AMERICAN
METEOROLOGICAL
SOCIETY

Monthly Weather Review

The DOI for this manuscript is doi: 10.1175/2010MWR3562.1

Seasonal modulations of the active MJO cycle characterized by nonlinear principal component analysis

Johannes Jenkner * and William W. Hsieh

University of British Columbia, Vancouver, BC, Canada

Alex J. Cannon

Environment Canada, Vancouver, BC, Canada

* *Corresponding author address:* Johannes Jenkner, Department of Earth and Ocean Sciences, University of British Columbia, 2219 Main Mall, Vancouver, BC, V6T 1Z4, Canada

E-mail: johannes.jenkner@alumni.ethz.ch

ABSTRACT

A novel methodology is presented for the identification of the mean cycle of the Madden–Julian Oscillation (MJO) along the equator. The methodology is based on a nonlinear principal component (NLPC) computed with a neural network model. The bandpass–filtered input data encompass 30 years with zonal winds on 850 hPa and 200 hPa plus outgoing longwave radiation (OLR). The NLPC is conditioned on a sufficiently strong MJO activity and is computed both for the pooled dataset and for the dataset stratified into seasons. The NLPC for all data depicts a circular mode formed by the first two linear principal components (LPCs) with marginal contributions by the higher–order LPCs. Hence, the mean MJO cycle throughout the year is effectively captured by the amplitude of the leading two LPCs varying in quadrature. The NLPC for individual seasons shows additional variability which mainly arises from a subordinate oscillation of the second pair of LPCs superimposed on the annual MJO signal. In reference to the all–year solution, the difference in resolved variability approximately accounts for 9% in solstitial seasons and 3% in equinoctial seasons. The phasing of the third LPC is such that convective activity oscillations over the Maritime Continent as well as wind oscillations over the Indian Ocean appear enhanced (suppressed) during boreal winter (summer). Also, convective activity oscillations appear more pronounced at the date line during both winter and summer. The phasing of the fourth LPC is such that upper–level westerlies over the Atlantic region are more persistent during boreal spring than during other seasons.

1. Introduction

a. Background information on the MJO

The Madden–Julian Oscillation (MJO) constitutes the dominant atmospheric oscillation pattern in the tropics and has been subject to many studies about its physical structure, dynamical influence and statistical predictability. The related intraseasonal variations encompass a large–scale envelope of disturbances with zonal wavenumbers between one and five and an oscillation period between 30 and 90 days (Wheeler and Kiladis 1999; Zhang 2005). The quasi–oscillatory cycle at low latitudes can be initiated at any longitude (Matthews 2008) but typically starts with a large–scale organization of convection over the Indo–Pacific (Madden and Julian 1972, 1994; Zhang 2005). Related convective clusters are quasi–stationary or slowly move eastward from the Indian towards the Pacific Ocean. When the anomaly reaches the date line, convection typically subsides due to cooler ocean currents (Hendon and Salby 1994), but can still linger on for some days in case of an amplifying El Niño (e.g. Roundy and Kravitz 2009). Afterwards, the dynamical signal of the MJO prevails and typically continues eastward towards the Atlantic Ocean and along the whole equatorial circumference (Madden and Julian 1994; Zhang 2005).

b. Seasonality of the MJO

The MJO cycle undergoes seasonal modulations in both the strength and the geographical position of related disturbances (Zhang 2005). The intraseasonal tropical variability peaks

over the Indo–Pacific during boreal winter (Madden and Julian 1994; Zhang 2005). At that time of the year, convective activity in the South Pacific convergence zone strongly enhances the MJO activity over the western Pacific (Zhang and Dong 2004). Typically, the convection is then reinforced by an equatorial Kelvin wave approaching from the west (Masunaga 2007). As a result, off–equatorial boundary layer convergence is primarily enhanced during boreal winter (Masunaga 2007) and frictional moisture convergence leads to strong convection (Maloney and Hartmann 1998; Kiladis et al. 2005).

The MJO features a latitudinal migration across the equator which produces a secondary maximum during boreal summer (Zhang and Dong 2004). This maximum resides north of the equator and can be split up into two peak regions, one from the Bay of Bengal east of India to the South China Sea and one over the eastern Pacific Ocean off the Central American coast. The transition into the regime of the former peak sets off the summer monsoon over India and Australia (Madden and Julian 1994). Wu et al. (2006) reproduced the latitudinal migration of the MJO in a circulation model which was forced by eastward propagating heating anomalies. They attributed the propagation between the solstices to seasonal changes in the tropical background state, i.e. changes in the vertical wind shear and the static stability. Hendon et al. (2007) found the MJO activity during boreal spring to be least asymmetric with respect to the equator and to be most sensitive to warm sea surface temperatures of the Pacific.

c. Motivation of nonlinear principal component analysis for the MJO

It has become common practice to identify the MJO with a linear principal component analysis (LPCA). The LPCA can be either univariate or multivariate (Waliser et al. 2009), i.e. the linear principal components (LPCs) can be defined either separately or jointly for the fields under investigation. For example, Maloney and Hartmann (1998) and Matthews (2000) used LPCs in the 850–hPa zonal winds whereas Slingo et al. (1999) and Kessler (2001) used LPCs in the outgoing longwave radiation (OLR). Wheeler and Hendon (2004, hereafter referred to as WH04) created the real–time multivariate MJO time series (RMM) which combines the zonal wind at 850 hPa (hereafter U850), the zonal wind at 200 hPa (hereafter U200) and OLR.

In general, the lowest order of a propagating wave is described by the leading pair of LPCs. A two–dimensional hyperplane is fitted to the data in the high–dimensional input space by LPCA and the MJO state is defined by a single point on this plane. Such a two–mode detection is convenient, since it decomposes the MJO signal into amplitude and phase (e.g. WH04). As the case may be, the MJO may feature some characteristics which only recur in a particular section of its cycle and are only resolved by one of the higher–order LPCs. These characteristics are hence missed in the LPCA, since they do not project onto the two leading LPCs. In contrast to LPCA, nonlinear principal component analysis (NLPCA) can account for more complex interactions of the input sources which can be illustrated, in the simplest case, by second–order mode interactions of a set of first–order mode interactions (e.g. Kramer 1991). Depending on the set of input modes, NLPCA can hence resolve a mixture of features which can project on all possible dimensions in the input space. In such a

way, transient recurring deviations in a particular stage of the MJO can be better resolved, if they persistently appear in a particular region on the two–dimensional hyperplane addressed above.

NLPCA outputs curved solutions between data clusters in the input data space (e.g. Kramer 1991; Hsieh 2001). In the present study, NLPCA is configured to seek a circular solution which provides a straightforward mapping of the archetypal MJO cycle. Mean states of the MJO cycle are solely prescribed by the circular phase of the nonlinear principal component (NLPC). In atmospheric sciences, NLPCA has been applied e.g. to ENSO in the tropical Pacific (Monahan 2001; An et al. 2005), to the quasi–biennial oscillation in the tropical stratosphere (Hamilton and Hsieh 2002) or to circulation regimes in the northern hemisphere (Teng et al. 2007). The NLPCA method is based on neural network modeling, which is increasingly used in the environmental sciences (Marzban and Stumpf 1996; Krasnopolsky 2007; Haupt et al. 2009; Hsieh 2009).

d. Outline of the current study

The objective of the current study is twofold. On the one hand, NLPCA is introduced as an innovative method to analyze zonal characteristics of the archetypal MJO cycle. On the other hand, it is strived to provide further evidence for significant seasonal alterations of the MJO cycle. To this end, fields of U850, U200 and OLR are stratified into seasons. NLPCA is applied to the four subsets and significant deviations are sought among the four solutions.

The data body in this study is analogous to the RMM of WH04. However, WH04 used a

combination of single filters with an eye toward the real–time application. In contrast, our study resorts to a conventional Lanczos filter (Duchon 1979) which provides a much sharper bandpass cut–off. Hence, more intraseasonal variability is captured by the leading LPCs which allows for the examination of a slightly larger subset of modes. Such a finer–scale analysis might be especially crucial for the Southeast Asian region of Malaysia, Indonesia and the Philippines, which is commonly denoted the Maritime Continent. The variability of convective structures is extraordinarily high over the scattered islands of the Maritime Continent (Wang and Rui 1990) and the MJO turns out to be strongly influenced by local synoptic–scale structures (Chang et al. 2005).

The structure of the present study is as follows: The neural network model used for NLPCA is briefly described in Section 2. The underlying datasets together with the approach used to extract the LPCs and later the NLPC are described in Section 3. The MJO cycle derived for the whole year is shown in Section 4. Seasonal characteristics are then highlighted in difference plots and examined carefully in Section 5. Finally, a discussion on the main findings is presented in Section 6. Details on the neural network model are presented in the Appendix.

2. Outline of nonlinear principal component analysis

Principal component analysis is a dimensionality reduction technique and identifies data axes which account for the maximum variance within a data cloud. Traditionally, LPCA is applied which seeks a straight–line approximation to a dataset. Nowadays, this linear framework can be extended to a nonlinear one which seeks a curve instead of a straight

line. A nonlinear mapping sequence to a curved solution can be provided by e.g. an auto–associative multilayer perceptron (MLP) neural network model.

In the common setup, the MLP consists of three hidden layers of neurons which are represented by small circles in Fig. 1. The bottleneck layer in the center of the MLP provides the dimensionality reduction. It is preceded by an "encoding" layer and succeeded by a "decoding" layer. Four layers of transfer functions (see Appendix) pass information between the input ($\mathbf{x}$), the hidden layers and the output ($\mathbf{x}'$).

A curve is constructed by allowing only for a single degree of freedom in the bottleneck layer. In case of a conventional NLPCA, the bottleneck contains a single neuron and the solution is restricted to an open curve (Kramer 1991). In case of a circular NLPCA (denoted NLPCA.cir, used in the present study), the bottleneck contains two neurons. Their values p and q are constrained to lie on a unit circle and to be centered around zero (see Appendix). Hence, a closed curve is obtained and the circular angle θ in the bottleneck remains the only degree of freedom (Kirby and Miranda 1996).

The complexity of the MLP is governed by the number of neurons in the "encoding" and "decoding" layers (denoted m, typically the same for both layers). In the present study, the information load is kept constant throughout the MLP layers, except for the bottleneck layer (as displayed in Fig. 1). Hence, m is chosen equal to the number of input variables. With this configuration, there are $2m^2 + 4m$ weight parameters (represented by connecting lines in Fig. 1) and $3m + 2$ offset parameters which are both determined by training.

The training is performed by minimizing the cost function J (see Appendix). To measure the departure between the input and the output of the MLP, J contains the mean squared error (MSE) between $\mathbf{x}'$ and $\mathbf{x}$. To enable a regularization during the training, J also contains

a weight penalty term which regulates the wiggliness of the curved solution (Hsieh 2001). The cost function is minimized with an ensemble of optimization runs carried out with a varying weight penalty. As proposed in Hsieh (2007), the most eligible run is then selected by a holistic information criterion H. The explicit formulation of H and more details on the entire optimization procedure are given in the Appendix. Further information about the architecture and the configuration of the MLPs used for NLPCA and NLPCA.cir can be found in Hsieh (2007, 2009).

3. Description of the data body and analysis

a. Input data and filtering

The tropical circulation of the MJO is driven by convective activity and baroclinic instability in the equatorial plane (Madden and Julian 1972, 1994). The related dynamical signal is discernible in various atmospheric proxies. Convection can be detected by OLR (inverse relationship), whereas baroclinity can be detected by a phase shift between the U850 and U200 fields. As exemplified by Rui and Wang (1990) or WH04, a combination of these three fields can be used to derive the amplitude and phase of the MJO. Waliser et al. (2009) recommended the MJO diagnostics to be likewise based on the three proxies plus precipitation. In this context, however, precipitation is only considered a necessary proxy in model simulations which typically fail to resolve convective clouds.

The OLR data are retrieved from the National Oceanic and Atmospheric Administration (NOAA), where daily OLR fields were estimated from advanced very high resolution

radiometer (AVHRR) measurements of polar orbiting satellites. Missing values which emanated from insufficient data recording or global satellite coverage were removed by a combined tempo–spatial interpolation (Liebmann and Smith 1996). The U850 and U200 data were taken from 12 UTC fields in the model reanalysis (Kalnay et al. 1996) jointly produced by the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). The analyzed data continuously cover 30 years from 1979 to 2008 with a total of 10958 days.

The signal of the MJO can be separated from other atmospheric circulation patterns by applying a bandpass time filter (e.g. Woolnough et al. 2000). To conform to the standard diagnostics given in Woolnough et al. (2000) and Waliser et al. (2009), we apply a Lanczos bandpass filter (Duchon 1979) with a cut–off at 20 and 100 days. As many as 500 filter points are used to provide a sharp filter response and to minimize the residual variance outside the prescribed frequency window. The filter is applied separately to each grid point and each variable. As a result of the filtering, synoptic weather is excluded on the shorter time scales, whereas seasonal to interannual to decadal climate variability is excluded on the longer time scales.

Subsequently, all data are averaged over 15°S–15°N which leaves a total of 144 meridional averages along the equator. Finally, OLR, U850 and U200 are standardized locally to zero mean and globally to unit variance by subtracting the individual mean and dividing by the overall standard deviation. Consequently, each of the three fields contributes equally to the definition of the LPCs and NLPCs described below.

b. Linear principal components

To compress the data for the NLPCA analysis, LPCs, with eigenvectors denoted empirical orthogonal functions (EOFs), are computed using all 10958 days. The four leading LPCs together explain more than 55% of the total variance. More precisely, they account for 21.9%, 21.2%, 6.8% and 5.8% of the daily variance, respectively. The physical processes in the higher–order modes are not consistent over time (not shown) and are not discussed henceforth.

The spatial structure of EOFs 1–4 is displayed in Fig. 2 with the eigenvectors weighted by the original field variances. The leading two EOFs only differ marginally from their RMM counterparts discussed in WH04. For a comparison, the eager reader is referred to Fig. 1 in WH04. For the RMM, a different type of filtering is used which leaves a significant proportion of variance with time periods beyond 100 days. Thus, the total variance (the MJO signal) is higher (weaker) in the LPCA of WH04 than in the LPCA of our study.

Physically, EOF1 depicts a pronounced convection anomaly with deviant OLR values from the eastern Indian Ocean over the Maritime Continent to the western Pacific Ocean (Fig. 2a). In accordance with enhanced (suppressed) convection, there is a convergent wind anomaly on lower (upper) levels and a divergent wind anomaly on upper (lower) levels. In other words, wind anomalies are of opposite sign on lower and upper levels and of opposite sign over the Indian and Pacific Ocean. EOF2 depicts a convective anomaly with changing sign near the location of the convection in EOF1. More precisely, EOF2 displays a strong OLR anomaly over the eastern Indian Ocean and a moderate OLR anomaly over the western Pacific Ocean (Fig. 2b). In comparison to EOF1, wind patterns are shifted to the east by

roughly 60°. In particular, U850 and U200 anomalies are of opposite sign over the Maritime Continent which contributes to subsidence and atmospheric lifting on either side of the peak wind anomalies. EOF3 displays an even more refined spatial pattern than EOF2. Convection is somewhat enhanced (suppressed) over parts of the Indian and Pacific Oceans while it is somewhat suppressed (enhanced) over the Maritime Continent (Fig. 2c). Most prominent in EOF3, U200 diverges considerably from the mean over the eastern Indian Ocean, parts of the Maritime Continent and the very east of the Pacific Ocean. Upper–level winds at the tropopause level can therefore provide an outflow from the convective active regions. In EOF4, the convective pattern is similar to EOF3, but displaced by roughly 30° to the west. Convection is slightly enhanced (suppressed) over the eastern Maritime Continent and slightly suppressed (enhanced) over the eastern Indian Ocean (Fig. 2d). The most interesting wind feature in EOF4 relates to a pronounced U200 anomaly over parts of South America, the Atlantic Ocean and parts of Africa.

Generally, spatial structures of convection and wind are somewhat similar, but shifted along the equator, in EOF1 and EOF2. The same applies to EOF3 and EOF4. EOF1 and EOF2 reveal wavenumber one structures whereas EOF3 and EOF4 reveal wavenumber two structures.

c. Nonlinear principal components

The four leading LPCs discussed above are used as input variables for NLPCA.cir with $m = 4$, as described in Section 2. Even though the LPCs are by construction uncorrelated, they are not statistically independent and the related spatial modes can interfere constructively

over time. The scaling of the daily time series is such that their variances over the entire 30 years equal the eigenvalues of the LPCs. To save computing time, the daily coefficients are aggregated to non–overlapping pentads, i.e. averaged over disjoint chunks of five days (as e.g. in Maloney and Hartmann 1998; Zhou and Miller 2005). During leap years, the 12th pentad is averaged over six days instead of five days. Since the present study has the focus on a sufficiently strong MJO signal, a simple criterion is applied for the exclusion of weak or quiescent MJO periods. Similarly to WH04, the MJO amplitude is limited to

$$\sqrt{\mathrm{LPC1}^2 + \mathrm{LPC2}^2}/\mathrm{std}(\mathrm{LPC1}) > 0.5\,, \tag{1}$$

where LPC1,2 refer to the time coefficients of the LPCs and std denotes the standard deviation. Note that the threshold of 0.5 in Eq. 1 is in the order of formerly selected values ranging from 0.4 (used in Matthews 2008) to 1 (e.g. used in WH04). The thresholding appears to be necessary, because the training of the neural network is elsewise too much affected by the noise during periods of weak MJO activity. The four LPCs are fed into the input layer $\mathbf{x}$ of NLPCA.cir (Fig. 1). The output LPC values are derived from $\mathbf{x}'$ and describe a closed curve in the 4–dimensional LPC space.

To investigate seasonal variations, the whole dataset is stratified into seasons whose definition relates to the northern hemisphere: winter (DJF, annual pentads 68–73, 1–12), spring (MAM, annual pentads 13–30), summer (JJA, annual pentads 31–49) and fall (SON, annual pentads 50–67). Firstly, NLPCA.cir is run for all seasons together to provide an all–year definition of the active MJO cycle. Secondly, NLPCA.cir is run for each season separately. The all–year solution is used as a reference and the seasonal solutions are compared to it. Due to a weak seasonal cycle of the MJO amplitude and the criterion shown in Eq. 1, the to-

tal sample size slightly varies across seasons: 411 pentads during winter, 442 pentads during spring, 402 pentads during summer and 405 pentads during fall. The information criterion H (given in the Appendix) determines the regularization in NLPCA.cir which turns out to be slightly stronger during winter and spring than during summer and fall.

The MJO phase, as defined by the angle θ of the circular bottleneck node, can be mapped over a full cycle to display the whole quasi–periodic NLPCA.cir solution. If multiple runs are to be compared, it is necessary to align the phase, i.e. to set a reference phase. In the present study, the zero phase $\theta = 0$ denotes LPC2 $= 0$ with LPC1 < 0 which agrees with the MJO phase definition in WH04 and reflects the Indian Ocean as the most common starting location for an active event. Accordingly, the phase groups 1–8 introduced in WH04 cover the range of $0 < \theta \leq 2\pi$ consecutively.

As a test for the consistency of our results, the seasonal NLPCA.cir was also run with LPCs defined separately over individual seasons instead of jointly over the whole year. Allowing for slight differences due to the nonlinear optimization in NLPCA.cir, the same results were obtained with the seasonal LPCs as with the annual LPCs.

The temporal consistency, i.e. the sample uncertainty, is investigated by a Monte–Carlo resampling. To this end, the pooled and the seasonally stratified data separately undergo a residual bootstrap. In other words, the residuals (input data minus the NLPCA.cir output) are randomly added to the output of NLPCA.cir (hereafter denoted the full solution). The time series of the residuals is stationary, but features weak autocorrelation. Therefore the residuals are resampled in blocks of 10 pentads. 250 bootstrap repetitions of the NLPCA.cir calculation are made available and confidence intervals are computed from bias–corrected percentiles (as explained in e.g. Efron 1987, with extensions described below).

Some computational complications emerge with the underlying bootstrap samples, as they typically emulate multimodal distributions, i.e. are heteroscedastic and diverge considerably from normality. To provide sufficiently consistent and continuous confidence intervals, the bootstrap samples are slightly smoothed with a Gaussian kernel. Note that such a smooth bootstrap was discussed by e.g. Hall et al. (1989) or De Angelis and Young (1992). The kernel bandwidth is defined by the normal reference bandwidth selector (as proposed in Zhang and Wang 2009) which explicitly accounts for departures from normality. Since the multimodality in the bootstrap samples precludes a transformation to the normal distribution (e.g. Efron 1987), the end points of the intervals are adjusted directly in the kernel density estimate rather than in a standard normal density. In such a way, the commonly applied bias correction (Efron 1987) is refined to work without any parametric assumptions.

4. All–year MJO

The use of NLPCA.cir provides a clearly arranged mapping of the active MJO cycle based on the angle in the circular bottleneck node. In the following, firstly, a time series of the circular angle θ is presented (Fig. 3). Secondly, the cycle $0 < \theta \leq 2\pi$ is broken down both into the output LPC values (Fig. 4) and the related physical fields (Fig. 5).

In Fig. 3, the MJO phase is shown for the active MJO episodes over the analyzed period. Note that the alignment of the phase is described at the end of Section 3. The intermittent time series of active pentads reflects the highly episodic nature of primary and successive MJO events, as described in Matthews (2008). Almost 76% of the pentads indicate a sufficiently active MJO which meets the criterion defined in Eq. 1. Over the years, the percentage

of active pentads varies between 60% in 1984 and 92% in 1979. As derived from the progression of the phase θ, the mean phase speed constitutes 4100 km per pentad or 9.5 ms^{-1}. To travel around the equator, the mean period for a full MJO cycle therefore is 48 to 49 days. The phase speed is not totally uniform over θ, but most phase segments are passed evenly and fractional deviations from the mean propagation speed are not significant over θ (not shown). As it can be seen in Fig. 3, the phase shows a steady and coherent temporal evolution. Except for few cases with a very short retrograde propagation, θ is strictly increasing over MJO periods, indicating a perpetual eastward motion. This demonstrates the ability of NLPCA.cir to model the quasi–periodicity of the MJO.

The degree of nonlinearity in the annual NLPCA.cir solution becomes evident in the interplay of the four output LPCs. The coefficients for the LPCs define the NLPC and are plotted with 90% confidence intervals over a full MJO cycle in Fig. 4. The leading pair of LPCs has an internal phase lag of $\frac{\pi}{2}$ and directly represents the RMM described in WH04. The second pair of LPCs is considered insignificant and therefore not discussed for the annual solution. The all–year mode is almost perfectly sinusoidal in LPC1,2 and exhibits very narrow confidence limits which signify low uncertainties. Thus, the solution can be well approximated by a circle in the LPC1,2 plane.

For a comparison with the methodology of WH04, the MJO phase is also derived from the angle which is formed solely by the leading pair of LPCs. This definition of the MJO phase directly corresponds to the RMM in WH04 except for the small deviations in the EOFs as discussed in Section 3. The differences between the two MJO definitions turn out to be unimportant over the whole year. The methodology of WH04 (with our time filtering) explains 99.9% of the variability of the MJO phase θ as derived from the NLPC.

If the output LPCs from the NLPCA.cir model are multiplied by the related EOFs, anomaly fields are reconstructed in the physical space. To this end, a Hovmoeller diagram is adopted to show the evolution of U850, U200 and OLR over θ (Fig. 5). Multiple studies already presented composites of dynamical fields for distinct MJO phases and seasons (e.g. Maloney and Hartmann 1998; Wheeler and Hendon 2004; Zhou and Miller 2005). These analyses are revisited and extended here.

Below, the physical fields of the mean annual MJO are discussed briefly. Generally speaking, wind anomalies extend throughout the eastern and western hemispheres. Convective anomalies (as detected by OLR) almost exclusively occur in the eastern hemisphere, but are also visible in a weakened form over South America. The zero phase is such that the cycle starts with a decaying MJO event over the Maritime Continent and a developing MJO event over Africa. For a comparison, here, the eager reader is referred to the seasonal composites in WH04. In the first half of the displayed cycle, strong convective activity develops over the Indian Ocean (Fig. 5c). After $\frac{3}{4}\pi$, convection peaks out to the south of India and subsequently spreads out over the Maritime Continent to the east. In the second half of the displayed cycle, high clouds gradually vanish over the Indian Ocean, as indicated by rising positive OLR anomalies. After $\frac{7}{8}\pi$, cloudiness bottoms out to the south of India. Subsequently, the Maritime Continent also increasingly receives conditions with a clear sky or merely low clouds.

The dynamics of the MJO are characterized by a strong coupling between the circulation patterns and the convective centers (Kiladis et al. 2005; Zhang 2005). More precisely, the zonal wind field is such that convection is supported by convergence on 850 hPa and divergence on 200 hPa. On the contrary, clear sky conditions are supported by divergence on 850

hPa and convergence on 200 hPa. Also, the sign of the U200 anomalies typically is opposite to the U850 anomalies which indicates strong baroclinity. The strongest wind oscillations on 850 hPa have an amplitude of more than 2 ms^{-1} and occur over the Maritime Continent (Fig. 5a). The strongest wind oscillations on 200 hPa have an amplitude around 4.5 ms^{-1} and occur over the western Indian Ocean and the far east of the Pacific Ocean (Fig. 5b).

The slope of the contours with respect to θ varies among variables and longitudes in Fig. 5. In the vicinity of the Maritime Continent, contours are steepest with respect to θ, indicating slowest propagation of the MJO signal. This confirms previous studies who found a phase speed as slow as 5 ms^{-1} over the Indo–Pacific (Milliff and Madden 1996; Matthews 2000). Over Africa and the western Indian Ocean, contours are comparatively flat with respect to θ, but steeper in U850 (Fig. 5a) than in U200 (Fig. 5b). This confirms previous observations that the vertical structure of the intraseasonal disturbances changes temporarily while the disturbances move over the Indian Ocean (Kiladis et al. 2005).

The gray–shaded areas in Fig. 5 delineate the phase space with a significant signal beyond the first pair of contours ($\pm$) closest to zero. (Note that the reference levels used are nonzero, i.e. the significance test is more stringent than if the zero reference level was used.) It can be seen that the physical fields reveal a consistent behavior and uncertainties in the location of both convective and circulation anomalies are small. Notedly, low–level wind anomalies are consistently very weak over the Atlantic (Fig. 5a). This confirms previous observations that the MJO only maintains a circumferential propagation in its upper–level fields (Zhang 2005).

5. Seasonal MJO

The stratified analysis outlined in Section 3 is discussed as follows: The output LPCs are evaluated for individual seasons (Fig. 6) and compared to the all–year solution. The characteristics of the physical fields are highlighted by means of difference plots (Fig. 8). Finally the variance of physical values is analyzed for individual longitudes (Fig. 9).

The phase speed of the MJO varies somewhat throughout the year. In our dataset, the mean period for a full MJO cycle changes from 52 days (winter and summer) to 43 days (boreal spring) and 49 days (boreal fall). The estimated confidence intervals turn out to be very narrow for the period length (not shown). As a result, the difference in period length is highly significant during springtime when the MJO cycle tends to be completed much faster.

The evolution of the leading pair of LPCs turns out to be consistently sinusoidal during boreal winter and spring, but exhibits a somewhat different phasing during boreal summer and fall (leftmost panels in Fig. 6). More precisely, the peak contribution of LPC1 is slightly shifted in the second half of the calendar year. The maximum in the coefficients appears at $\theta = \frac{7}{6}\pi$ during boreal summer and fall which is $\frac{\pi}{6}$ later than during the other two seasons. The minimum in the coefficients has the same lag during boreal summer, i.e. appears at $\theta = \frac{\pi}{6}$, but it lacks a shift during boreal fall. This observation shows that the transition from the convective peak over the Indian Ocean to that one over the Maritime Continent is somewhat slower during boreal summer and fall.

In contrast to the all–year solution (Section 4), the second pair of LPCs contributes to the mean MJO cycle during individual seasons (rightmost panels in Fig. 6). Most notably, the coefficients for LPC3 are significantly different from zero and reverse sign semiannually in

solstitial seasons. LPC3 contributes with negative (positive) coefficients for $\theta = \pi$ ($\theta = 2\pi$) during boreal winter. Contrary to this phasing, LPC3 contributes with positive (negative) coefficients for $\theta = \pi$ ($\theta = 2\pi$) during boreal summer. Also notably, the coefficients for LPC3 are significantly different from zero for $\theta = \frac{\pi}{2}$ and $\theta = 2\pi$ during boreal fall. The seasonal influence of LPC4 is weaker than that of LPC3. The coefficients for LPC4 are close to zero but still entail significant values for $\theta = \frac{3}{2}\pi$ during boreal winter, for $\theta = \pi$ during boreal spring and for $\theta = \frac{\pi}{2}$ and $\theta = \frac{7}{4}\pi$ during boreal summer. Thus, oscillations in LPC3 and LPC4 jointly modulate the MJO cycle during solstitial seasons whereas only one of the two modes has a weak impact during equinoctial seasons.

The degree of nonlinearity turns out to be highest during boreal winter when the second pair of LPCs contributes most effectively. In Fig. 7, it is shown how the wintertime solution spears through the data cloud. The four–dimensional input space is displayed by means of four mutually perpendicular cross–sections. Fig. 7a reveals a hole in the data cloud due to the thresholding of LPC1,2 and shows that the circle in the LPC1,2 plane is only approximately uniform. Fig. 7b demonstrates that the solution oscillates back and forth on the same path in the LPC1,3 plane, i.e. that the solution is not circular in this projection. The coefficients of LPC3 appear to be negatively correlated with those of LPC1. Fig. 7c reveals that the solution is elliptical in the LPC2,3 plane and Fig. 7d reveals that the solution varies along an irregular closed curve in the LPC1,4 plane.

The all–year solution explains more than 90% of the intraseasonal variance within the seasonal solutions. Percentages account for 90.7% during boreal winter, 97.4% during boreal spring, 91.3% during boreal summer and 96.1% during boreal fall. Thus, the MJO during boreal winter (spring) entails the most (least) additional intraseasonal variability which is

not contained in the all–year definition of the MJO. The added value in explained variance is around 9% during solstitial seasons but only around 3% during equinoctial seasons. Notedly, the seasonally defined MJO phase diverges marginally from the conventional approach of WH04, since around 99% of the variability in θ can be explained by the angle in the LPC1,2 plane.

To highlight MJO modulations throughout the year, the differences to the all–year solution are displayed for the physical fields. The offset between the full solutions (seasonal minus all–year) is outlined with contours in Fig. 8. Clearly, seasonal variations depend on the variable under investigation and occur at varying longitudes. Seasonal differences in U850 and OLR are most evident over the Indo–Pacific. Seasonal differences in U200 are most evident over the Indo–Pacific and the far eastern Pacific Ocean. The observed wind differences typically reverse their sign from low to high levels which indicates that most seasonal irregularities are baroclinic in nature.

To detect both significant and physically relevant seasonal modulations, the statistical tests for the physical fields are applied with a reference level other than zero. In Fig. 8, significant seasonal differences beyond the pair of contours ($\pm$) closest to zero, are marked by a gray shading. Related areas cover the centers of seasonal irregularities and show that more than half of the peaks in the difference fields are significant.

During boreal winter, three significant differences with a zonally alternating sign are present in the OLR over the Indo–Pacific (Fig. 8c). At the same time, a single significant difference is present in the wind field in between the Indian Ocean and the Maritime Continent (Fig. 8a and Fig. 8b). Evidently, convection in the MJO cycle tends to establish earlier around Malaysia during boreal winter than during the rest of the year. Also, the onset of

enhanced low–level westerlies tends to be earlier than usual over the Bay of Bengal east of India. During boreal summer, there is an east–west dipole of significant differences in U850 and OLR over the Indo–Pacific (Fig. 8g and Fig. 8i). Simultaneously, a single significant difference emerges in U200 over the Indian Ocean (Fig. 8h). If differences in Fig. 8i are compared to absolute anomalies in Fig. 5c, it becomes evident that clouds tend to clear up less after a major convective outbreak over the Maritime Continent during boreal summer.

MJO modulations in upper–level winds clearly reflect the bimodal wind structure of EOF3 with coherent anomalies around Indonesia and the eastern Pacific Ocean (e.g. $\theta = 2\pi$ during boreal winter or $\theta = \pi$ during boreal summer). In some stages of the MJO, the related anomaly over the eastern Pacific Ocean is diminished by other modes (e.g. $\theta = \pi$ during boreal winter or $\theta = \frac{\pi}{2}$ during boreal fall). If differences in Figs. 8b, 8e, 8h and 8k are compared to absolute anomalies in Fig. 5b, it becomes evident that the seasonal modulations can make the MJO signal both stronger and weaker, as seen e.g. in the middle of the MJO cycle ($\theta = \pi$) over the Indian Ocean (stronger during boreal winter and weaker during boreal summer).

MJO modulations mostly remain unimportant during boreal spring, except for a significant upper–level wind difference which emerges for $\theta = \frac{5}{4}\pi$ between the Atlantic Ocean and Africa (Fig. 8e). The modification of U200 can be attributed to the wind anomaly in EOF4 and brings about an earlier onset of enhanced upper–level westerlies over western Africa.

To identify longitudes with a noticeably enhanced or suppressed MJO cycle, the impact of the transient seasonal departures in Fig. 8 is summed up in Fig. 9. In other words, seasonal irregularities in the standard deviation of physical values are displayed along the equator (thick central lines in Fig. 9). As before, bootstrap confidence intervals are employed to

estimate significances (thin lower and upper lines in Fig. 9).

Over the eastern hemisphere, the standard deviation of the mean MJO cycle tends to be highest during boreal winter indicating an on average enhanced MJO signal (first panel in Fig. 9). Wintertime OLR fluctuations are significantly increased in the mean MJO cycle around 120°E and 180°. The standard deviation in OLR is raised by roughly 3 Wm^{-2} at the longitude of the Philippines whereas it is still raised by more than 2 Wm^{-2} at the date line. At the same time, wind fluctuations are significantly increased in the mean MJO cycle over the eastern Indian Ocean where the standard deviation of U200 is raised by 1 ms^{-1}.

During boreal summer, local oscillations embedded in the mean MJO cycle are clearly suppressed over the eastern hemisphere whereas they are somewhat enhanced over the western hemisphere (third panel in Fig. 9). In particular, convective activity oscillations are reduced over the Maritime Continent during boreal summer. Over Indonesia and Malaysia, the standard deviation in OLR is more than 2 Wm^{-2} lower than usual. Simultaneously, the variance in the wind fields is weaker over the Indian Ocean. Here, upper–level wind fluctuations are diminished by roughly 0.8 ms^{-1}. Convective activity fluctuations appear unusually strong around the date line during boreal summer. However, the local increase in the variance at 180° is smaller and slightly less significant than during boreal winter. Interestingly, upper–level wind fluctuations are significantly increased in the mean MJO cycle over the eastern Pacific Ocean during boreal summer. The local increase in the standard deviation amounts up to 0.6 ms^{-1} in U200.

During equinoctial seasons, significant departures in the standard deviation are weak in the mean MJO cycle (second and fourth panels in Fig. 9). During boreal spring, convective activity oscillations are somewhat intensified (weakened) over the Indian Ocean (western

Pacific Ocean). Also, springtime upper–level wind oscillations are weakly but significantly enhanced over South America. During boreal fall, upper–level wind oscillations are somewhat weakened from the Atlantic region over Africa to the Indian Ocean. However, the latter only turn out to be significant over the Indian Ocean.

6. Discussion of methodology and results

In the present study, NLPCA has been based on a neural network with a circular bottleneck node (denoted NLPCA.cir). NLPCA has been introduced to assign a circular phase to the MJO cycle and to identify the exemplary MJO structure along the equator. Thirty years of meridionally averaged zonal wind (U850, U200) and convection (OLR) have been conditioned on a sufficiently active stage of the MJO. The information content has been compressed into four LPCs which have been fed into NLPCA.cir with the dataset both as a whole and stratified into seasons. Conclusions on the seasonal modulations of the MJO have been drawn from the differences between the seasonal subsets and the entire dataset. Statistical inference has been based on a Monte–Carlo resampling, i.e. on a residual bootstrap and bias–corrected confidence intervals.

The question arises as to what added value is offered by the new methodology which displays the zonal structure of the mean MJO cycle in a single mode solution. As a matter of fact, NLPCA.cir only offers one degree of freedom, i.e. only leaves the phase θ varying over time. A threshold is utilized for the MJO activity, but all other attributes of the MJO are aggregated over time, i.e. smoothed out. In particular, the vertical structure is fixed in the output of NLPCA.cir and the horizontal structure is partly fixed by the meridional

average of the input data. In reality, the MJO constitutes a complex tempo–spatial phenomenon with multiple variations in not only amplitude and phase but also baroclinity and vertical structure (Kiladis et al. 2005), multiscale structure (e.g. Zhang 2005) or regionality and meridional propagation (Wu et al. 2006; Masunaga 2007). In such a way, the presented approach does not capture the momentary layout of the MJO as ideally required, but the approach successfully captures the mean MJO cycle, i.e. the typical MJO evolution along the equator.

Given some new input data, the NLPCA model both outputs the phase θ and the coefficients for the output LPCs. It is hence possible to link any θ directly to a spatial pattern along the equator. Accordingly, the presented approach does not require any compositing technique to keep track of the average appearance of the MJO at any stage of its cycle. This is regarded as the major advantage of NLPCA.cir in comparison to other MJO identification methods.

A linear superposition of the leading LPCs has proven of value to determine the current state of the MJO (e.g. WH04, Waliser et al. 2009). It is convenient to use only two leading LPCs which account for amplitude and phase and can be displayed in a graphical two–dimensional MJO chart (e.g. Fig. 7 in WH04). If the sample size is large enough, more than two LPCs can be sufficiently well–separated from those with neighboring eigenvalues (North et al. 1982). Accordingly, four LPCs have been analyzed in the present study. The leading pair explains 3.4 times more daily variance than the second pair. More surprisingly, LPC1,2 explain over 99% of the pentad variance in the all–year NLPC which questions the influence of LPC3,4 on the mean MJO cycle. Tests (not shown) with NLPCA.cir and only LPC3,4 as input indicate that the second pair of LPCs features a secondary quasi–periodic oscillation

with a much more irregular time period than the leading pair. LPC4 then exactly lags LPC3 by $\frac{\pi}{2}$. Since the oscillations of LPC1,2 and LPC3,4 are typically not phase–locked, positive and negative interference occur for individual seasons but cancel each other out throughout the year. The influence of LPC3 reverses its sign from one solstice to the other whereas the influence of LPC4 changes from one season to the other (Fig. 6). In such a way, the present study shows that the wavenumber–two oscillations in LPC3,4 alter the appearance of the mean MJO cycle over seasons. In particular, they most likely contribute to a meandering of the MJO envelope which is clearly associated with an interaction with ENSO (Kessler 2001). However, it remains striking how well a quadrature amplitude modulation of the first two LPCs captures the mean signal over the course of a year.

The change in phase speed, is found to be significant during boreal spring. The mean period of the MJO cycle turns out to be about 8 days shorter than during the rest of the year. Note that this finding is new, as no seasonal changes in the average period length could be detected so far (e.g. Madden and Julian 1994). The increase in phase speed might be related to the least meridional asymmetry of the MJO structures during springtime (Hendon et al. 2007). If the involved convective clusters are then most centered at the equator, their eastward translocation might be least obstructed by opposing trade winds.

Seasonal variations in the equatorial MJO were formerly linked to the regional formation of convective clusters and superclusters as well as their meridional translocation (Zhang 2005; Wu et al. 2006; Masunaga 2007). Given the meridional average of the input data, only zonal variations in the MJO structure have been considered in the present study. The three main findings on seasonal changes of the mean zonal MJO are outlined in the following.

Firstly, the MJO is most active over the Indo–Pacific during boreal winter (Madden and

Julian 1994; Zhang and Dong 2004; Zhang 2005; Masunaga 2007). As evident from Fig. 9, the variations in the mean MJO cycle are significantly enhanced over most parts of the eastern hemisphere during boreal winter. Two separate wintertime maxima appear during boreal winter: one centered around Indonesia and one centered around the date line. Wang and Rui (1990) clearly identified the equatorial Indian Ocean and the western Pacific convergence zone as preferred geographic locations for the development of intraseasonal convection anomalies. Likewise, Wang and Rui (1990) identified the Maritime Continent and the Central Pacific Ocean as regions of dissipation. The bimodal structure of wintertime MJO variability in the present study clearly confirms their observations.

Secondly, changes in the MJO activity split up into different regimes during boreal summer. As evident from Fig. 9, the variations of the mean are significantly reduced at most parts of the eastern hemisphere with respect to both OLR and the zonal wind. In contrast, the variations of the mean are significantly enhanced at the date line with respect to OLR and over the eastern Pacific Ocean with respect to the zonal wind. As a result, the summertime maximum of the MJO activity over the eastern Pacific off the Central American coast (Zhang and Dong 2004) is resolved by the stratified analysis with NLPCA.cir. In contrast, the summertime maximum over the Bay of Bengal (Zhang and Dong 2004) is not resolved. Tests (not shown) with two–dimensional LPCs show that the maximum over the Bay of Bengal mostly occurs to the north of 15°N and hence is missed in the current approach.

Thirdly, some weak modifications of the MJO behavior are highlighted during equinoctial seasons. Most interestingly, upper–level wind anomalies appear to be temporarily intensified between the Atlantic Ocean and Africa during boreal spring (Fig. 8e). The springtime wind increment causes the upper–level westerlies to be significantly intensified and turns out to

be the only highly significant seasonal modulation over the Atlantic region.

The seasonality of the MJO has been identified as an important driver for global climate modes such as ENSO (Zhang and Gottschalck 2002; Hendon et al. 2007; Roundy and Kravitz 2009). It remains to investigate, if ENSO is more influenced by the seasonal modulations of the mean MJO cycle or by individual MJO events occurring in a specific season.

Acknowledgments.

We cordially thank Matthew Wheeler for answering our queries at the beginning of the project. Our work was supported by a project grant from the Canadian Foundation for Climate and Atmospheric Sciences and a discovery grant from the Natural Sciences and Engineering Research Council of Canada.

APPENDIX

Configuration and optimization of the neural network model

With the input variables forming the 0th layer of the neural network (Fig. 1), a neuron $v_j^{(i)}$ at the ith layer ($i = 1, 2, 3, 4$) receives its value from the neurons $\mathbf{v}^{(i-1)}$ in the preceding layer, i.e.

$$v_j^{(i)} = f^{(i)}(\mathbf{w}_j^{(i)} \cdot \mathbf{v}^{(i-1)} + b_j^{(i)}) \,, \tag{A1}$$

where $\mathbf{w}_j^{(i)}$ is a vector of weight parameters and $b_j^{(i)}$ an offset parameter, and the activation or transfer functions $f^{(1)}$ and $f^{(3)}$ are the hyperbolic tangent functions, while $f^{(2)}$ and $f^{(4)}$ are simply the identity functions.

To obtain closed curve solutions, the bottleneck of the neural network (Fig. 1) holds two activation values p and q. An angular degree of freedom is defined by a circle constraint on p and q (Kirby and Miranda 1996; Hsieh 2001). To this end, the pre-states p_o and q_o are calculated by

$$p_o = \mathbf{w}_1^{(2)} \cdot \mathbf{v}^{(1)} + b_1^{(2)} \,, \quad \text{and} \quad q_o = \mathbf{w}_2^{(2)} \cdot \mathbf{v}^{(1)} + b_2^{(2)} \,, \tag{A2}$$

with

$$r = (p_o^2 + q_o^2)^{1/2} \,. \tag{A3}$$

Subsequently, the circular node is defined by

$$p = p_o/r\,, \quad \text{and} \quad q = q_o/r\,, \tag{A4}$$

which satisfies the unit circle equation $p^2 + q^2 = 1$.

To configure the MLP, the cost function J is minimized with

$$J = \langle \|\mathbf{x} - \mathbf{x}'\|^2 \rangle + \langle p^2 \rangle + \langle q^2 \rangle + P \sum_j \|\mathbf{w}_j^{(1)}\|^2\,. \tag{A5}$$

On the right hand side of Eq. A5, the first term is the mean squared error (MSE, with $\langle \cdots \rangle$ denoting a sample or time mean). The second and third terms force the bottleneck values to have an expectation of 0 which is applicable for the restricted configuration of NLPCA.cir (Hsieh 2001). The final term is a regularization term, with P the weight penalty parameter. Hsieh (2001) found that penalizing just the first layer of weights is sufficient to limit the degree of nonlinearity in the model.

The minimization of J is carried out with the quasi-Newton algorithm fminunc.m in the MATLAB Optimization Toolbox. To find the global minimum of J, an ensemble of 30 runs with a random initialization is computed separately for allocated choices of P with $0 \leq P \leq 1$. In each run, 15% of the data is randomly selected as test data and withheld from the training. If the MSE is higher in the test data than in the training data, ensemble members are discarded to avoid local overfitting.

Finally, the non–discarded ensemble members of the MLP are rated with a holistic information criterion H (Hsieh 2007) with

$$H = \left(1 - \frac{1}{2}\left[C(\hat{p}, \check{p}) + C(\hat{q}, \check{q})\right]\right) \text{MSE}\,, \tag{A6}$$

where $C(\hat{p}, \check{p})$ and $C(\hat{q}, \check{q})$ denote the correlation coefficients of the bottleneck values computed for all the nearest neighbor input data pairs $(\hat{\mathbf{x}}, \check{\mathbf{x}})$. The idea is to favor runs with

close neighbors in $\mathbf{x}'$, if they are likewise close in $\mathbf{x}$. The run with the lowest H is elected the final solution.

REFERENCES

An, S. I., W. W. Hsieh, and F. F. Jin, 2005: A nonlinear analysis of the ENSO cycle and its interdecadal changes. *J. Climate*, **18**, 3229–3239.

Chang, C. P., P. A. Harr, and H. J. Chen, 2005: Synoptic disturbances over the equatorial South China Sea and western Maritime Continent during boreal winter. *Mon. Wea. Rev.*, **133 (3)**, 489–503.

De Angelis, D. and G. A. Young, 1992: Smoothing the bootstrap. *Int. Stat. Rev.*, **60 (1)**, 45–56.

Duchon, C. E., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor.*, **18**, 1016–1022.

Efron, B., 1987: Better bootstrap confidence intervals. *J. Am. Stat. Assoc.*, **82 (397)**, 171–185.

Hall, P., T. J. DiCiccio, and J. P. Romano, 1989: On smoothing and the bootstrap. *Ann. Stats.*, **17 (2)**, 692–704.

Hamilton, K. and W. W. Hsieh, 2002: Representation of the quasi–biennial oscillation in the tropical stratospheric wind by nonlinear principal component analysis. *J. Geophys. Res.*, **107 (D15)**, doi:10.1029/2001JD001250.

Haupt, S. E., A. Pasini, and C. Marzban, (Eds.) , 2009: *Artificial Intelligence Methods in the Environmental Sciences*. Springer, 424 pp.

Hendon, H. H. and M. L. Salby, 1994: The life–cycle of the Madden–Julian Oscillation. *J. Atmos. Science.*, **51 (15)**, 2225–2237.

Hendon, H. H., M. Wheeler, and C. Zhang, 2007: Seasonal dependence of the MJO–ENSO relationship. *J. Climate*, **20 (3)**, 531–543.

Hsieh, W. W., 2001: Nonlinear principal component analysis by neural networks. *Tellus*, **53A**, 599–615.

Hsieh, W. W., 2007: Nonlinear principal component analysis of noisy data. *Neural Net.*, **20 (4)**, 434–443.

Hsieh, W. W., 2009: *Machine Learning Methods in the Environmental Sciences*. Cambridge Univ. Press, 349 pp.

Kalnay, E., et al., 1996: The NCEP/NCAR 40–year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77 (3)**, 437–471.

Kessler, W. S., 2001: EOF representations of the Madden–Julian Oscillation and its connection with ENSO. *J. Climate*, **14 (13)**, 3055–3061.

Kiladis, G. N., K. H. Straub, and P. T. Haertel, 2005: Zonal and vertical structure of the Madden–Julian Oscillation. *J. Atmos. Science.*, **62 (8)**, 2790–2809.

Kirby, M. J. and R. Miranda, 1996: Circular nodes in neural networks. *Neural Comput.*, **8**, 390–402.

Kramer, M. A., 1991: Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.*, **37**, 233–243.

Krasnopolsky, V. M., 2007: Neural network emulations for complex multidimensional geophysical meteor. atmos. phys.pings: Applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling. *Rev. Geophys.*, **45 (3)**.

Liebmann, B. and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.

Madden, R. A. and P. R. Julian, 1972: Description of global–scale circulation cells in tropics with a 40–50 day period. *J. Atmos. Science.*, **29 (6)**, 1109–1123.

Madden, R. A. and P. R. Julian, 1994: Observations of the 40–50–day tropical oscillation – A review. *Mon. Wea. Rev.*, **122 (5)**, 814–837.

Maloney, E. and D. L. Hartmann, 1998: Frictional moisture convergence in a composite life cycle of the Madden–Julian Oscillation. *J. Climate*, **11 (9)**, 2387–2403.

Marzban, C. and G. J. Stumpf, 1996: A neural network for tornado prediction based on Doppler Radar–derived attributes. *J. Appl. Meteor.*, **35 (5)**, 617–626.

Masunaga, H., 2007: Seasonality and regionality of the Madden–Julian Oscillation, Kelvin wave, and equatorial Rossby wave. *J. Atmos. Science.*, **64 (12)**, 4400–4416.

Matthews, A. J., 2000: Propagation mechanisms for the Madden–Julian Oscillation. *Quart. J. Roy. Meteor. Soc.*, **126 (569, Part A)**, 2637–2651.

Matthews, A. J., 2008: Primary and successive events in the Madden–Julian Oscillation. *Quart. J. Roy. Meteor. Soc.*, **134 (631, Part B)**, 439–453.

Milliff, R. F. and R. A. Madden, 1996: The existence and vertical structure of past, eastward–moving disturbances in the equatorial troposphere. *J. Atmos. Science.*, **53 (4)**, 586–597.

Monahan, A. H., 2001: Nonlinear principal component analysis: Tropical Indo–Pacific sea surface temperature and sea level pressure. *J. Climate*, **14**, 219–233.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110 (7)**, 699–706.

Roundy, P. E. and J. R. Kravitz, 2009: The association of the evolution of intraseasonal oscillations to ENSO phase. *J. Climate.*, **22 (2)**, 381–395.

Rui, H. and B. Wang, 1990: Development characteristics and dynamic structure of tropical intraseasonal convection anomalies. *J. Atmos. Science.*, **47 (3)**, 357–379.

Slingo, J. M., D. P. Rowell, K. R. Sperber, and E. Nortley, 1999: On the predictability of the interannual behaviour of the Madden–Julian Oscillation and its relationship with El Niño. *Quart. J. Roy. Meteor. Soc.*, **125 (554, Part B)**, 583–609.

Teng, Q., J. C. Fyfe, and A. H. Monahan, 2007: Northern Hemisphere circulation regimes: observed, simulated and predicted. *Clim. Dynam.*, **28 (7–8)**, 867–879.

Waliser, D., et al., 2009: MJO simulation diagnostics. *J. Climate*, **22 (11)**, 3006–3030.

Wang, B. and H. Rui, 1990: Synoptic climatology of transient tropical intraseasonal convection anomalies: 1975–1985. *Meteor. Atmos. Phys.*, **44 (1–4)**, 43–61.

Wheeler, M. and H. H. Hendon, 2004: An all–season real–time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132 (8)**, 1917–1932.

Wheeler, M. and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Science.*, **56 (3)**, 374–399.

Woolnough, S. J., J. M. Slingo, and B. J. Hoskins, 2000: The relationship between convection and sea surface temperature on intraseasonal timescales. *J. Climate*, **13 (12)**, 2086–2104.

Wu, M. C., S. D. Schubert, M. J. Suarez, P. J. Pegion, and D. Waliser, 2006: Seasonality and meridional propagation of the MJO. *J. Climate*, **19 (10)**, 1901–1921.

Zhang, C., 2005: Madden–Julian Oscillation. *Rev. Geophys.*, **43 (2)**.

Zhang, C. and M. Dong, 2004: Seasonality in the Madden–Julian Oscillation. *J. Climate*, **17 (16)**, 3169–3180.

Zhang, C. and J. Gottschalck, 2002: SST anomalies of ENSO and the Madden–Julian Oscillation in the equatorial Pacific. *J. Climate*, **15 (17)**, 2429–2445.

Zhang, J. and X. Wang, 2009: Robust normal reference bandwidth for kernel density estimation. *Statistica Neerlandica*, **63 (1)**, 13–23.

Zhou, S. T. and A. J. Miller, 2005: The interaction of the Madden–Julian Oscillation and the Arctic Oscillation. *J. Climate*, **18 (1)**, 143–159.

List of Figures

3 Time series of the MJO phase θ. The zero phase $\theta = 0$ corresponds to LPC2 = 0 with LPC1 < 0. The black curve depicts the analyzed active MJO periods whereas the gaps denote quiescent periods which are not investigated in our study. The steady progression of θ with time demonstrates the temporal coherency of the active MJO. 41

4 All–year NLPCA.cir solution displayed with its output LPCs. The phase θ $(0 \ldots 2\pi)$ is aligned as for Fig. 3. All coefficients are standardized by the standard deviation of LPC1. The central lines stand for the coefficients of the full solution, while the gray–shaded envelopes indicate pointwise 90% confidence intervals. 42

5 Spatial structure of the NLPCA.cir solution computed for the whole year. The panels show the anomalies of a) U850 (ms^{-1}), b) U200 (ms^{-1}) and c) OLR (Wm^{-2}). The ordinates show the phase θ progressing downward. Positive contours are drawn in black whereas negative contours are drawn in dark gray. The zero contour is omitted. Gray–shaded areas indicate the MJO signal to be significant beyond the first contour level ($\pm 0.25\mathrm{ms}^{-1}$ for U850, $\pm 0.5\mathrm{ms}^{-1}$ for U200, $\pm 2.5\mathrm{Wm}^{-2}$ for OLR) at the 95% significance level (one–sided test). 43

6 Seasonal NLPCA.cir solutions displayed with their output LPCs for a) winter, b) spring, c) summer and d) fall. The phase θ $(0 \ldots 2\pi)$ is aligned as for Fig. 3. All coefficients are standardized by the standard deviation of LPC1. The central lines stand for the coefficients of the full solutions, while the gray–shaded envelopes indicate pointwise 90% bootstrap confidence intervals. 44

7 NLPCA.cir solution (thick black line) displayed with the input data cloud (gray points) in varying planes. The graphs relate to boreal winter. The presented surfaces are spanned by a) LPC1 and LPC2, b) LPC1 and LPC3, c) LPC2 and LPC3 and d) LPC1 and LPC4. 45

8 Spatial variations of the NLPCA.cir solution computed for individual seasons. Contours depict the difference to the all–year solution presented in Fig. 5 (seasonal solution minus all–year solution). The panels a–c relate to winter, the panels d–f relate to spring, the panels g–i relate to summer and the panels j–l relate to fall. In each row, the first panel shows U850 (ms^{-1}), the second panel shows U200 (ms^{-1}) and the third panel shows OLR (Wm^{-2}). The ordinates show the MJO phase θ progressing downward. Positive (negative) contours are displayed in black (gray) and denote that the local value is higher in the seasonal (all–year) solution. Gray–shaded areas indicate differences beyond the first contour level ($\pm 0.25 ms^{-1}$ for U850, $\pm 0.5 ms^{-1}$ for U200, $\pm 2.5 Wm^{-2}$ for OLR) which are significant at the 95% level (one–sided test). 46

9 Departures in local variance as measured by differences in the standard deviation (std) along the equator (std of the seasonal solution minus std of the all–year solution). The black lines depict U850 (solid) and U200 (dashed–dotted) with differences in ms^{-1} (left scale). The dashed magenta lines depict OLR with differences in Wm^{-2} (right scale). The thick central lines correspond to the full solutions whereas the thin upper and lower lines refer to 90% bootstrap confidence intervals. 47

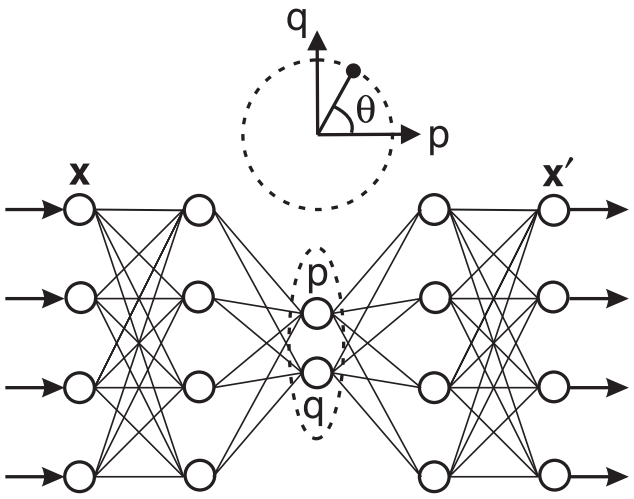

FIG. 1. A schematic diagram illustrating the neural network model for calculating the NLPCA with a circular node at the bottleneck (NLPCA.cir). The model is a standard feedforward neural network or multilayer perceptron with variables called neurons (marked by small circles). Four layers of activation or transfer functions map from the input layer ($\mathbf{x}$) to the output layer ($\mathbf{x}'$). The three hidden layers in between contain the so–called bottleneck layer sandwiched by the "encoding" and "decoding" layers. Two values p and q are assigned to the bottleneck nodes in the center and are forced to lie on a unit circle with a single degree of freedom described by θ. The connection between individual neurons is accomplished by weight parameters (marked by connecting lines). Additionally, offset parameters are included in the mapping functions (not displayed).

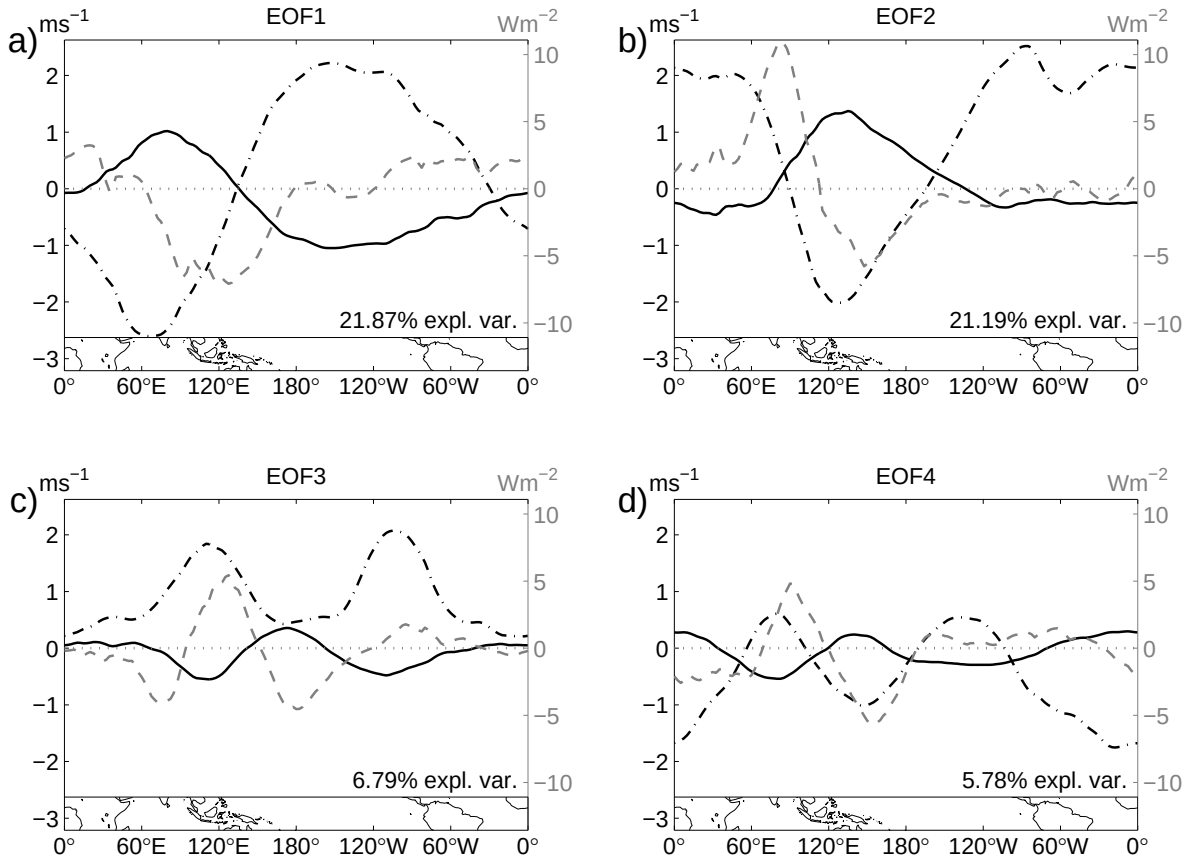

FIG. 2. Spatial structure of the leading empirical orthogonal functions (EOF1–4) obtained from the linear principal component analysis. The solid black and the dashed–dotted black lines represent U850 and U200, respectively (ms^{-1}, left scale). The dashed gray line depicts OLR (Wm^{-2}, right scale). The map at the bottom of the panels shows the area (15°S–15°N) over which the data were averaged meridionally.

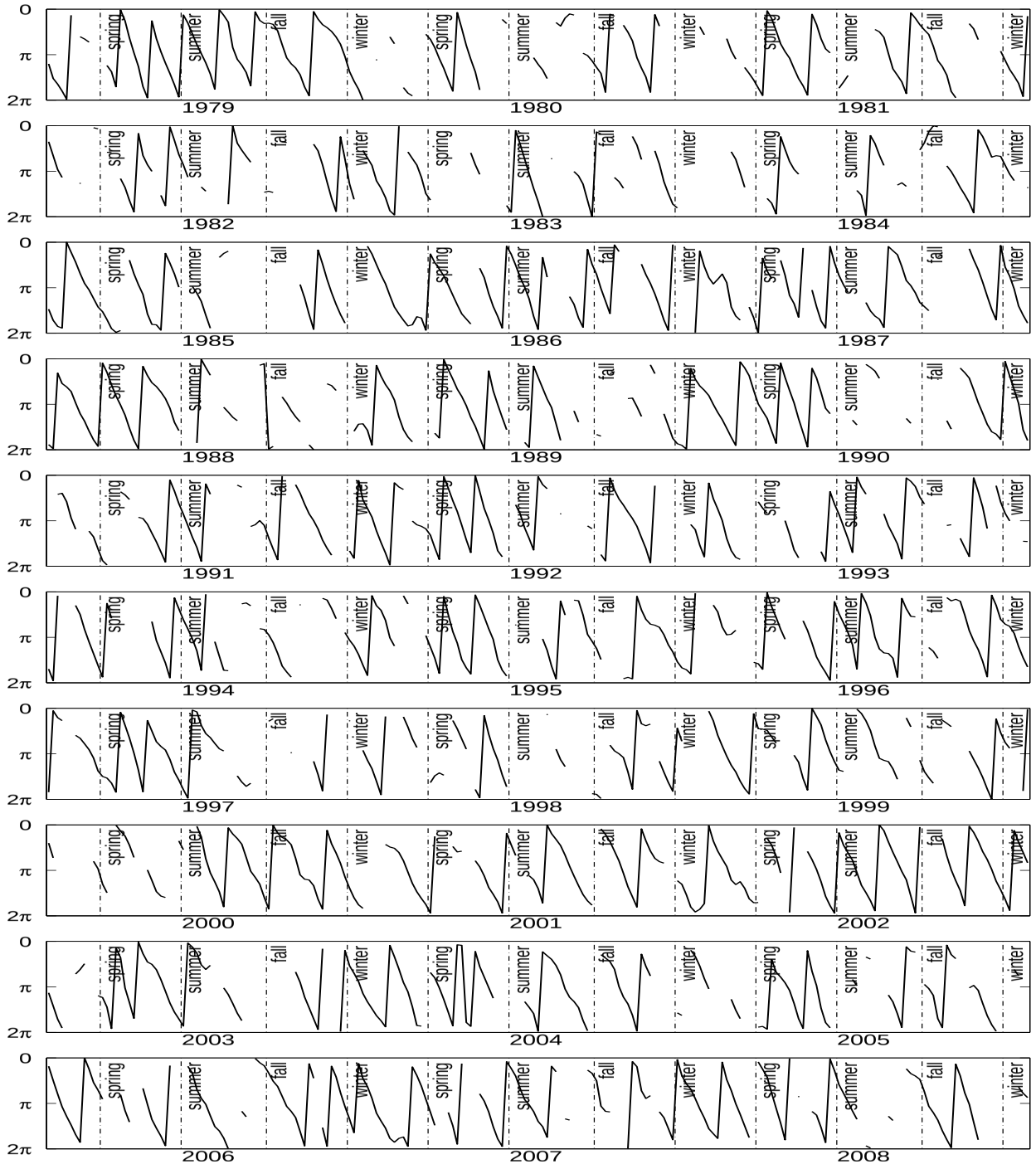

FIG. 3. Time series of the MJO phase θ. The zero phase $\theta = 0$ corresponds to LPC2 $= 0$ with LPC1 < 0. The black curve depicts the analyzed active MJO periods whereas the gaps denote quiescent periods which are not investigated in our study. The steady progression of θ with time demonstrates the temporal coherency of the active MJO.

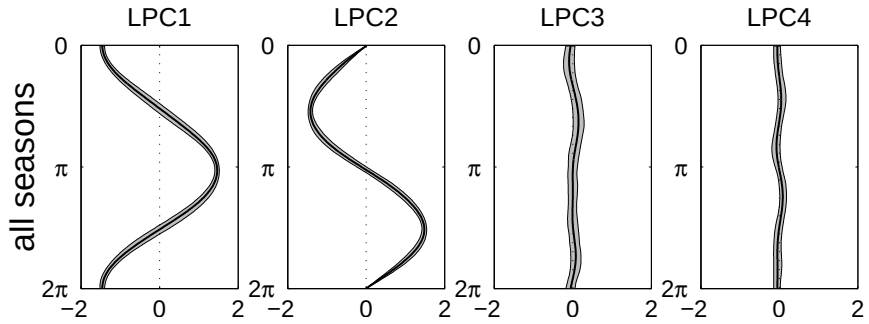

FIG. 4. All–year NLPCA.cir solution displayed with its output LPCs. The phase θ $(0 \ldots 2\pi)$ is aligned as for Fig. 3. All coefficients are standardized by the standard deviation of LPC1. The central lines stand for the coefficients of the full solution, while the gray–shaded envelopes indicate pointwise 90% confidence intervals.

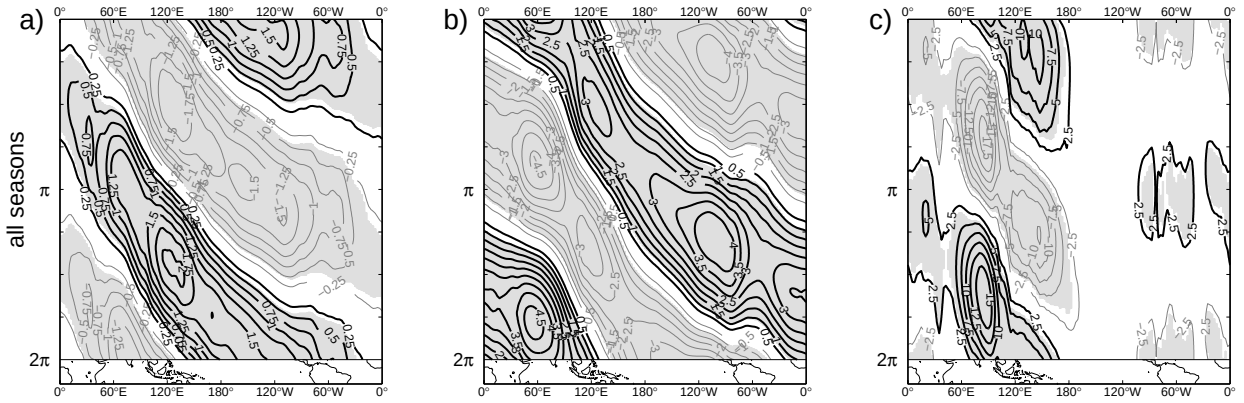

FIG. 5. Spatial structure of the NLPCA.cir solution computed for the whole year. The panels show the anomalies of a) U850 (ms^{-1}), b) U200 (ms^{-1}) and c) OLR (Wm^{-2}). The ordinates show the phase θ progressing downward. Positive contours are drawn in black whereas negative contours are drawn in dark gray. The zero contour is omitted. Gray–shaded areas indicate the MJO signal to be significant beyond the first contour level ($\pm 0.25 ms^{-1}$ for U850, $\pm 0.5 ms^{-1}$ for U200, $\pm 2.5 Wm^{-2}$ for OLR) at the 95% significance level (one–sided test).

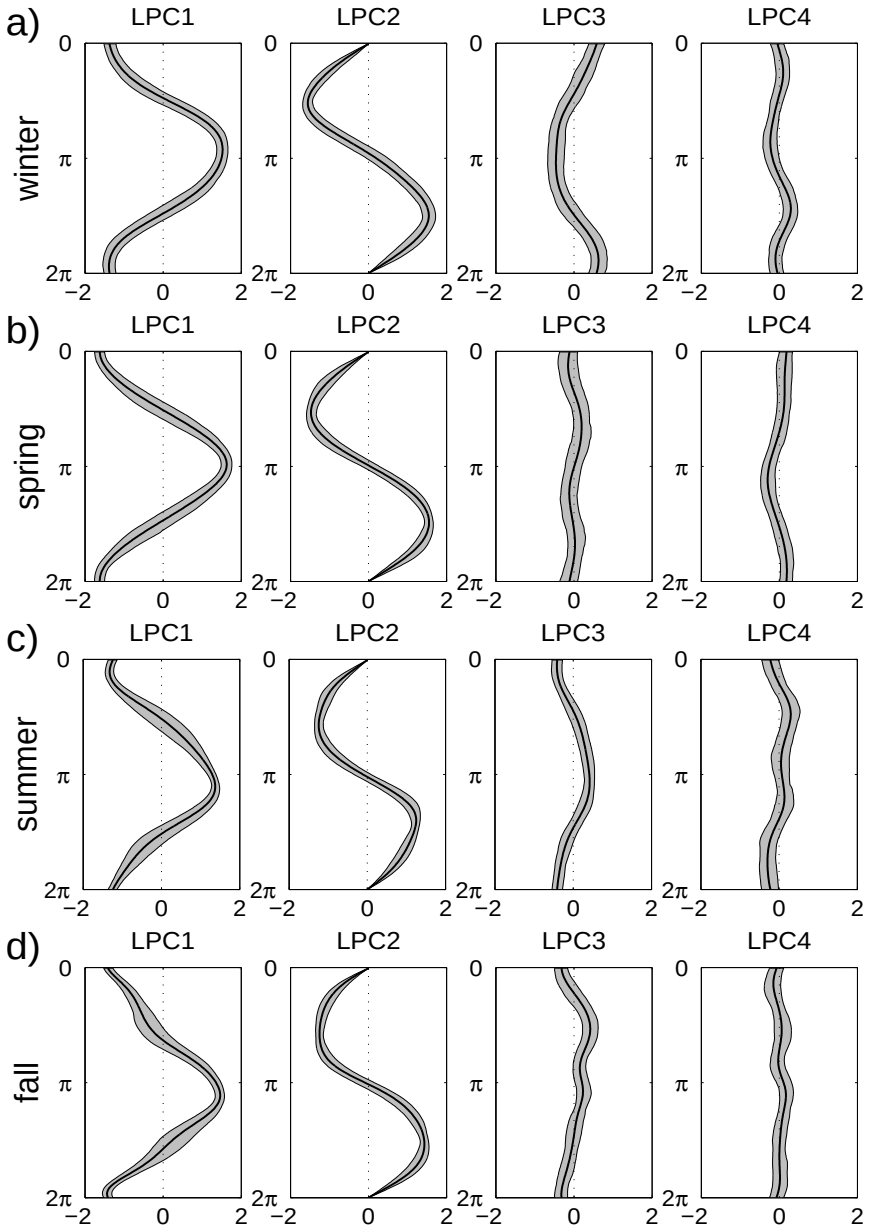

FIG. 6. Seasonal NLPCA.cir solutions displayed with their output LPCs for a) winter, b) spring, c) summer and d) fall. The phase θ ($0 \ldots 2\pi$) is aligned as for Fig. 3. All coefficients are standardized by the standard deviation of LPC1. The central lines stand for the coefficients of the full solutions, while the gray–shaded envelopes indicate pointwise 90% bootstrap confidence intervals.

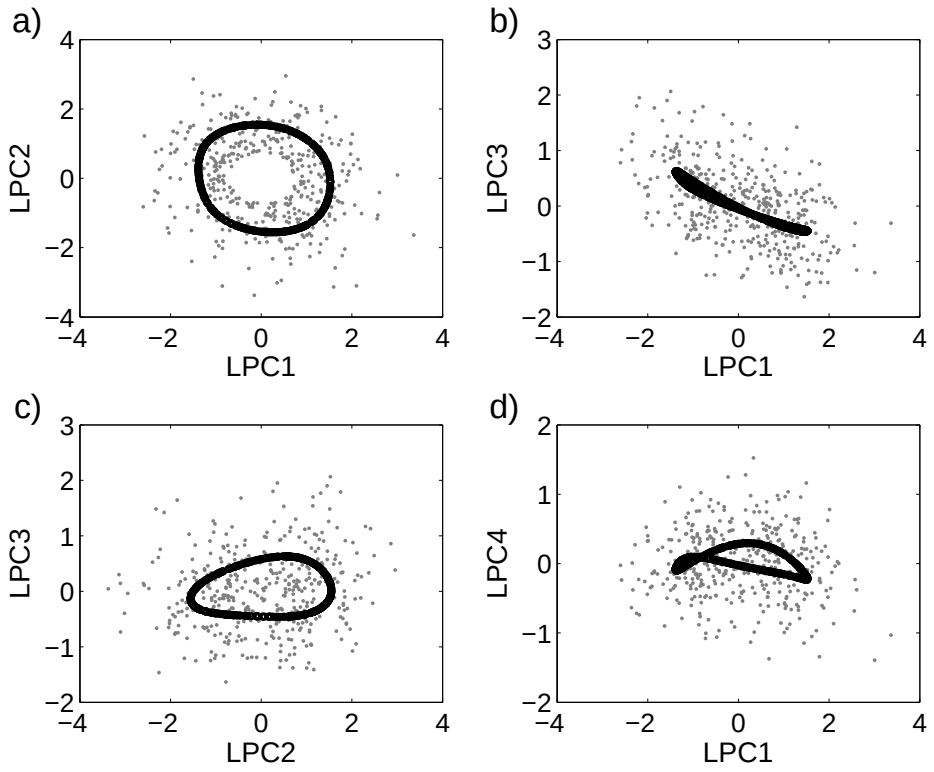

FIG. 7. NLPCA.cir solution (thick black line) displayed with the input data cloud (gray points) in varying planes. The graphs relate to boreal winter. The presented surfaces are spanned by a) LPC1 and LPC2, b) LPC1 and LPC3, c) LPC2 and LPC3 and d) LPC1 and LPC4.

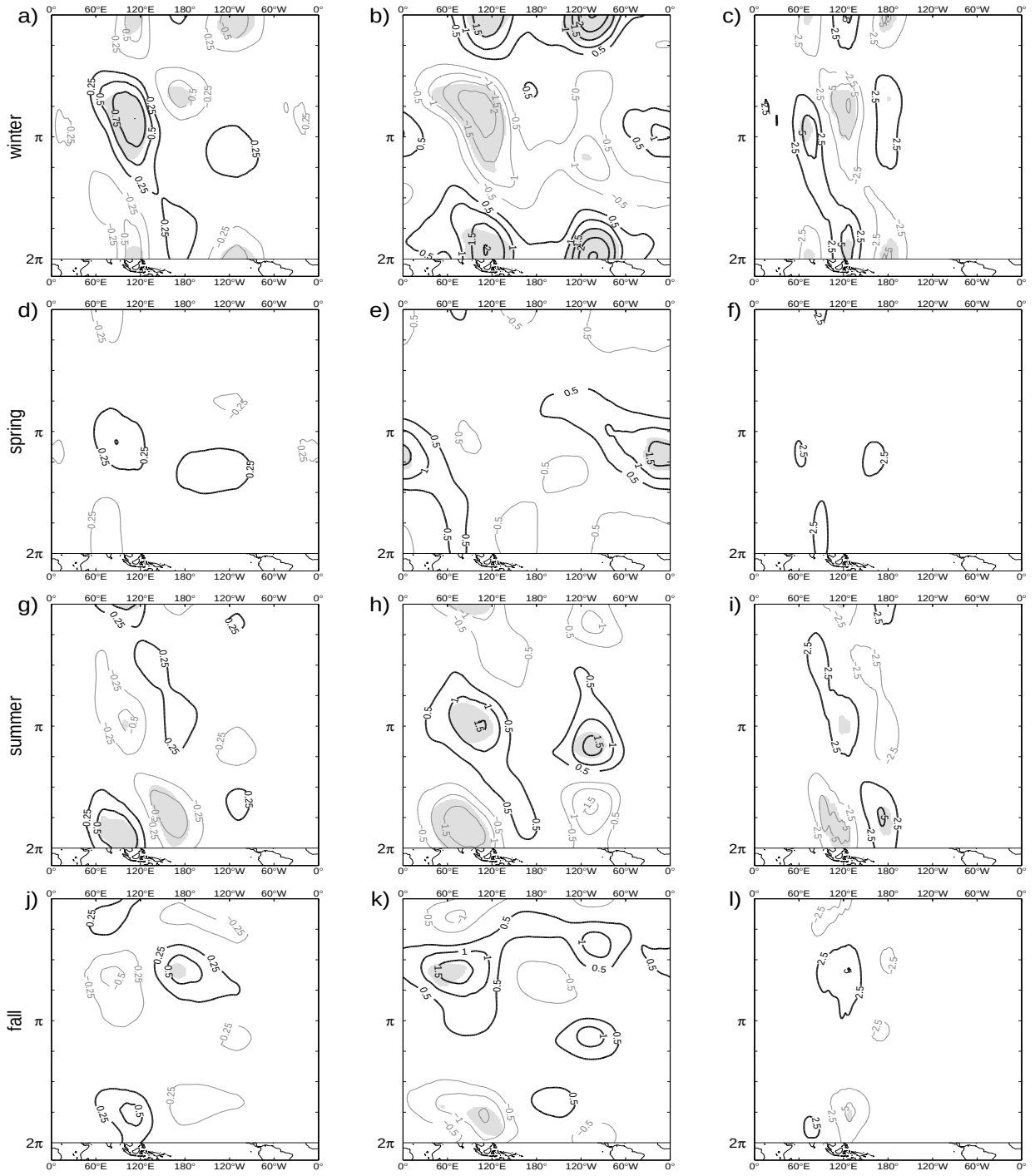

FIG. 8. Spatial variations of the NLPCA.cir solution computed for individual seasons. Contours depict the difference to the all–year solution presented in Fig. 5 (seasonal solution minus all–year solution). The panels a–c relate to winter, the panels d–f relate to spring, the panels g–i relate to summer and the panels j–l relate to fall. In each row, the first panel shows U850 (ms^{-1}), the second panel shows U200 (ms^{-1}) and the third panel shows OLR (Wm^{-2}). The ordinates show the MJO phase θ progressing downward. Positive (negative) contours are displayed in black (gray) and denote that the local value is higher in the seasonal (all–year) solution. Gray–shaded areas indicate differences beyond the first contour level ($\pm 0.25\mathrm{ms}^{-1}$ for U850, $\pm 0.5\mathrm{ms}^{-1}$ for U200, $\pm 2.5\mathrm{Wm}^{-2}$ for OLR) which are significant at the 95% level (one–sided test).

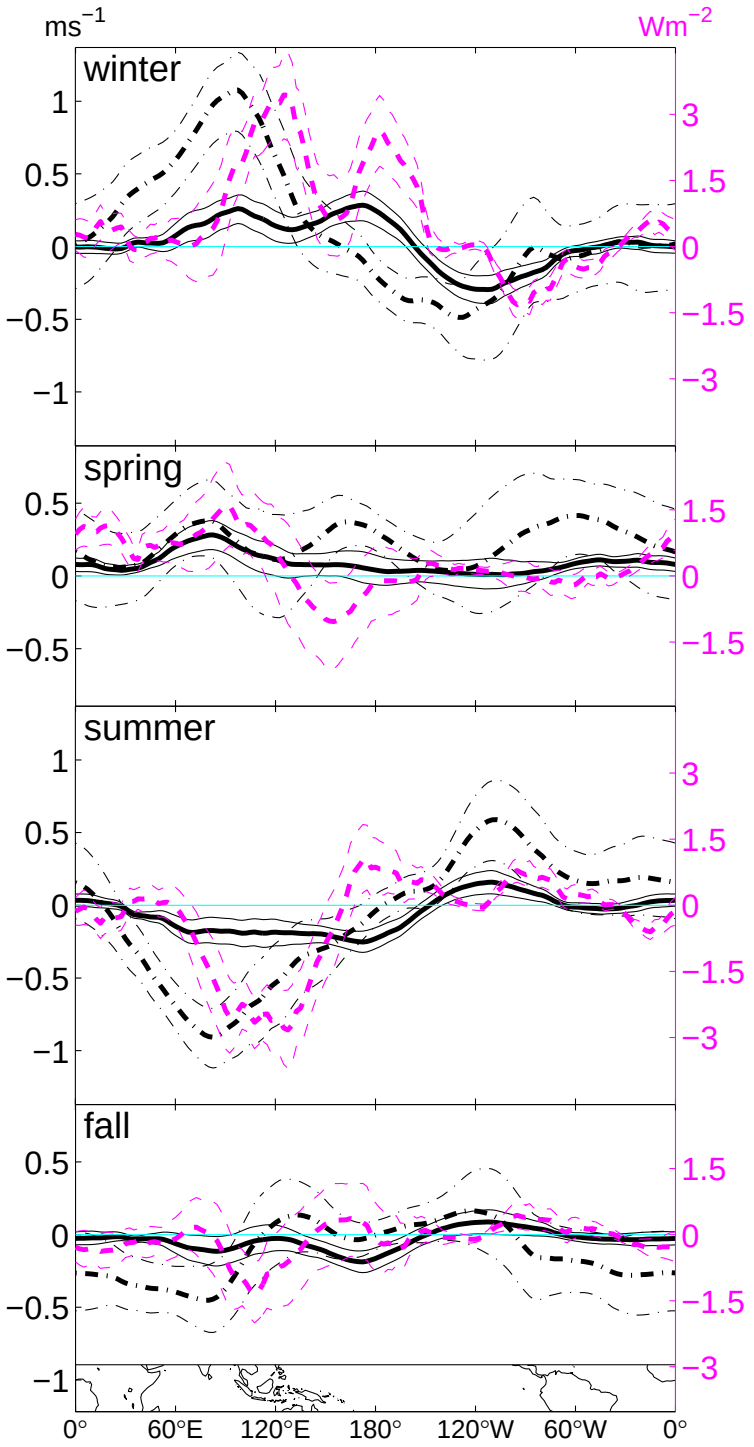

FIG. 9. Departures in local variance as measured by differences in the standard deviation (std) along the equator (std of the seasonal solution minus std of the all–year solution). The black lines depict U850 (solid) and U200 (dashed–dotted) with differences in ms^{-1} (left scale). The dashed magenta lines depict OLR with differences in Wm^{-2} (right scale). The thick central lines correspond to the full solutions whereas the thin upper and lower lines refer to 90% bootstrap confidence intervals.