0.67 |
| | W | 4.60 | 4.63 | 4.96 |
| Leading structure | Q | 0.62 | 0.48 | 0.31 |
| | DIV | 0.36 | 0.24 | 0.21 |