



Preferred intra-seasonal circulation patterns of the Indian summer monsoon and active-break cycles

A new view of the active-break cycle

David M. Straus¹

Received: 19 May 2021 / Accepted: 7 November 2021 / Published online: 2 December 2021
© The Author(s) 2021

Abstract

Intra-Seasonal circulation regimes are identified from a cluster analysis of 5-day mean anomaly fields of 850 hPa horizontal winds from the ERA-Interim reanalysis for the boreal summer season (June–Sept. for 1979–2018) over the region (50°–100° E; 5° S–35° N). The k-means method was applied to the leading 6 principal components yielding k clusters. The degree of clustering is significant compared to synthetic data sets for any value of $k > 3$. The circulation is most likely to stay in the same cluster from one pentad to the next; significant transitions (with 95% confidence level) form a cycle. The similarity between the cycle depicted from 4 or 5 clusters and the active-break cycle, as well as the 45-day oscillation, is established by composites of 850 hPa winds, 200 hPa divergence, 500 hPa vorticity and vertical pressure velocity, precipitable water, diabatic heating and rainfall over India: Strong convection over the subtropical Indian Ocean moves to the central Bay of Bengal and central India, subsequently to the northern Bay of Bengal and west Bengal, and then further north into the Himalayas. We also find preferred transitions in which the convection moves equatorward from central India. The number of complete cycles found in 40 summers is 7 in the 4-cluster analysis. The number of times the system undergoes four (three) consecutive legs of the cycle is 16 (31). For 5 clusters only 3 complete cycles are found. sequences of five, four and three consecutive legs occur 10, 11 and 28 times, respectively.

Keywords Indian monsoon · Active and break cycles · Circulation regimes · Boreal summer intra-seasonal oscillation

1 Introduction

The atmospheric circulation associated with the Indian monsoon has a very rich structure on intra-seasonal time scales. Fluctuations in the wind field over the broad Indian region strongly modulate the occurrence of synoptic systems and the resulting rainfall, so that prolonged episodes (several days to one week) of strong rainfall (and of no rainfall) are observed over central India and surrounding regions (Murakami et al. 1984; Annamalai and Slingo 2001; Rajeevan et al. 2010) Since this well-known “active-break cycle” has a large impact on agriculture (Prasanna 2020) the fidelity with which weather and climate models treat this cycle (and

the circulation in which it is embedded) is clearly of concern (Pattanaik et al. 2020). These active and break periods are often characterized by compositing fields based on their time lagged relationship with a single, regionally averaged index of extreme rainfall, as in Rajeevan et al. (2010).

This cycle is embedded in a complex of larger scale intra-seasonal fluctuations which extend over the broad Indian Ocean—western Pacific region. There is general agreement that two broad-band oscillations, the 30–60 day boreal summer intra-seasonal oscillation, or BSISO (Murakami et al. 1984; Lau and Chan 1986; Kemball-Cook and Wang 2001; Annamalai and Sperber 2005; Hazra and Krishnamurthy 2014) and the 10–20 day (quasi bi-weekly, or QBW) oscillation (Krishnamurthy and Bhalme 1976; Krishnamurthy and Arduinay 1980; Chatterjee and Goswami 2004) dominate this variability. In order to identify these oscillations and determine their character, a number of complex spatio-temporal spectral analysis techniques have been used, including

✉ David M. Straus
dstrauss@gmu.edu

¹ AOES Department, Center for Ocean-Land-Atmosphere Studies, George Mason University, Fairfax, VA 22030, USA

extended empirical orthogonal functions (Suhas et al. 2013; Sahai et al. 2015), multi-channel singular spectrum analysis (Krishnamurthy and Shukla 2000, 2007, 2008), principal oscillation patterns (Annamalai and Slingo 2001) and cyclo-stationary empirical orthogonal functions (Annamalai and Sperber 2005).

Comparisons of the BSISO in simulations (Neena et al. 2017) and forecasts (Lee et al. 2015) with those from reanalyses have been carried out in the EEOF framework, although this process is somewhat cumbersome.

In contrast to the local index-based approach used to diagnose the active-break cycle, and the complex mode techniques used to isolate the BSISO and QBW, this article analyzes the large-scale variability of the circulation field, focusing on circulation patterns (“regimes”) which are by some measure preferred states of the system. This approach can be implemented via the general technique of cluster analysis, which while very often used in studies of mid-latitude intra-seasonal variability (see Straus et al. 2017), has not been widely applied to the Indian monsoon circulation. Chu et al. (2017) use self-organizing maps (a form of cluster analysis) to characterize the intra-seasonal variability of outgoing long-wave radiation and low-level zonal winds in the very broad region extending from East Africa to the western Pacific, and are able to identify both stationary and propagating patterns. The purpose of this article is to explore the structure of preferred circulation patterns over the Indian region (50°–100° E, 0°–35° N) and to relate transitions between them to the active-break cycle over India. This is meant to provide a framework in which the model forecasts (and simulations) can be evaluated with regard to the both the circulation regimes and, via their transitions, to the active-break cycle. Recent interest in the fidelity of predictions and simulations of the active-break cycle is evidenced in the work of Neena et al. (2017) and Pattanaik et al. (2020).

Cluster analysis partitions every state of the atmosphere (whether daily or time-filtered anomalies) between a small number of clusters on the basis of a common structure. The average of all states in a cluster defines a *centroid* (also called a circulation regime). The complex evolution of the circulation is “coarse-grained” and so is described by the residence time within a regime and the probability of transition into another regime. Active-break cycles within this approach emerge as preferred transition paths between regimes. A fuller picture of the cycle and its components is available by compositing a number of variables over states assigned to each regime.

In Sect. 2, the data set and the analysis methods are described. Methods of estimating statistical significance are described in Sect. 3, which can be skipped by the reader without loss of continuity. Section 4 gives a number of composite fields over the broad South Asian region

for the characteristic states (centroids) from each cluster for a 4-cluster analysis and a 5-cluster analysis, while Sect. 5 describes the composite diabatic heating over the global tropical belt. The frequency of occurrence and lifetime of the regimes, as well as evidence for preferred transition paths between them, are given in Sect. 6. A Discussion is given in Sect. 7, and the Conclusions in Sect. 8.

2 Data sets and analysis methods

2.1 Data

Once-daily fields of the horizontal wind (u , v) at 200, 500 and 850 hPa and the vertical pressure velocity (ω) at 500 hPa were obtained from the ERA-Interim reanalysis (Dee et al. 2011) for the 120-day period starting 01 June for the 40 years 1979 through 2018. These data were interpolated from the original (512 × 256) (lon × lat) Gaussian grid to a (128 × 64) (lon × lat) Gaussian grid. 24 consecutive five-day means (pentads) of the winds, divergence (D), vorticity (ζ) and vertical pressure velocity (ω) were formed for use in the cluster analysis and composites.

Additionally, four-times daily data of the horizontal winds (u , v), temperature T , and pressure vertical velocity ω were obtained from the ERA5 reanalysis (Hersbach et al. 2020) at 37 pressure levels for the purpose of estimating the diabatic heating. These fields were interpolated from the original 0.25° grid to a (256 × 128) (lon × lat) Gaussian grid prior to calculating the heating. The diabatic heating, whose calculation is described in Sect. 2.3 was vertically integrated into nine layers, and pentads formed.

Daily Indian precipitation data for June through September of the years 1979–2018 were obtained from the India Meteorological Department (Pai et al. 2014). This dataset, which has been updated regularly, covers only the Indian land region on a 0.25° (lon × lat) grid. The daily data were averaged into pentads, as described above.

For all the fields described above, the seasonal cycle was computed by projecting the seasonal time series of pentads (for each grid point and year) onto the first three Legendre polynomials, equivalent to a parabola in time, as in Straus (1983). The polynomial coefficients were then averaged over all 40 years, and a single smooth seasonal cycle reconstructed. Anomalies were obtained by removing this seasonal cycle.

2.2 Cluster analysis

The goal of the cluster analysis is to describe the slowly varying, large-scale circulation field over the Indian region in terms of a relatively few characteristic patterns. Although it is possible to apply a cluster analysis to the raw pentad

fields of (u, v) directly, it is convenient to pre-filter the data using a principal component analysis, which provides a set of patterns (the empirical orthogonal functions or EOFs), and their time coefficients (the principal components, or PCs). The PCs provide a new representation of the data set in a reduced-dimensional coordinate space. It is in this space that the cluster analysis is carried out.

We apply a principal component analysis to the vector field of (u, v) at 850 hPa over the wide Indian region (50.63°–98.44° E, 1.40 S°–34.88° N). The leading 6 (12) modes explain 65% (79%) of the total space-time variance. Since the variance explained drops off rapidly with mode number, we retain only 6 modes in the cluster calculations to be described, although we have verified that the composite circulation and rainfall for each cluster are nearly unchanged when 12 modes are retained.

The particular method of cluster analysis used is the k -means method (Deday and Simon 1976; Desbois et al. 1982; Michelangeli et al. 1995; Straus et al. 2017), which takes as input the leading principal component time series, and seeks to approximately identify regions in this phase space in which the local density is higher than would be expected based on a purely multi-normal distribution. Given a number k of clusters to be sought, the method assigns each observed state (pentad) uniquely to one of the k clusters, each consisting of points “bunched” together in the phase space. Given these assignments, the average PC coordinates of all states in a given cluster (the so-called “centroid”) define a single set of PC coefficients which corresponds (using their respective EOF patterns) to a unique pattern, termed a circulation regime. The criterion that determines how the states are assigned is a global one, based on a single parameter R , called the variance ratio, which is defined as the ratio of the average sum of squares of the centroid coordinates (weighted by the cluster population) to the average intra-cluster variance. This ratio is used both to monitor the convergence of the algorithm, and in the assessment of its significance. Technical details of the k -means clustering algorithm and more interpretation of the results can be found in Straus et al. (2017).

While each individual circulation regime consist of a linear combination of the leading EOFs, the regimes are not constrained to be symmetric in sign: a regime pattern often does not have the opposite pattern represented in the set of k regimes. This is in distinct contrast to the output of principal component analysis, which describes modes which are equally likely to have either sign.

It should be pointed out that there is no unique method of choosing what k , the number of regimes, should be. Retaining too many regimes generally leads to very poor reproducibility to sub-sampling, while retaining too few regimes may obscure important distinctions in local behavior. In this paper we report results for both $k = 4$ and $k = 5$ regimes.

The cluster analysis yields an assignment of each 5-day period to one of k regimes. Composites of the full (not truncated) fields of anomalies of low-level (850 hPa) circulation u, v, ω, D, ζ and rainfall over each regime are examined. The frequency of transitions between regimes, and the possible sequences of regimes, together give a coarse-grained approximation to trajectories in phase space.

2.3 Diabatic heating

The diabatic heating rate Q is estimated from the four-times daily ERA5 data at 37 pressure levels as a residual in the thermodynamic equation. The heating is evaluated on the (256×128) (lon \times lat) Gaussian grid, using a residual method as reported in Swenson and Straus (2021): The heating rate Q is obtained from:

$$c_p \Pi \left(\frac{\partial \theta}{\partial t} + \nabla \cdot (\mathbf{v} \theta) - \theta (\nabla \cdot \mathbf{v}) + \omega \frac{\partial \theta}{\partial p} \right) = Q \quad (1)$$

Here θ is potential temperature, p is pressure, $\omega = \frac{dp}{dt}$, c_p the specific heat at constant pressure per unit mass, \mathbf{v} the horizontal velocity field, and $\Pi = \left(\frac{p_0}{p} \right)^\kappa$, with $p_0 = 1000$ hPa and $\kappa = \frac{R}{c_p}$, R being the gas constant. Q is transformed from the grid representation to spherical harmonics, truncated to a triangular (T159) representation, and transformed back to the grid.

3 Significance of results

The assessment of significance of the cluster analysis is a global one, applying to the entire assignment of clusters, and not the local density in the PC space. The idea is that if the time series of one the PCs of the observed states (assumed to have Gaussian statistics) is statistically independent of the time series of all other PCs, then any maximum in probability density not at the origin (identifying a cluster centroid) can only occur due to sampling error: in the limit of a large sample, the joint probability density of all the PCs should be a multi-normal distribution (no clustering). Based on this idea, surrogate stochastic Gaussian time series are constructed for each individual PC, with the same length as the observed time series. The surrogate time series are constructed in such a way that the time lag-correlation statistics of each observed PC are preserved, but the time series of different PCs are independent (Christiansen 2007). When the clustering algorithm is applied to the surrogate time series, it will always give a set of cluster assignments (for a given k) and a variance ratio R , but the apparent clustering can only be due to the finite size of the sample. Repeating the application of the cluster analysis on 100 independently generated

sets of surrogate time series, we determine how many times the value of R exceeds the observed value. The corresponding fraction α yields the p -value, so that if $\alpha = 0.05$ then we are 95% confident that the clustering of the observed time series is not due only to sampling error.

We have carried out this analysis for a range of k , from 2 to 6, and find that for $k = 4, 5$, or 6 the confidence level is 99%, indicating very little chance that the degree of clustering in the original data is due to sampling error. Common to many methods of clustering, the final choice for the value of k is to some extent arbitrary (Christiansen 2007). We used both $k = 4$ and 5 in our analysis, but focused on the choice of 5 clusters, as explained in Sect. 4. We should note that for highly skewed data, Christiansen (2007) has shown that k -means algorithm may incorrectly identify multiple clusters. In our case this is not a concern: the leading two PCs have small skewness (-0.143 and 0.125).

In order to test the robustness of the cluster centroids to sub-sampling, we repeated the cluster analysis using 1000 randomly chosen half-length data sets. For each such half-length dataset, the variance ratio R was computed and the resulting histogram formed into a probability distribution (not shown). The distribution is sharply peaked at a value near 1.01 for $k = 4$ and $k = 5$, and at a slightly higher value for $k = 6$. Another measure of the agreement between the clusters in the half-length data sets and those in the full data set is the pattern correlation between the cluster centroids. For each half-length data set and value of k , each cluster centroid was matched to one of the k centroids in the full data set based on pattern correlation. The k values of correlation were then averaged (using the Fisher transformation) to assign a single value to that half-length data set. In all cases almost the entire distribution lies above correlation values of 0.8, with a small tail below this for $k = 4$ and $k = 6$.

Preferred transition paths between the clusters are assessed with the use of the transition matrix, whose i th row and j th column gives the number of times the system makes a transition between cluster i and cluster j . The off-diagonal elements define transitions, whose statistical significance is determined by comparison to the results of a large number (here 1000) of scrambled data sets. In each scrambled data set the cluster assignments are resampled using a bootstrap with replacement technique: the cluster assignment of each state (pentad p of year x) is replaced with that of a randomly chosen pentad q and year y . If a block of n pentads $q + 0, q + 1, \dots, q + n - 1$ of year y are all assigned to the same cluster, then the pentads associated with the entire block of n days are assigned to pentads $p, p + 1, \dots, p + n - 1$ of year x .

The significance of the composite anomalies of the vorticity, divergence, vertical velocity and diabatic heating were assessed with a two-sided t-test. Significance is indicated for all anomalies exceeding 95% confidence.

4 Circulation regimes

Composites of a number of anomaly fields are presented here for the cases of $k = 4$ and $k = 5$. The 850 hPa horizontal wind field (u, v), shown using the full anomalies (not truncated in PC-space), is helpful in indicating the location of anomalies in the Monsoon trough. This is also indicated by the 500 hPa vorticity ζ . Both the 200 hPa divergence D and the 500 hPa vertical pressure velocity ω indicate regions of anomalous upward motion associated with convection. Finally, we show the rainfall anomalies (available only over the Indian land region).

4.1 Four regimes

Composites for $k = 4$ are shown in Figs. 1, 2, and 3. The composite figures are ordered (from (a) through (d)) to correspond to a preferred cyclical transition path $(a) \rightarrow (b) \rightarrow (c) \rightarrow (d) \rightarrow (a)$ which will be discussed in Sect. 6.

In Fig. 1, Cluster (a) shows low-level cyclonic anomalies over the equatorial Indian Ocean associated with enhanced upper-level divergence (indicative of deep convection), and a suppression of the Monsoon trough over India itself (with Easterly anomalies) associated with a negative divergence anomaly in the Bay of Bengal. The mid-level ζ and ω anomalies for the same cluster, shown in Fig. 2a, indicate enhanced vorticity and upward motion over the equatorial Indian Ocean, with suppressed vorticity and downward motion over India and the Bay of Bengal. Consistent with this, the corresponding rainfall anomalies over India, shown in Fig. 3 show suppressed rainfall over the Western Ghats and over the Indo-Gangetic plain.

Cluster (b) presents a sharp contrast: the Monsoon trough is now enhanced, with positive upper-level divergence anomalies over the Bay of Bengal and central India (Fig. 1b), associated with enhanced mid-level vorticity and upward motion (Fig. 2b) and strongly enhanced rainfall over the Western Ghats and over central India (Fig. 3b). The enhanced rainfall over the west coast is consistent with the enhanced on-shore low-level winds, seen in Fig. 3.

In cluster (c), the enhanced cyclonic low-level anomaly (and the accompanying upper-level divergence) are now located further to the northeast compared to their positions in cluster (b), and are also weaker (compare Fig. 1b and c). Enhanced mid-level vorticity (along with rising motion) is seen over the northern Bay of Bengal, with a large area of sinking motion and decreased vorticity further to the southwest (Fig. 2c). As expected from the circulation anomalies, the rainfall is modestly enhanced over northeastern India but is suppressed over central India (Fig. 3c).

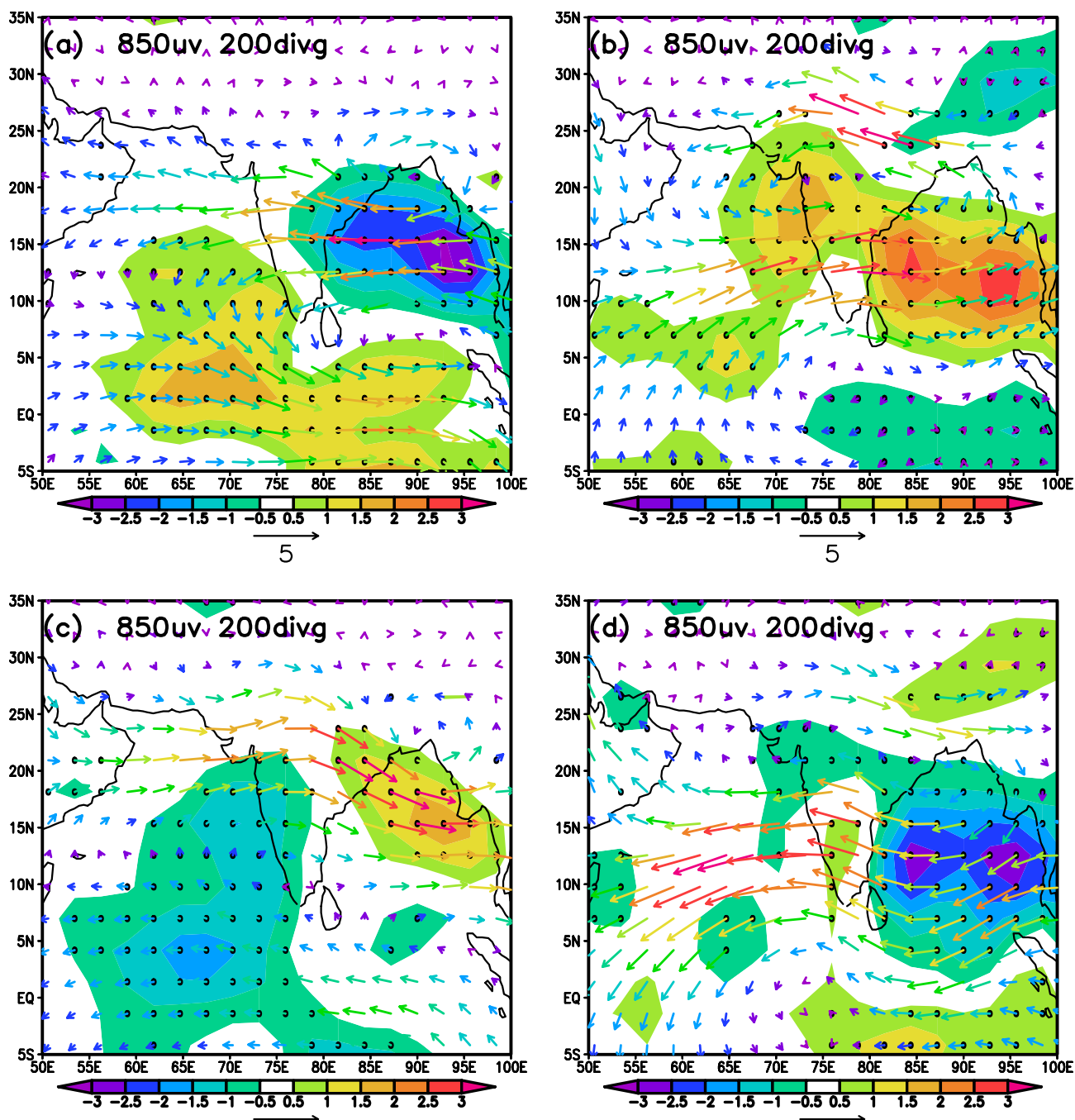


Fig. 1 Composite anomalies of horizontal winds (u, v) at 850 hPa for each of 4 clusters ($k = 4$), shown in vectors. Composite anomalies of divergence at 200 hPa shown in shading, in units of 10^{-6} s^{-1} . Sig-

nificance of the divergence at the 95% level is indicated by the stippling. The clusters are labeled (a), (b), (c), and (d) corresponding to the panels

Finally, in cluster (d) the enhanced low-level cyclonic anomaly and associated upper level divergence (and rising motion) have moved further northward to the eastern Himalayan mountain range (see Figs. 1d and 2d), while anomalous sinking motion and negative vorticity are seen over the Bay of Bengal and across much of India into the Arabian Sea. Over the tropical Indian Ocean and southern

India, increased convection, with upper level divergence and rising motion are apparent in Figs. 1d and 2d. The rainfall anomalies (Fig. 3d) are consistent with the circulation, with suppressed rainfall over much of central India, but enhanced rainfall in the Himalayas and over southern India.

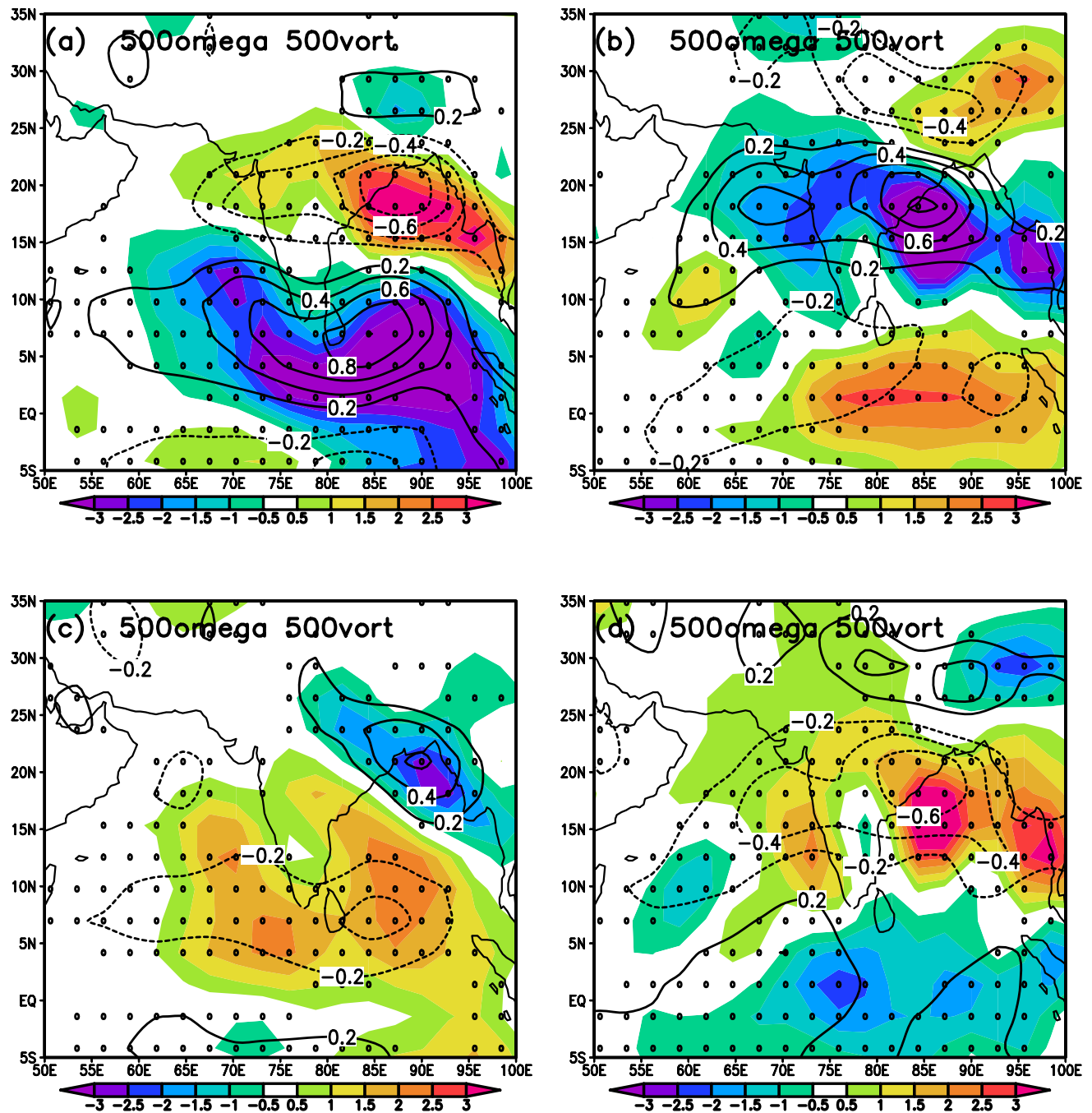


Fig. 2 Composite anomalies of vorticity ζ at 500 hPa for each of 4 clusters ($k = 4$), shown in contours in units of 10^{-5} s^{-1} . Composite anomalies of vertical pressure velocity (ω) at 500 hPa shown in shad-

ing, in units of $1.0^{-2} \text{ Pa s}^{-1}$. Significance of the vorticity at the 95% level is indicated by the stippling. The clusters are labeled (a), (b), (c), and (d) corresponding to the panels

4.2 Five regimes

Composites for $k = 5$ are shown in Figs. 4, 5, and 6. The composite figures are ordered (from (a) through (e)) to correspond to a preferred cyclical transition path (a) \rightarrow (b) \rightarrow (c) \rightarrow (d) \rightarrow (e) \rightarrow (a) which will be discussed in Sect. 6.

The low-level circulation and upper-level divergence anomalies for cluster (a) shown in Fig. 4a are very similar to cluster (a) for $k = 4$ (compare Fig. 1a), although the center of upper-level divergence over the tropics is displaced southwards slightly in Fig. 4a, while a small area of positive divergence is seen over the eastern Himalayas. These

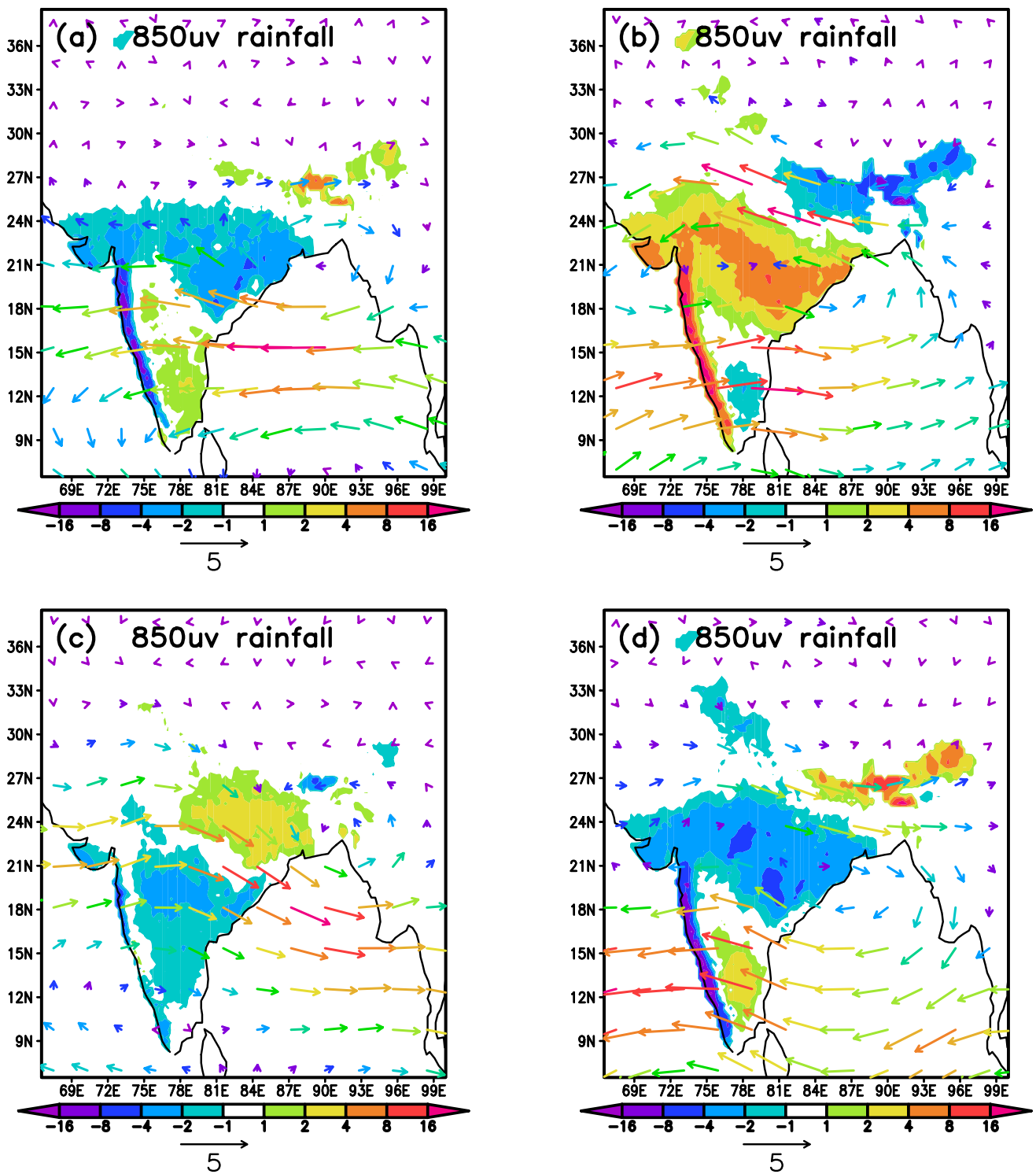
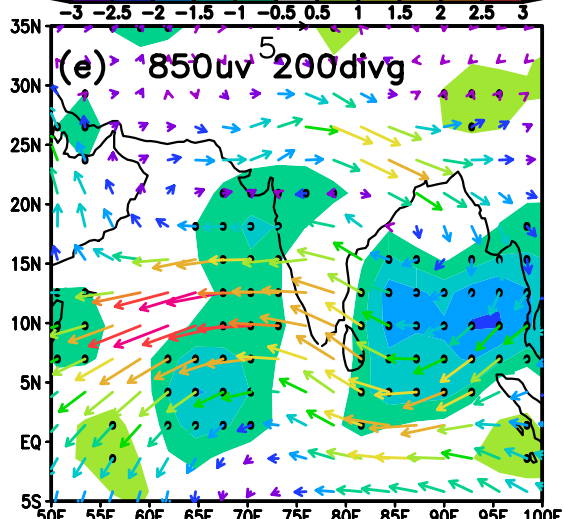
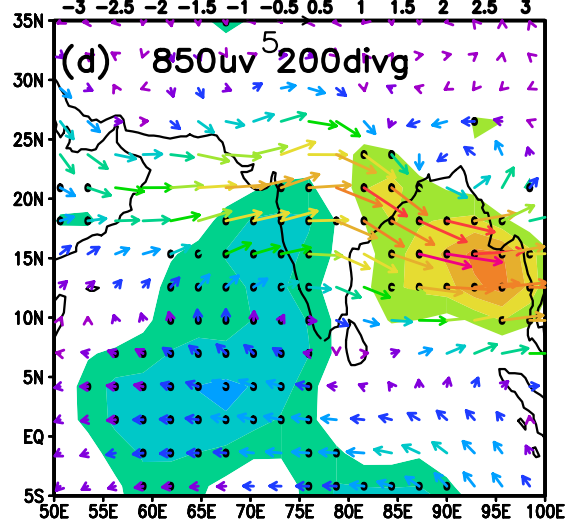
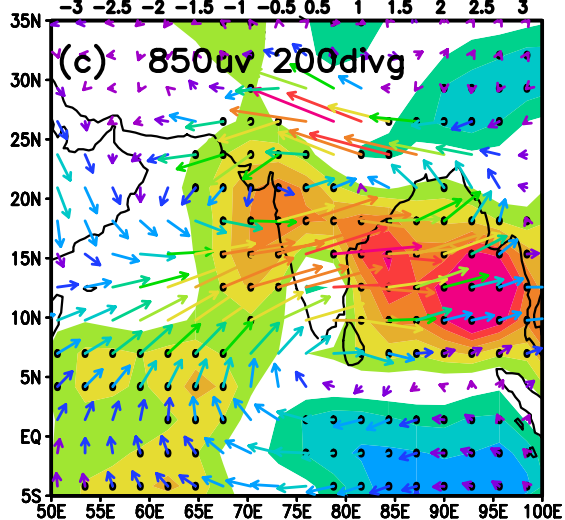
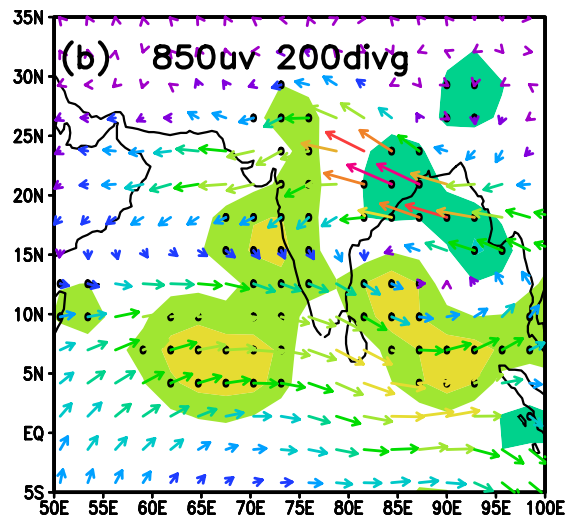
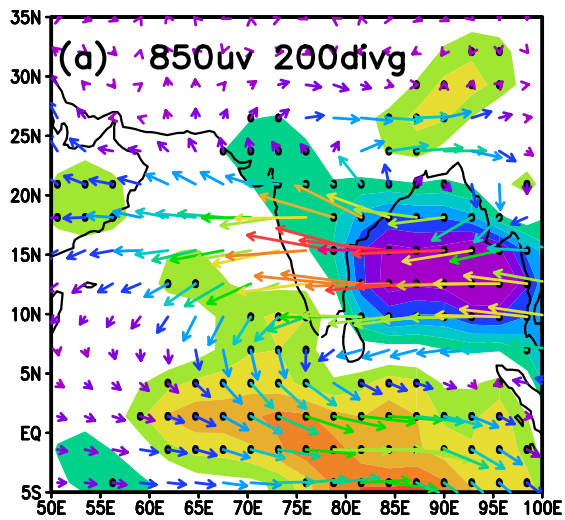


Fig. 3 Composite anomalies of horizontal winds (u, v) at 850 hPa for each of 4 clusters ($k = 4$), shown in vectors. Composite anomalies of rainfall shown in shading, in units of 10^{-3} md^{-1} . Only values of the

precipitation significant at the 95% level are plotted. The clusters are labeled (a), (b), (c), and (d) corresponding to the panels



5

5

Fig. 4 Composite anomalies of horizontal winds (u , v) at 850 hPa for each of 5 clusters ($k = 5$), shown in vectors. Composite anomalies of divergence at 200 hPa shown in shading, in units of 10^{-6}s^{-1} . Significance of the divergence at the 95% level is indicated by the stippling. The clusters are labeled (a), (b), (c), (d) and (e) corresponding to the panels

observations also hold for the regions of enhanced mid-level vorticity and ω (compare Figs. 2a and 5a).

Cluster (b) for $k = 5$, showing enhanced upper-level divergence, cyclonic anomalies and rising motion in both the Bay of Bengal and the Arabian Sea, with modest subsidence further to the northeast (Figs. 4b and 5b) has no obvious counterpart in the results for $k = 4$. This cluster appears as a (new) transition state between cluster (a) and cluster (c). The latter (Figs. 4c and 5c) shows strong divergence, upward motion and enhanced mid-level vorticity over the Bay of Bengal and across central India into the Arabian Sea, where this feature extends southwestward into the tropics. Regions of subsidence are seen in the tropics (east of 65° E) and further north over the eastern Himalayas. Note that this cluster is very similar to clusters (b) for $k = 4$ seen in Figs. 1b and 2b. Similarly, cluster (d) for $k = 5$ (Figs. 4d and 5d), with positive upper-level divergence, rising motion and positive mid-level vorticity in the northern Bay of Bengal and the opposite circulation over a broad band south of 15° N, is very similar to cluster (c) for $k = 4$ (Figs. 1c and 2c). Finally, cluster (e) for $k = 5$ (Figs. 4e and 5e) shows suppression of convection (downward motion) over a band extending from Myanmar westward over central India, and into the Arabia Sea, with some rising motion to the south of this, and over the extreme northeastern portion of the domain. The patterns are similar to those of cluster (d) for $k = 4$ (Figs. 1d and 2d); however there is a clear northward shift in the $k = 4$ cluster (d) compared to the $k = 5$ cluster (e).

The cluster (a) rainfall composite for $k = 5$ shown in Fig. 6a indicates suppressed rainfall over central India and the western Ghats with enhanced rainfall over the Himalayas and south central India, in rough agreement with cluster (a) for $k = 4$ (Fig. 3a), although the positive anomalies in the $k = 5$ cluster are clearly stronger. Cluster (b) (Fig. 6b) shows a weak enhancement of rainfall over central India and only a small remnant of the dryness which has moved northeastward (compared to cluster (a)). In cluster (c), strong positive anomalies of rainfall appear over central India and the western Ghats, with a suppression seen over the eastern Himalayas (Fig. 6c), very similar to cluster (b) for $k = 4$ (Fig. 3b). The positive rainfall anomalies weaken and move northeastward in cluster (d) (Fig. 6d), matching cluster (c) for $k = 4$ (Fig. 3c). At the end of the sequence, cluster e (Fig. 6e) shows widespread suppression of rainfall over both central India and the western Ghats, and a small area of

enhanced rainfall over the eastern Himalayas, a pattern very similar to that seen for $k = 4$ in cluster (d) (Fig. 3d).

5 Diabatic heating

Composites of the anomalies of diabatic heating (vertically integrated from 1000 to 50 hPa) are given for the four clusters ($k = 4$) over the global tropical belt in Fig. 7. Again we shall discuss the clusters in the framework of a cyclical transition path (a) \rightarrow (b) \rightarrow (c) \rightarrow (d) \rightarrow (a).

In cluster (a), prominent tropical Indian Ocean heating anomalies to the south of the equator are connected to a northwest-to-southeast (“tilted”) band of heating extending from the Arabian Sea, across southern India into the southern Bay of Bengal and further east over the Maritime continent. A parallel, similarly oriented band of cooling lies about 15 degrees further to the north. Heating anomalies are also seen further north over the Himalayas. The tilted band of heating moves northward in cluster (b) so that it covers central and northern India, and extends southeastwards into the tropical western Pacific. The Himalayan area of heating in cluster (a) is replaced by cooling in cluster (b), as is the tropical Indian Ocean heating. The tilted heating bands moves further northward in cluster (c), reaching the Himalayas in cluster (d), in which heating is again seen in the tropical Indian Ocean. Generally bands of heating are separated by bands of cooling. Of interest is the alternating areas of cooling (in clusters (a) and (b)) and heating (clusters (c) and (d)) over the eastern tropical Pacific.

Heating anomalies for five clusters are shown in Fig. 8, labeled (a) through (e) as in Sect. 4.2. Cluster (a) is very similar to cluster (a) for the 4 cluster case (Fig. 7), while cluster (d) and (e) for 5 clusters are very similar to clusters (c) and (d) for 4 clusters, respectively. It is clusters (b) and (c) in the 5-cluster case which are merged into cluster (b) in the 4-cluster case (compare Figs. 7b and 8b). The far eastern Pacific cooling noted in the 4-cluster case is seen in clusters (a) and (b) in Fig. 8, whereas the previously noted eastern Pacific heating is seen in clusters (d) and (e).

6 Regime lifetimes and transition paths

6.1 Regime lifetimes

The average lifetime in a cluster (regime) is given for each of the $k = 4$ clusters in Table 1 and for each of the $k = 5$ clusters in Table 2, in both pentads and the corresponding number of days. Also given is the total time the system resides in each cluster. The average regime lifetime for $k = 4$ is 8.6 days, while that for $k = 5$ is 7.8 days. For $k = 4$ the system spends 58% of time in regimes (b) and (c), the regimes which show

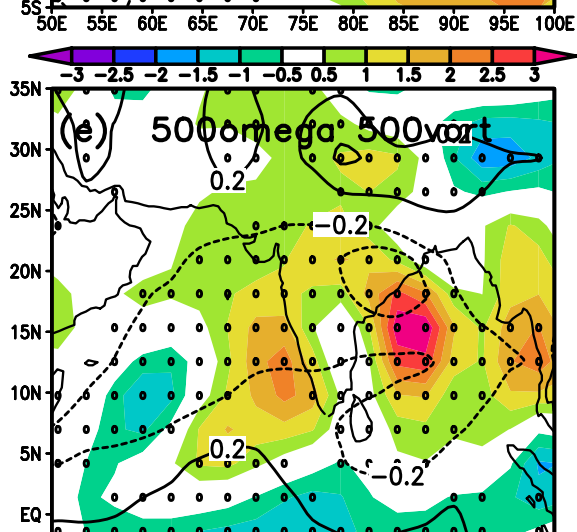
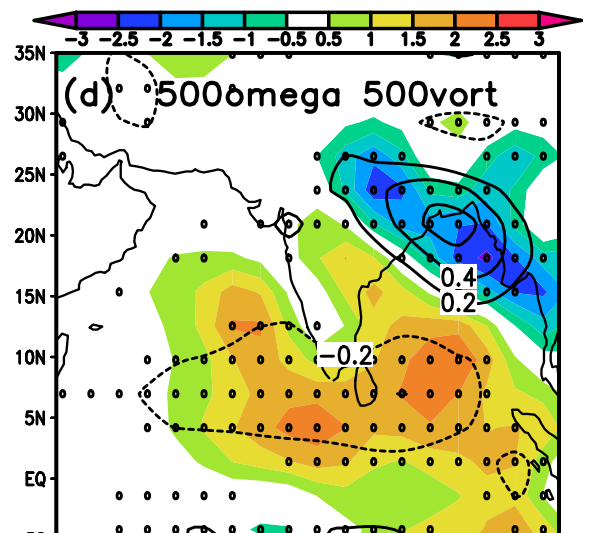
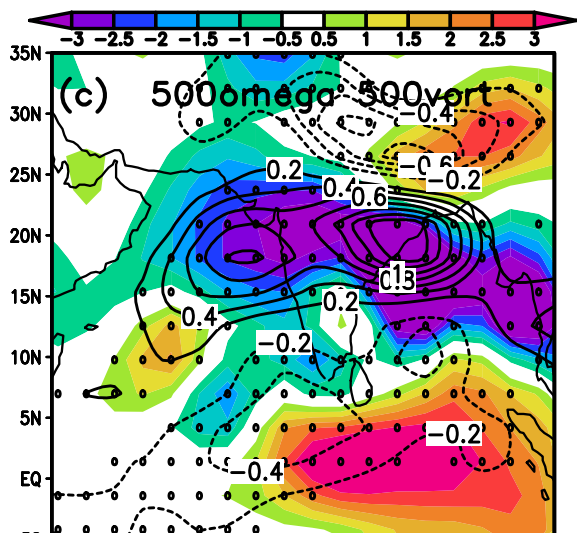
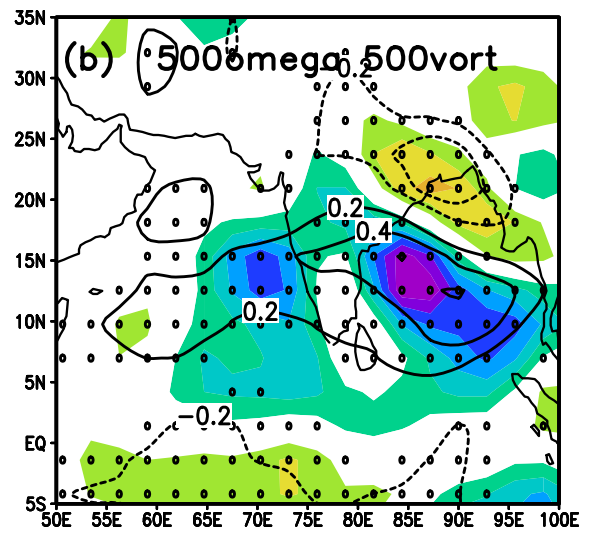
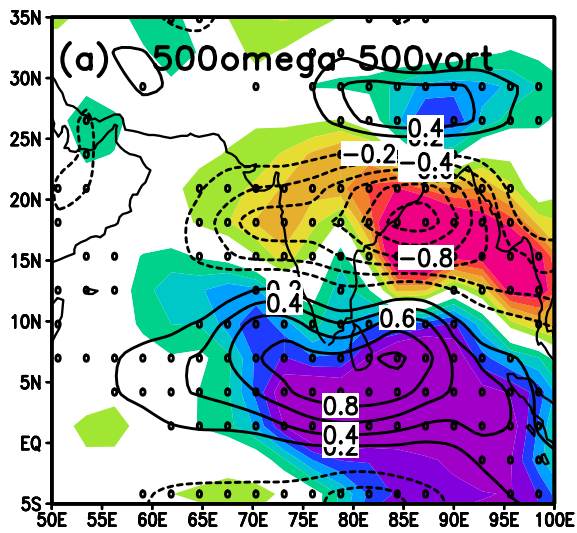


Fig. 5 Composite anomalies of vorticity ζ at 500 hPa for each of 5 clusters ($k = 5$), shown in contours in units of 10^{-5}s^{-1} . Composite anomalies of vertical pressure velocity (ω) at 500 hPa shown in shading, in units of $1.0^{-2} \text{Pa s}^{-1}$. Significance of the vorticity at the 95% level is indicated by the stippling. The clusters are labeled (a), (b), (c), (d) and (e) corresponding to the panels

rising motion and upper-level divergence over the Bay of Bengal as in Fig. 1b and c. The remaining 40% is divided nearly evenly between the regime with divergence closer to the tropics or over the Himalayas (Fig. 1a and d.) Considering 5 clusters, the system spends 61% of time in regimes (b)–(d) with upper-level divergence over the Bay of Bengal (Fig. 4b through d). The remaining two regimes (a) and (e) with rising motion to the north and south of India (Fig. 4a and e) are occupied for the remaining 39% of the time.

6.2 Regime transitions

Table 3 presents the total number of transitions between individual clusters for $k = 4$. The number of transitions between cluster i and cluster j is given in the i th row and j th column. The rows and columns are labeled (a) through (d) to correspond to the composites of Figs. 1, 2 and 3. The diagonal elements give the number of pentads for which the system stayed in the same cluster, and this is largest for clusters (b) and (c). The off-diagonal elements are generally smaller, and give the number of transitions. Bold values are significant at the 95% confidence level using the boot-strap method detailed in Sect. 3. Significant transitions include the complete cycle: $(a) \rightarrow (b)$, $(b) \rightarrow (c)$, $(c) \rightarrow (d)$, and $(d) \rightarrow (a)$. In addition, the number of transitions from cluster $(c) \rightarrow (b)$ is not only significant, but is also larger than the $(c) \rightarrow (d)$ leg of the complete cycle. Table 4 shows the number of episodes during which 3, 4 or 5 legs of the cycle occur consecutively, and the average length of each episode. Such episodes are counted as long as residence in one of the clusters is followed (after one or more pentads) by residence in the next cluster in the cycle. All possible sequences within the entire cycle are counted; for example the entry in the table for 3 legs includes all sequence $(a) \rightarrow (b) \rightarrow (c)$, $(b) \rightarrow (c) \rightarrow (d)$, $(c) \rightarrow (d) \rightarrow (a)$, and $(d) \rightarrow (a) \rightarrow (b)$.

The entire cycle from (a) through (d) back to (a) occurs only 7 times in the 40 years; 3 consecutive legs occur on average for over three quarters of the years. The average cycle length for the entire cycle is about 48 days.

6.3 Five clusters

Table 5 presents the total number of transitions between individual clusters for $k = 5$, in the same form as Table 3. The significant transitions encompass the complete cycle:

$(a) \rightarrow (b)$, $(b) \rightarrow (c)$, $(c) \rightarrow (d)$, and $(d) \rightarrow (e)$, and $(e) \rightarrow (a)$. In addition to the transitions encompassing the cycle, one transition in the opposite direction ($(b) \rightarrow (a)$) is seen. This transition will be discussed further in Sect. 7. Table 6 shows the number of episodes during which 3, 4, 5 or 6 legs of the cycle occur consecutively, and the average length of each episode. The episodes are counted in a way analogous to those in Table 4.

The entire cycle from (a) through (e) back to (a) occurs only 3 times in the 40 years, with an average length of 42 days. Subsets of the cycle occur increasingly more often as the number of required legs of the cycle decreases; 3 legs occur 70% of the time, taking about 23 days.

7 Discussion

7.1 Comparison of 4 and 5 clusters

In general, the cyclical sequence of states for $k = 5$ from $(a) \rightarrow (b) \rightarrow (c) \rightarrow (d) \rightarrow (e) \rightarrow (a)$ is very similar to the sequence for $k = 4$ of $(a) \rightarrow (b) \rightarrow (c) \rightarrow (d) \rightarrow (a)$. The states in the $k = 4$ sequence seem to occur at slightly earlier times than the corresponding ones in the $k = 5$ sequence. For example, the appearance of rising motion and upper-level divergence over the Himalayas appear at the “end” of the cycle using 4 clusters, but appear both at the “end” of the $k = 5$ cycle and the “beginning” of the next one. Thus we see slightly stronger rising motion over the Himalayas in Fig. 5a compared to 2a, as well as the weaker rising motion over the equatorial Indian Ocean in in Figs. 5e compared to 2d. Remnants of positive rainfall anomalies over the Himalayas and south India are seen in the $k = 5$ cluster (a) composite (Fig. 6a) which appear at end of the “previous” cycle for the $k = 4$ cluster (d) (Fig. 3d).

7.2 Relationship to active/break episodes and intra-seasonal oscillations

The sequences of rainfall composites presented here can be compared to composites of the active and break episodes presented by Rajeevan et al. (2010) (hereafter R). These authors define the dates of these isolated episodes on the basis of rainfall anomalies averaged over the Monsoon core region exceeding $\pm 1\sigma$ (standard deviation) for 3 consecutive days, using a high resolution data set of precipitation over India. The active phase lag-0 composite of R, showing large positive rainfall anomalies over the Monsoon core region of India, over the western Ghats, along with suppressed rainfall over southern India and to the northeast over the Himalayas, is very similar to cluster (c) for $k = 5$ (Fig. 6c), as well as with cluster (b) for $k = 4$ (3b). The break composite presented in R shows completely the reverse pattern, similar

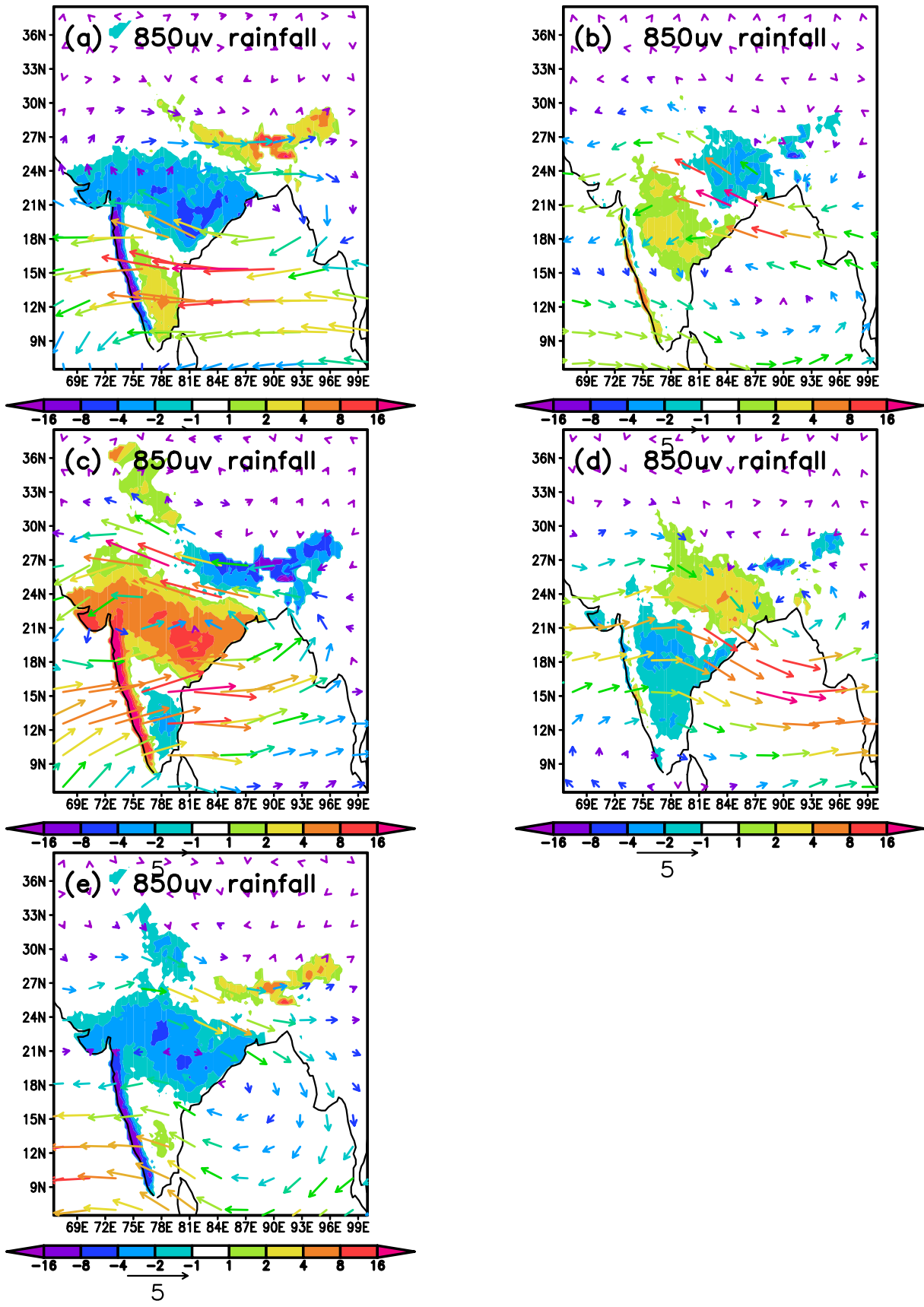


Fig. 6 Composite anomalies of horizontal winds (u , v) at 850 hPa for each of 5 clusters ($k = 5$), shown in vectors. Composite anomalies of rainfall shown in shading, in units of 10^{-3} md^{-1} . Only values of the precipitation significant at the 95% level are plotted. The clusters are labeled (a), (b), (c), (d) and (e) corresponding to the panels

to the $k = 5$ cluster (a) (and to some extent cluster (e)) in Fig. 6, and cluster (d) for $k = 4$ (Fig. 3d). It should be noted that the longer life-span of break phases compared to active phases found by R is not reflected in the average persistence times seen for the clusters. For example, Table 2 shows that for $k = 5$ cluster (c) (active phase) has the same average persistence (7.8 days) as clusters (a) and (e), which encompass the break phase.

Krishnamurthy and Shukla (2008), hereafter KS8, show rainfall composites based on 8 phases of both the 45-day oscillation and 28-day oscillation obtained from application of the MSSA technique to observed outgoing long-wave radiation. The composites for phases 7 and 8 of the 45-day oscillation (from Fig. 4 of KS8), with large negative OLR over India, correspond well to the circulation in regime (c) for $k = 5$ shown in Figs. 4 and 5. The opposite phases of KS8 (1 and 2) correspond to the circulation in regime (a) in Figs. 4 and 5. The correspondence of our results with the 28-day oscillation of KS8 is less clear. This oscillation shows a dipolar structure (in latitude) of OLR over the Indian longitudes, in contrast to the more tripolar patterns seen in both the 45-day oscillation and our cluster composites. The association of our results with the 45-day oscillation is consistent with the average length of time it takes for the entire preferred sequence of $k = 4$ or $k = 5$ clusters to complete, 42–48 days, with partial sequences completing in about 34 days (Tables 4 and 6).

The phases of the 45-day oscillation in diabatic heating, shown in Hazra and Krishnamurthy (2014) also map onto the 5 clusters shown in Fig. 8, with cluster (a) corresponding to phases 2 and 3 of the 45-day oscillation, cluster (c) corresponding to phase 6, and cluster (e) to phases 8 and 1. Note that as the strong equatorial Indian Ocean (EIO) vertically integrated heating in cluster (a) gives way to EIO diabatic cooling in cluster (c), the negative precipitation anomalies seen over India in Fig. 6 for cluster (a) become positive anomalies for cluster (c). This association with lack of convection in the EIO with positive Indian Monsoon rainfall is also seen on the inter-annual time scale, as discussed by Gadgil et al. (2004).

7.3 Mechanisms of transitions between clusters

The mechanisms of northward and eastward propagation seen in the BSISO have been discussed extensively in the literature in terms of Rossby wave dynamics and air-sea interaction (e.g. Kemball-Cook and Wang 2001; Zhang et al.

2018; Hazra and Krishnamurthy 2018). Hazra and Krishnamurthy (2018) emphasize the role of moisture preconditioning, in which large-scale circulations create positive moisture anomalies leading to convection, followed by and negative moisture anomalies before suppressed conditions.

In order to indicate the moisture availability in the set of 5 clusters, we show the precipitable water (PW) for the five clusters in Fig. 9. The northeastward propagation of moist anomalies, closely tracking the upper level divergence seen in Fig. 4, is very clear. Figure 10 shows the changes in PW just prior to each of the six significant transitions indicated in Table 5. For each regime transition, the figure shows the difference between the PW for the last pentad prior to the transition and the mean PW of the regime being left. For example, the PW difference in regime (a) five days prior to transitions to regime (b) (seen in Fig. 10a) shows an enhancement of moisture over the Arabian Sea and Bay of Bengal north of the area of large PW in regime (a). The enhancement leads to a northward movement of convection and corresponding areas of high PW as the system moves to regime (b). The change in PW prior to the transition from (b) to (c) consists of more of a strengthening and widening of the existing area of high PW, rather than a migration. The strengthening is consistent with the difference in moisture between regimes (b) and (c). The remaining transitions (c) to (d), (d) to (e) and (e) to (a) are all presaged by a movement of the PW anomalies.

The PW prior to the transition from (b) back to (a), which is not part of the active-break cycle, is shown in panel (f) of Fig. 10. A strong drying is seen extending from the Bay of Bengal west-northwestward across central India and intensifying in the Arabian Sea, while moistening occurs over the equatorial Indian Ocean to the south. This south-north dipole in moisture changes presages the drying over India and the southward movement of moisture leading back to regime (a).

8 Summary and conclusions

The cluster analysis presented here categorizes the evolution of the full circulation in time, rather than focus on only those components which are oscillatory. The preferred transition paths between the circulation regimes clearly reflect aspects of both the active/break episodes and the 45-day oscillation which is associated with northeastward propagation of the circulation. Yet significant transitions (from cluster (c) to (b) in Figs. 2 and 3, and from cluster (b) to cluster (a) in Figs. 5 and 6) are characterized by *equatorward* propagation, and there are a number of other (not statistically significant) transitions between clusters. Such transitions occur during periods when regular northward propagation is not seen.

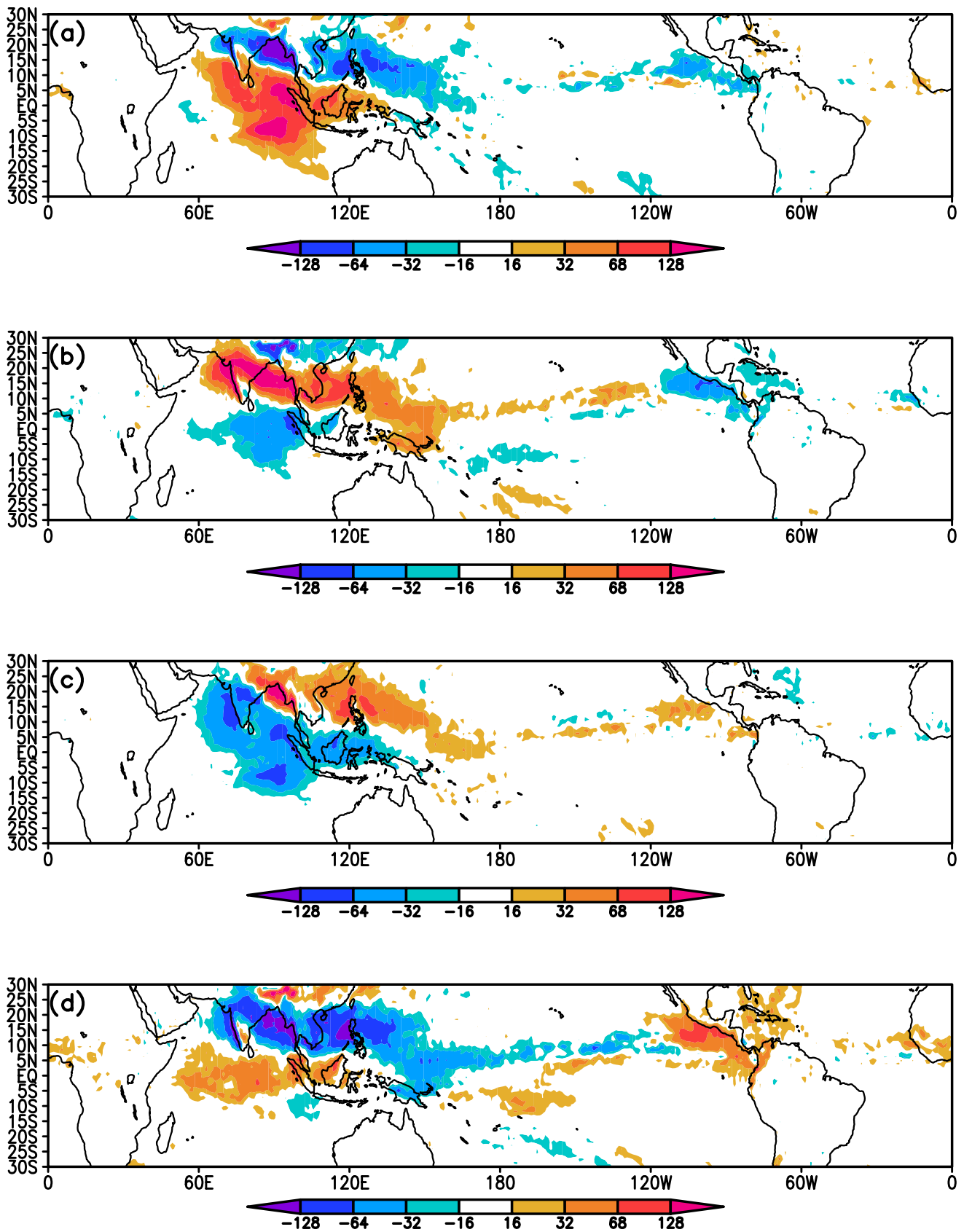
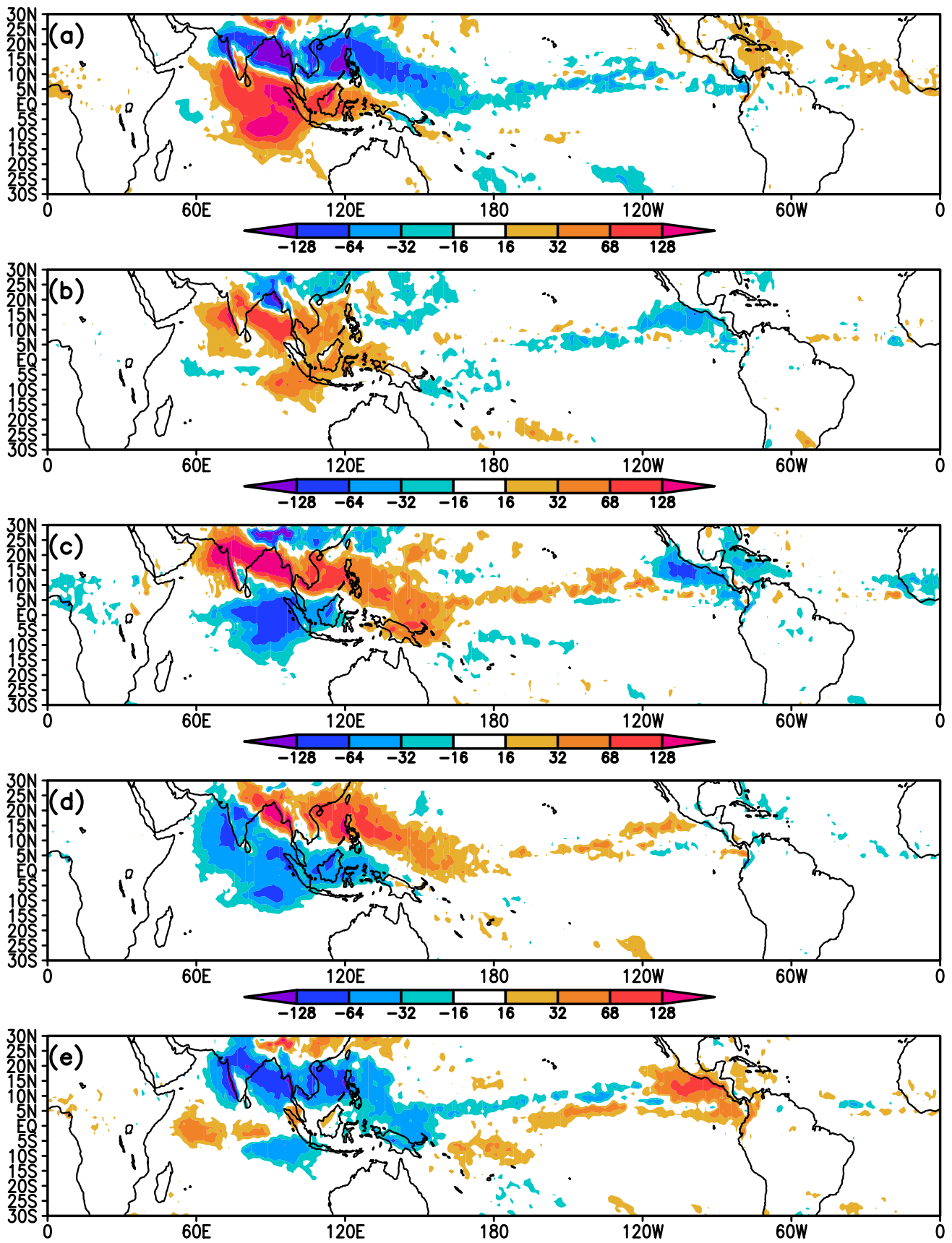


Fig. 7 Composite anomalies of vertically integrated diabatic heating for each of 4 clusters ($k = 4$), in units of W m^{-2} . The clusters are labeled (a), (b), (c), and (d) corresponding to the panels



◀ **Fig. 8** Composite anomalies of vertically integrated diabatic heating for each of 5 clusters ($k = 5$), in units of $W m^{-2}$. The color scale applies to all panels. The clusters are labeled (a), (b), (c), (d) and (e) corresponding to the panels

Table 1 Statistics of cluster episodes for 4 clusters: average length and total time

Cluster	Ave length		Total time	
	(Pentads)	(Days)	(Pentads)	(Percent)
(a)	1.627	8.1	205	21.4
(b)	1.838	9.2	283	29.5
(c)	1.731	8.7	277	28.9
(d)	1.696	8.5	195	20.3

Table 2 Statistics of cluster episodes for 5 clusters: average length and total time

Cluster	Ave length		Total time	
	(Pentads)	(Days)	(Pentads)	(Percent)
(a)	1.557	7.8	165	17.2
(b)	1.444	7.2	205	21.4
(c)	1.558	7.8	148	15.4
(d)	1.648	8.2	234	24.4
(e)	1.564	7.8	208	21.7

Table 3 Transition table for 4 clusters

COLUMN	(a)	(b)	(c)	(d)
ROW M				
(a)	79	62	28	29
(b)	26	129	89	26
(c)	43	59	117	44
(d)	47	27	35	80

Elements give number of transitions from cluster M (row) to column (N). The off-diagonal elements significant at the 95% confidence level are given in bold font

Table 4 Analysis of complete and partial cycles: (a) → (b) → (c) → (d)

No. of legs	Number		Avg. length	
	Per year	Total	Pentads	Days
3	0.77	31	5.16	25.8
4	0.40	16	6.75	33.8
5	0.17	7	9.75	47.9

Table 5 Transition table for 5 clusters

COLUMN	(a)	(b)	(c)	(d)	(e)
ROW M					
(a)	59	45	8	12	34
(b)	30	63	47	33	23
(c)	6	18	53	54	13
(d)	19	39	23	92	48
(e)	43	31	16	36	75

Elements give number of transitions from cluster M (row) to column (N). The off-diagonal elements significant at the 95% confidence level are given in bold font

Table 6 Analysis of complete and partial cycles: (a) → (b) → (c) → (d) → (e)

No. of legs	Number		Avg. length	
	Per year	Total	Pentads	Days
3	0.70	28	4.57	22.9
4	0.28	11	6.45	32.3
5	0.25	10	6.90	34.5
6	0.08	3	8.33	41.7

The evaluation of forecast and reforecast anomalies using the clustering approach presented has a number of advantages, including: (i) the focus on large-scale circulation patterns that are frequently seen in reanalyses; (ii) a natural way to describe the components of the active/break cycle in forecasts as well as the transitions that are not part of this cycle, and (iii) its applicability to short forecasts, for which application of time-filtering is not an option. The assignment of each forecast pentad anomaly to one of the observed regimes can be done on the basis of pattern correlation or mean-squared error, and leads to a coarse-grained representation of the (re)-forecast evolution. Ensembles of such (re)-forecasts can then meaningfully summarized in terms of the evolution of regime probabilities. To evaluate the fidelity of long simulations, another option is to apply the clustering procedure described in this paper to determine the realism of the preferred circulation states as well as their transition paths. In this way the presence (or absence) of a simulated active-break cycle can be obtained without the use of spectral time filtering.

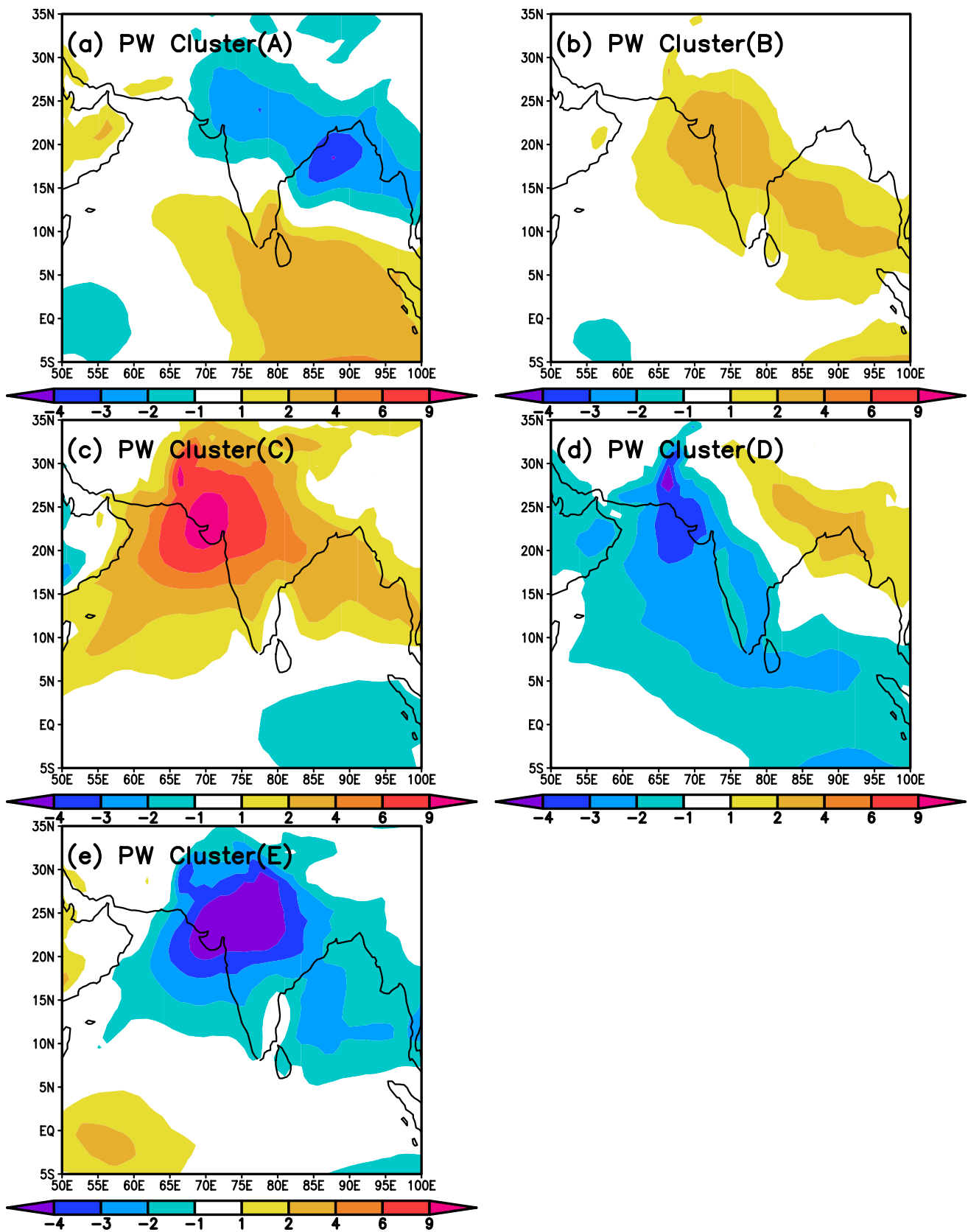


Fig. 9 Composite anomalies of precipitable water for each of 5 clusters ($k = 5$), shown in units of kg m^{-2} . The clusters are labeled (a), (b), (c), (d) and (e) corresponding to the panels

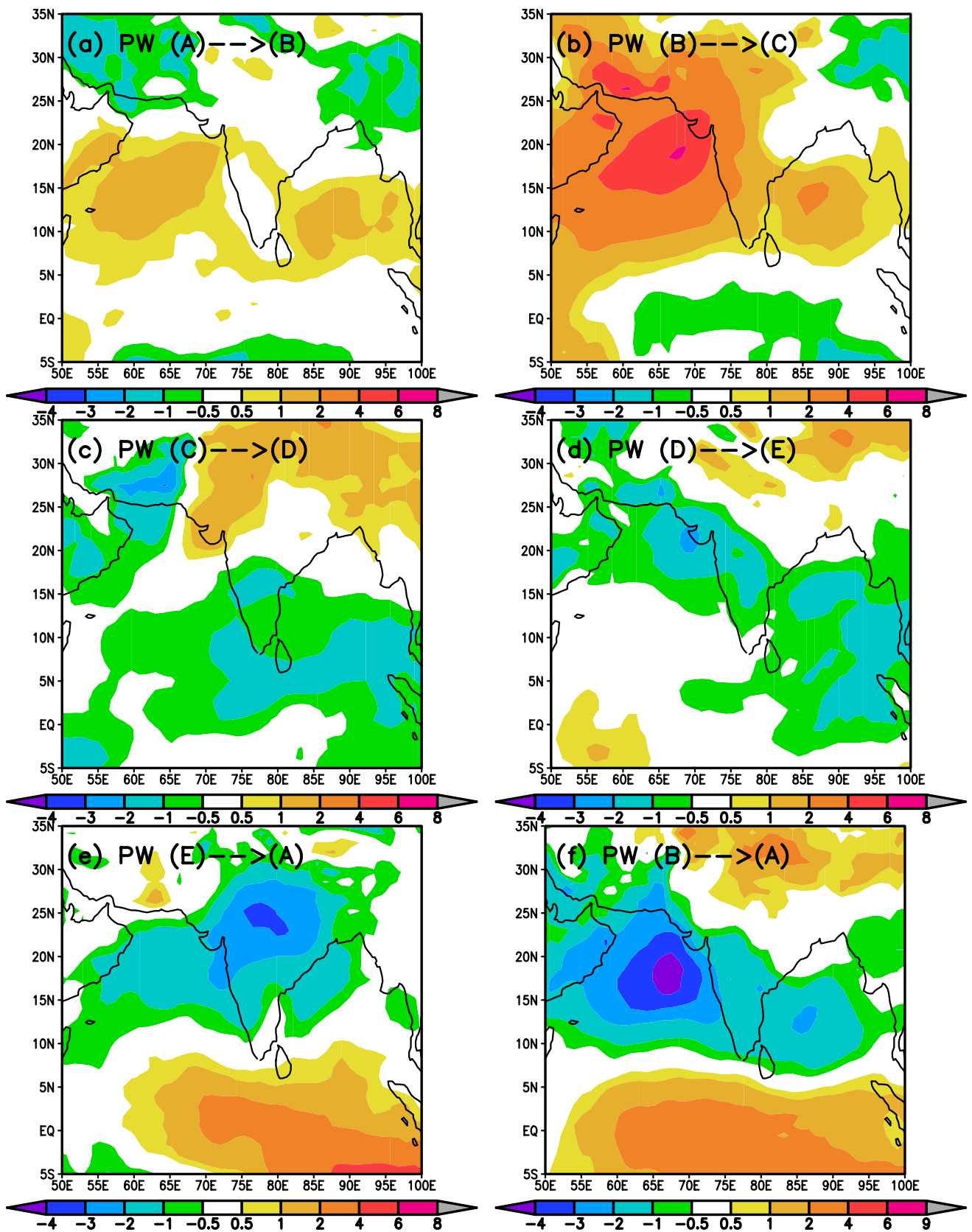


Fig. 10 Composite precipitable water (PW) differences prior to significant transitions for 5 clusters ($k = 5$), (units of kg m^{-2}). Each panel shows the difference between the last pentad prior to a transi-

tion and the mean PW of the regime being transitioned from. The transitions are labeled in the panels

Acknowledgements The authors gratefully acknowledge the Financial support given by the Earth System Science Organization, Ministry of Earth Sciences, Government of India (Grant no. IITM/MM-II/UnivGMU USA/2018/I) to conduct this research under the Monsoon Mission. The authors have no competing interests to declare that are relevant to the content of this article.

Funding The authors gratefully acknowledge the Financial support given by the Earth System Science Organization, Ministry of Earth Sciences, Government of India (Grant no. IITM/MM-II/UnivGMU USA/2018/I) to conduct this research under the Monsoon Mission.

Data and code availability The datasets generated during and/or analysed during the current study, as well as the Fortran codes, are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Annamalai H, Slingo J (2001) Active/break cycles: diagnosis of the intraseasonal variability of the asian summer monsoon. *Clim Dyn* 18:85–102. <https://doi.org/10.1007/s003820100161>
- Annamalai H, Sperber KR (2005) Regional heat sources and the active and break phases of boreal summer intraseasonal (30–50 day) variability. *J Atmos Sci* 62:2726–2748
- Chatterjee P, Goswami BN (2004) Structure, genesis and scale selection of the tropical quasi-biweekly mode. *Q J R Meteorol Soc* 130:1171–1194. <https://doi.org/10.1256/qj.03.133>
- Christiansen B (2007) Atmospheric circulation regimes: Can cluster analysis provide the number? *J Clim* 20:2229–2250
- Chu JE, Wang B, Lee JY, Ha KJ (2017) Boreal summer intraseasonal phases identified by nonlinear multivariate empirical orthogonal function-based self-organizing map (esom) analysis. *J Clim* 30:3513–3528
- Deday E, Simon JC (1976) Clustering analysis. In: Fu K (ed) *Communication and cybernetics 10 digital pattern recognition*. Springer-Verlag, pp 47–94
- Dee DP et al (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597
- Desbois M, Seze G, Swejwach G (1982) Automatic classification of cloudfs on METEOSAT imagery: application to high-level clouds. *J Appl Meteorol* 21:401–412
- Gadgil S, Vinayachandran PN, Francis PA, Gadgil S (2004) Extremes of the Indian summer monsoon rainfall, ENSO and equatorial Indian Ocean oscillation. *Geophys Res Lett* 31(L12):213. <https://doi.org/10.1029/2004GL019733>
- Hazra A, Krishnamurthy V (2014) Space-time structure of diabatic heating in monsoon intraseasonal oscillation. *J Clim* 28:2234–2255
- Hazra A, Krishnamurthy V (2018) Seasonality and mechanisms of tropical intraseasonal oscillation. *Clim Dyn* 50:179–199
- Hersbach H et al (2020) The ERA5 gobar reanalysis. *Q J R Meteorol Soc* 146:1999–2049. <https://doi.org/10.1002/qj.3803>
- Kemball-Cook S, Wang B (2001) Equatorial waves and air–sea interaction in the boreal summer intraseasonal oscillation. *J Clim* 14:2923–2942
- Krishnamurthi TN, Arduay P (1980) The 10 to 20-day westward propagating mode and “Breaks in the Monsoons.” *Tellus* 32:15–26. <https://doi.org/10.1111/j.2153-3490.1980.tb01717.x>
- Krishnamurthi TN, Bhalme HN (1976) Oscillations of a monsoon system. Part I. Observational aspects. *J Atmos Sci* 33:1937–1954. [https://doi.org/10.1175/1520-0469\(1976\)033%3C1937:OOAMSP%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033%3C1937:OOAMSP%3E2.0.CO;2)
- Krishnamurthy V, Shukla J (2000) Intraseasonal and interannual variability of rainfall over India. *J Clim* 13:4366–4377. [https://doi.org/10.1175/1520-0442\(2000\)013<0001:IAIVOR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0001:IAIVOR>2.0.CO;2)
- Krishnamurthy V, Shukla J (2007) Intraseasonal and seasonally persisting patterns of indian monsoon rainfall. *J Clim* 20:3–20. <https://doi.org/10.1175/JCLI3981.1>
- Krishnamurthy V, Shukla J (2008) Seasonal persistence and propagation of intraseasonal patterns over the Indian monsoon region. *Clim Dyn* 30:353–369. <https://doi.org/10.1007/s00382-007-0300-7>
- Lau KM, Chan PH (1986) Aspects of the 40–50 day oscillation during the northern summer as inferred from outgoing longwave radiation. *Mon Wea Rev* 114:1354–1367
- Lee SS, Wang B, Waliser DE, Neena JM, Lee JY (2015) Predictability and prediction skill of the boreal summer intraseasonal oscillation in the Intraseasonal Variability Hindcast Experiment. *Clim Dyn* 45:2123–2135. <https://doi.org/10.1007/s00382-014-2461-5>
- Michelangeli PA, Vautard R, Legras B (1995) Weather regimes: recurrence and quasi-stationarity. *J Atmos Sci* 52:1237–1256
- Murakami T, Nakazawa T, He J (1984) On the 40–50 day oscillation during 1979 Northern Hemisphere summer. Part 1: phase propagation. *J Meteorol Soc Jpn* 62:440–468
- Neena JM, Waliser D, Jiang X (2017) Model performance metrics and process diagnostics for boreal summer intraseasonal variability. *Clim Dyn* 48:1661–1683. <https://doi.org/10.1007/s00382-016-3166-8>
- Pai DS, Sridhar L, Rajeevan M, Sreejith OP, Satbhai NS, Mukhopadhyay B (2014) Development of a new high spatial resolution (0.25 deg x 0.25 deg) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. MAUSAM pp 1–18
- Pattanaik DR, Sahai AK, Muralikrishna R, Mandal R, Dey A (2020) Active-break transitions of monsoons over india as predicted by coupled model ensembles. *Pure Appl Geophys* 177:4391–4422. <https://doi.org/10.1007/s00024-020-02503-2>
- Prasanna V (2020) Impact of monsoon rainfall on the total foodgrain yield over india. *J Earth Syst Sci* 123:1129–1145
- Rajeevan M, Gadgil S, Bhatte J (2010) Active and break spells of the indian summer monsoon. *J Earth Syst Sci* 119:229–247
- Sahai AK, Abhilash S, Chattopadhyay R, Borah N, Joseph S, Sharmila S, Rajeevan M (2015) High-resolution operational monsoon forecasts: an objective assessment. *Clim Dyn* 44:3129–3140. <https://doi.org/10.1007/s00382-014-2210-9>
- Straus DM (1983) On the role of the seasonal cycle. *J Atmos Sci* 40:303–313
- Straus DM, Molteni F, Corti S (2017) Atmospheric regimes: the link between weather and climate. In: Franzke CLE, O’Kane TJ (eds)

- Nonlinear and stochastic climate dynamics. Cambridge University Press, pp 105–134
- Suhas E, Neena JM, Goswami BN (2013) An indian monsoon intra-seasonal oscillations (MISO) index for real time monitoring and forecast verification. *Clim Dyn* 40:2605–2616. <https://doi.org/10.1007/s00382-012-1462-5>
- Swenson ET, Straus DM (2021) A modelling framework for a better understanding of the tropically-forced component of the indian monsoon variability. *J Earth Syst Sci* 130. <https://doi.org/10.1007/s12040-020-01503-z>
- Zhang L, Han W, Li Y, Maloney ED (2018) Role of north indian ocean air–sea interaction in summer monsoon intraseasonal oscillation. *J Clim* 31:7885–7908. <https://doi.org/10.1175/JCLI-D-17-0691.1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.