

**Local versus regional-scale characteristics of monsoon onset and post-onset rainfall over  
Indonesia**

Vincent Moron<sup>°,\*</sup>, Andrew W. Robertson\*, Jian-Hua Qian\*,

<sup>°</sup> CEREGE, UMR 6635 CNRS, Aix-Marseille University, Aix en Provence, and Institut  
Universitaire de France, France

\* IRI, Columbia University, Palisades, USA

January 2009

**REVISED VERSION**

submitted to Climate Dynamics

## **Abstract**

The austral summer monsoon onset and post-onset rainfall and their associated low-level winds are analyzed during the August-February season over Indonesia from 1979 to 2006 using surface and satellite products as well as reanalyses and regional climate model simulations. Onset date is defined using a local agronomic definition. Its leading empirical orthogonal function is found to exhibit a regional-scale spatially-coherent signal across “monsoonal” Indonesia, i.e. mostly south of the Equator, with an asymmetric temporal behavior, such that delayed onsets are more intense than early ones. Associated anomalies in rainfall tend to weaken quickly after mid-to-late November or early December, especially over islands, while they tend to persist over ocean. This weakening is shown to be associated with the evolution of distinct weather types revealed by a *k*-means cluster analysis. In particular, late onsets—usually related to warm El Niño Southern Oscillation (ENSO) events—are found to be accompanied by an increased prevalence of a weather type characterized by weak low-level daily-averaged winds across monsoonal Indonesia and increased (decreased) rainfall over most of the island orography and southern and western coasts (seas). The regional model simulations provide evidence that this land-sea rainfall contrast could be associated with an enhanced diurnal sea-land breeze circulation.

## **1. Introduction**

Indonesia experiences a marked seasonal cycle in precipitation characteristic of a monsoon climate, especially near and south of the equator with the principal rainy season centered on December–February (DJF), and the dry season peaking in July–August (Aldrian and Susanto 2003; Giannini et al. 2007). The main rainy season is associated with the Australasian monsoon, i.e. the large-scale shift of the Inter-Tropical Convergence Zone (ITCZ) to the southern hemisphere with a northwesterly monsoon flow at low levels south of the equator (i.e. Chang et al 2004 2005; Wheeler and McBride 2005). The islands close to or north of the equator (such as Sumatra, Borneo and Sulawesi) have a less pronounced seasonal cycle with a decrease of rainfall rather than a real dry season around June–August (i.e. Aldrian and Susanto 2003; Aldrian et al 2005 2007). This regional-scale (i.e. corresponding roughly to the whole of Indonesia) nature of the monsoon is complicated at local-scale scales due to the complex island topography, warm waters (usually above 27°C), pronounced diurnal sea- and mountain-breeze circulations, and their interaction with eastward-propagating intraseasonal oscillations (i.e. Sakurai et al 2005; Shibagaki et al 2006; Kodama et al 2006), that produce large local-scale rainfall contrasts on distances less than 50–100 km (i.e. Hadi et al 2002; Wu et al 2003; Mori et al 2004; Ichikawa and Yasunari 2006; Qian, 2008).

Interannual variability over Indonesia is strongly related to the El Niño Southern Oscillation (ENSO) with anomalously low rainfall during warm events (Hackert and Hastenrath 1986; Ropelewski and Halpert 1987 1996; Kiladis and Diaz 1989; Hamada et al 2002; McBride et al 2003; Giannini et al 2007), associated with anomalous large-scale subsidence, a weakened Walker circulation (Klein et al 1999) with anomalous surface easterlies, and anomalously cool sea surface temperatures (SST) around Indonesia (Hendon 2003). However, the picture becomes more complex, at regional and finer spatial scales, especially during the rainy season. Spatial

coherence (Moron et al 2006 2007) of interannual rainfall anomalies and teleconnections with ENSO are stronger during the “dry” (JJA) and “transition” (SON) seasons, than during the peak rainy season (DJF) (Haylock and McBride 2001; Aldrian et al 2004 2005 2007; Giannini et al 2007). Hendon (2003) suggested that the weakening of the relationship between ENSO and Indonesian rainfall from JJA to DJF is linked to the transition between surface southeasterlies to northwesterlies. The easterly anomaly superimposed on the climatological mean winds during a warm ENSO event acts to reduce the wind speed during the monsoon season, damping the pre-monsoon local cold SST anomalies, thus reducing the negative DJF rainfall anomalies. Chang et al (2004) hypothesized that the weak correlation between Niño 3 SST index and CMAP rainfall during DJF over a triangular area between southern of Borneo, southeastern Sumatra and west-central Java, is related to the rain-shadow effect of the northwest-southeast oriented mountain chain along the southwest coast of Sumatra (i.e. *Pegunungan Bukit Barisan*) that weakens the influence of an anomalous Walker circulation. The decreased spatial coherence of DJF rainfall anomalies may also be partly due to regional-scale modes of variability, such as the north-south dipolar rainfall anomaly identified over Java by Giannini et al (2007), with less rainfall in the north and more in the south during warm ENSO events..

Consistent with the temporal modulation of the observed intensity of ENSO-related rainfall anomalies, the skill of seasonal rainfall forecasts is usually higher in the “dry” and transition seasons and rather low for the peak of the rainy season (Aldrian et al 2005 2007). This modulation has potential implications for rice production over the highly populated island of Java (Harger 1995; Naylor et al 2001 2002 2007). In that context the phase of onset of the austral summer monsoon season is particularly critical, due to its impact on planting dates, which also influence the second planting of rice at the end of the rainy season (Naylor et al 2007). Moron et al (2009a) recently argued that the spatially-coherent component of seasonally-average

(September–December, SOND) of local rain-gauge anomalies is largely due to the timing of the monsoon onset, and in particular with delayed onsets during warm ENSO events. Post-onset interannual rainfall anomalies were found to be far less spatially coherent, with some weak coherence limited to a southwest-northeast band extending from southern Sumatra and Java to southern Sulawesi (Moron et al 2009a). Thus, SOND seasonal rainfall anomalies, at least over islands, owe their spatial coherence (and potential predictability) to the systematic delay or advance of large-scale monsoon onset.

In this paper, the issue of the spatial coherence of onset and post-onset rainfall at interannual time scale is further investigated using several observed datasets, together with high-resolution regional climate model simulations driven with reanalysis data. We revisit the spatio-temporal structure of rainfall variability and its relationship with monsoon onset. The data sets and model simulations are described in section 2. Section 3 begins with an analysis of the interannual variability of onset date, using empirical orthogonal functions (EOFs). The seasonal evolution of rainfall anomalies are then examined for early and late onset years. A cluster analysis of daily reanalysis low-level wind fields is then used to identify the relationships between seasonal rainfall anomalies and modes of atmospheric variability in section 4. Section 5 provides a discussion and concludes the paper.

## **2. Datasets**

### *a. “Global Summary Of the Day” dataset*

Daily rainfall totals were extracted for Indonesia for the period 1979–2004 from the NOAA Climate Prediction Center (CPC) Global Summary of the Day (GSOD) dataset, archived at the National Center for Atmospheric Research (NCAR), and originating through the WMO Global Telecommunication System (GTS). Note that these data are not checked for inhomogeneities

and/or wrong reports. They should be then cautiously analyzed and viewed as the best available, albeit imperfect, estimate of the “true” rainfall. There are 91 available stations but many data are missing (> 40% for the October–March season). Respectively, 72, 57 and 38 stations have at least 25%, 50% and 75% of available daily observations. We then selected the stations having at least 50% of available data; these stations are concentrated over Java, with a moderate sampling of Sumatra and Borneo, and almost nothing elsewhere, except for few stations over Timor (Fig. 1). There are 20.5% of missing entries in the remaining 57-station network, mostly scattered in time. The missing entries were filled with a simple stochastic generator using the persistence of wet-to-wet and dry-to-wet days and the shape and scale of a gamma distribution as parameters (Wilks 1999); these parameters were computed separately for each calendar month. This method does not enhance any spatial coherence between stations, and does not attempt to reconstruct the interannual variability of missing periods.

Figure 2 illustrates the daily rainfall sequences for two stations located in central Java, 97 km apart from each other, indicated as red dots on Figure 1: Jatiwangi (108.27°E, 6.75°S) at the northern foot of *Gunung Ceremé* – a 3078 m high volcano – and Tegal, (109.15°E, 6.85°S) on the north coast of Java. Two years with almost complete records are illustrated: the strong El Niño year 1997–1998, and the ENSO-neutral year 2003–2004. The inland station (Jatiwangi) is clearly wetter than the coastal one (Tegal), with mean annual rainfall of 2254 mm and 1733 mm respectively. The seasonal cycle here is quite obvious with almost no rain between the end of July and late October, especially during the El Niño year. Dry sequences longer than 3–4 days during the rainy season are rare. False starts, defined as occurrence of a 5-day wet spell receiving > the mean rainfall amount of 5-day wet sequence in August–February, i.e. 39 mm at Tegal and 48 mm at Jatiwangi, followed by a 10-day dry spell, are rather frequent as seen in Tegal in 1997 (Fig. 2b) and 2003 (Fig. 2d). The “real” onset of the rainy season (see section 3a) at Tegal —defined in the

same way, but without the occurrence a 10-day dry spell ( $< 5$  mm) within the following 30 days—is delayed during both years by 36 days in 1997 (Fig. 2b) and by 16 days in 2003 (Fig. 2d) relatively to the 1979-2003 mean, i.e. November 4 at Jatiwangi and November 28 at Tegal). In consequence the negative anomaly of seasonal amount of rainfall is clearly larger in 1997/98 (-20%) relatively to 2003/04 (-5%), mostly due to the decreased length of the rainy season and the concomitant decrease of the frequency of wet days  $> 1$  mm; the mean daily intensity of rain is almost identical in both years. The difference between these two years is also clear in Jatiwangi with a delay of 35 days in 1997 (Fig. 2a) reduced to 10 days in 2003 (Fig. 2c). But in this case such distinct onsets are not necessarily related to a large negative anomaly in seasonal amount. The daily correlation between the square-rooted rainfall at each station is only 0.33 during August–February, and 0.22 during November–February, computed over all years. Considering (square-rooted) pentads increases the correlations to 0.51 for August–November season, but to 0.23 only for the core of the rainy season (November–February).

This example illustrates (i) the challenge of defining the onset from daily rainfall amounts, (ii) the large interannual variability in onset date, not necessarily related to the seasonal rainfall total, and (iii) the large spatial inhomogeneities in rainfall over very short distances at intraseasonal time scales.

#### *b. CMORPH dataset*

The CPC morphing technique (CMORPH) uses precipitation estimates from low-orbiter satellite microwave observations, and whose features are evolved between observations via spatial propagation information obtained from geostationary infrared satellite data (Joyce et al 2004). We used the version at  $0.25^\circ$  resolution from December 7 2002 to May 31, 2007. This dataset is

primarily used as reference for the regional climate model simulations, which have similar horizontal resolution (see section 2e).

Ebert et al (2007) indicated a good correspondence between the CMORPH estimates and observations over tropical Australia (i.e. north of 25°S) at 0.25° horizontal scale, even in DJF when rain is largely convective in nature. The correlations between square-rooted of daily (pentad) amounts for Jatiwangi and Tegal and their corresponding CMORPH grid-points are 0.35 (0.73) and 0.43 (0.73) respectively, during November 2003–February 2004, suggesting that the CMORPH is more reliable at the pentad scale. Similar values were obtained in other years.

*c. The pentad CMAP data*

The primary dataset used in the following analysis is the pentad CPC Merged Analysis of Precipitation (CMAP) available on a 2.5° x 2.5° grid since 1979 (Xie and Arkin 1997 1998). The version used here (see for example <http://www.cdc.noaa.gov/cdc/data.cmap.html>) ends in February, 24 2007 and the 28 seasons from October 1979 to February 2007 are used in the following analysis. Values are derived from 5 kinds of satellite estimates (GPI, OPI, SSM/I scattering, SSM/I emission and MSU) together with in-situ surface observations. There are typically 1–3 rain-gauge station per grid-squares (Xie, personal communication). The version without reanalyses estimates is used to avoid any bias related to parameterization. There are no missing values in this region.

To pursue our limited comparison using Jatiwangi and Tegal both stations lie within the same 2.5° CMAP grid-box. The correlation between square-rooted pentad amounts for November–February season equals 0.59 for Jatiwangi and 0.46 for Tegal and reaches 0.66 when the average of both stations is used instead of each station.

#### *d. Reanalysis wind dataset*

The daily 850 hPa NCEP I winds (Kalnay et al 1996) are extracted on August-February 1979-2007 on 12°S-6°N, 90°E-130°E on a 2.5° grid.

#### *e. Regional climate model simulations*

We make use of simulations made with the RegCM3 regional climate model driven by NCEP reanalyses, at 60 km and 25 km resolution in 1979-2006 period. RegCM3 is the Abdus Salam International Centre for Theoretical Physics (ICTP) regional climate model version 3 (Pal et al 2007; Qian 2008), a mesoscale atmospheric model including physics parameterizations of land-surface, radiation, and moist processes suitable for long-term climate simulation. These simulations are referred to as RegCM-25 and RegCM-60 in the following. These experiments have been detailed elsewhere (Qian 2008). Figure 3 shows the topography of RegCM-60 (panel b) and RegCM-25 (panel c), compared with observation (panel a). The main mountain ranges extend along the western and southern coasts of Sumatra and Java, while they are located in the central northern part of Borneo. Sulawesi is mountainous almost everywhere (Fig. 3a). The geography of the mountain ranges is rather well reproduced in RegCM-25 (Fig.3c) although the tallest peaks, above 2000m, are strongly attenuated. However, only the main ranges, strongly smoothed, are resolved in RegCM-60 (Fig. 3b). The mountain ranges are usually composed of groups of volcanoes separated by lowlands. The relatively small isolated volcanoes as *Gunung Baru* (3726 m) on Lombok island (Sunda islands) are not represented in RegCMs.

### **3. Spatial coherence of monsoon onset-date and post-onset rainfall**

#### *a. Definition of onset date*

From an agronomic perspective, monsoon onset is defined by the first “significant” rains, allowing sowing crops, without any long dry spells then after, potentially damaging during the first stages of growth of crops (Stern et al 1981; Sivakumar 1988). This definition is sensitive to the amplitude of daily rainfall intensities and any systematic model biases, as well as to topography which can induce considerable small-scale variability; it does not filter out the noise associated with the convective nature of tropical rainfall. The intensity and length of initial wet spell and following dry spell should ideally rely on local crop and soil properties, but values of 20-40 mm in 2-5 days (initial wet spell) and 1-5 mm in 7-15 days (following dry spell) are usually considered (Stern et al 1981; Sivakumar 1988; Omotosho 1992; Moron et al 2009a,b). A recent study (Marteau et al 2009) indicated that the interannual variability of local-scale onset is largely insensitive to the choice of thresholds used to define the initial wet spell and following dry spell.

We considered previously the onset in Indonesia as the first wet day ( $> 1$  mm) of a 5-day wet sequence receiving at least 40 mm (Moron et al 2009a). We adopt here a more flexible approach by considering the first wet day  $> 1$  mm of a 5-day wet sequence receiving more than the climatological mean amount of rainfall received in 5-day wet sequences ( $> 1$  mm) in August-February, not followed by a 10-day dry sequence receiving less than 5 mm in the following 30 days counted from onset. Mean rainfall amounts during 5-day wet sequences are displayed on Fig. 4. The values are fairly homogeneous for CMAP and GSOD (Fig. 4a) with  $> 80\%$  of values between 25 and 40 mm in both datasets; GSOD values often exceed those of CMAP over Borneo. The CMORPH (Fig. 4b) is rather similar despite a different and shorter period taken into account. The eastern part is drier than in CMAP with relatively dry areas surrounding the southern and eastern islands. In CMORPH and RegCMs, heavy rainfall is concentrated over islands, with highest rainfall located over mountains. Qian (2008) found that this is caused by the convergence

of sea breezes, reinforced by valley breezes, over the islands during the afternoon hours. The cumulus merger process also acts to multiply rainfall over the islands. A narrow strip surrounding the islands, relatively drier than the outer seas, is visible in the CMORPH and the RegCMs, located in between the day-time rainfall maximum over the islands and the nocturnal rainfall maximum over the seas. However, in the RegCMs, (Fig. 4c,d), the land-sea contrast appears to be overestimated. This suggests that the convective scheme underestimates the strength of deep convection over ocean and that the diurnal land-sea breeze is probably overestimated (Qian 2008). The bias could be from physics parametrization of convection, or those associated with boundary layer and surface-air interaction. Aldrian et al (2005) found that their simulated rainfall is also biased in a regional atmospheric model over the Maritime Continent, albeit with overestimation over sea rather than over land. By coupling their atmospheric model to an ocean model, their simulated rainfall over ocean was improved.

*b. Mean onset date*

Maps of the mean onset date are plotted in Fig. 5. The three observational and satellite datasets exhibit similar patterns (Fig. 5a,b), despite the different data periods and very different horizontal resolutions. The mean onset dates, as well as the general NW-SE shift north of the equator followed by a more zonal propagation south of the equator, are in good agreement with previous estimates, drawn from different datasets and/or methods (Janowiak and Xie 2003; Naylor et al 2007). The larger spatial noise in CMORPH relative to CMAP could be related the much higher spatio-temporal resolution of the CMORPH, and its shorter 5-yr record. The spatial distribution of mean onset date in RegCMs compares well with the observed datasets, with the main distinguishing feature being the earlier model onsets over Java and the southeastern chain of islands extending to Timor. This is more marked in the 25-km resolution model, where earlier onsets occur over topography. While these earlier onsets are probably exaggerated in the model, they may be underestimated in the CMORPH, while GSOD stations do not tend to be located

over mountains. Higher model resolution leads to later onsets over the South China Sea (Fig. 5d vs. 5c), as also seen in the CMORPH but not CMAP (Fig. 5b vs 5a). It should be noted that most of the area north of the equator has a different seasonal cycle (Aldrian and Susanto 2003) and “onset” is merely, at best, an increase of rainfall. Its exact phase is thus more sensitive to the methods and parameters used to define it.

*c. Interannual variability of onset*

The leading EOF of onset-date is plotted in Fig. 6, for the CMAP, GSOD and RegCM-25 datasets, and provides a measure of the spatially coherent component of variability. It accounts for 35%, 33%, and 26% of onset-date variance in the three datasets respectively (30% in RegCM-60), while the second mode explains less than 12.2% in all cases. These similar variance fractions, demonstrate the dominance of the EOF 1 patterns, which are also similar across the datasets: in CMAP an almost continuous band of high loadings extends from the tropical southeast Indian Ocean to Sulawesi across southern Sumatra, Java and southern Borneo—i.e. “monsoonal” Indonesia (Aldrian and Susanto 2003)—with a weak decrease toward the southeast and a stronger decrease to near zero over northern Sumatra, the Malay Peninsula and northern Borneo (Moron et al 2009a). GSOD rain-gauge loadings (filled circles in Fig. 6a) are largely consistent with those of CMAP although a slightly larger amount of noise is visible over islands. These large EOF loadings signify substantial spatial coherence and a seasonally predictability component of onset date (Moron et al 2009a). The leading EOF of RegCM-60 exhibits a very similar pattern to that of CMAP (not shown).

The leading EOF of RegCM-25 onset date (Fig. 6b) also exhibits a southwest-northeast band of large loadings over ocean, but this band is broken up over land, where loadings are often much

smaller. However, this local-scale noise over land is found to largely cancel when a simple 5-point spatial smoothing is applied (not shown).

The temporal behavior of the respective leading PCs of the different datasets (Fig. 6c) is highly coherent, except for the periods 1983–1987, and 1998 onwards (correlations between pairs of PC are always  $> 0.87^{***}$  (one, two and three asterisks indicate significant correlations at the two-sided 90%, 95% and 99% level according to a random-phase test, Janicot et al 1996) on the common period 1979-2003). The time evolution is asymmetrical with more marked delayed onsets, and can be interpreted using indices of central and western Pacific SST anomalies (Fig. 6d), consistent with previous studies (e.g. Haylock and McBride 2001; Hendon 2003). The correlations between the average of standardized leading PC of onset date from the four datasets and July SST, one to four months before the onset date across Indonesia, equals  $0.75^{***}$  for central Pacific SST index and  $-0.71^{***}$  for western Pacific SST index, both indices being correlated at  $-0.44^{**}$  (all correlations computed on 1979-2003). Both SST regions appear to act constructively, although the western Pacific SST alone seems to be decisive in some years (e.g. 1994) and the ENSO signal in central Pacific in others (1987). The SST tendency between July and October in the central Pacific is also significantly related to the average of standardized leading PC of onset date ( $r = 0.62^{***}$ ), but not in the western Pacific ( $r = -0.11$ )

#### *d. Mean seasonal evolution of rainfall for late and early onset years*

The previous section suggests that the interannual variability of onset date is largely in-phase across the monsoonal Indonesia, especially over ocean, while small-scale noise is larger over islands, especially north of the Equator where the onset is merely an increase of the rainfall rather than an abrupt transition (Aldrian and Susanto 2003; Aldrian et al 2005). Moron et al (2009a) hypothesized that most of the spatially-coherent signal of rain-gauge SOND seasonal amounts is

conveyed by the phase of the regional-scale onset while post-onset rainfall anomalies lack spatial coherence and are largely unpredictable. This section explores this issue in greater depth by analyzing the evolution of rainfall anomalies from before the onset to the seasonal peak. Two sets of years are defined using the simple average of the standardized leading PC (Fig. 6c) derived from each of the 4 datasets (CMAP, GSOD, RegCM-60 and RegCM-25) over the common period (1979-2003). The lower and upper quartile categories of the average of the four standardized PC time series on the common period (i.e. 1979-2003) in Fig. 6c are used to define the early onsets (i.e. 1980, 1988, 1995, 1996, 1998, 1999) and late onsets (i.e. 1982, 1987, 1991, 1994, 1997, 2002) respectively. The remaining 13 years (i.e. within the inter-quartile range of the averaged PC time series) are taken to define neutral-onset years.

The domain-averaged August–February evolution of running 30-day rainfall anomalies are plotted in Fig. 7, for the sets of early and late onset years vs the long-term mean. In the observed datasets (Fig. 7a), the positive rainfall anomaly during early onset years (black curves) decreases from about 50% of the mean in late August, to less than 20–30% thereafter, and no significant anomalies are observed from about mid-November. In contrast, the negative rainfall anomalies during late onset years are larger and last longer: CMAP anomalies remain systematically negative through the rainy season, while the GSOD anomalies lose statistical significance from late-November, i.e. the beginning of the peak rainfall season. These differences between CMAP and GSOD may be related to land-sea sampling differences, since GSOD rain gauges are restricted to land.

The evolution of RegCM-25 simulated rainfall anomalies (Fig. 7b) show similar characteristics, although the composite anomalies appear more stable over time; a similar picture is obtained for RegCM-60 (not shown). The high-resolution model allows us to separate out the island and sea

components. We here focus on the area south of the equator, where the onset is better defined, and where most of the spatially-coherent large-scale interannual signal is located (Fig. 6a,b). The mean rainfall anomalies over land points (Fig. 7c) tend to be less than 20% from the mean after late-September in early onset years and late-October in late onset years, matching well the GSOD behavior (Fig. 7a). In contrast, the mean rainfall anomalies over ocean (Fig. 7d) tend to be much larger and more stable across the season. Note that the evolution of RegCM-60 is highly similar (not shown).

These analyses suggest that (i) the large-scale signal (i.e. negative rainfall anomalies during late onsets and positive anomalies during early onsets) tends to decrease in amplitude within the core of the rainy season, and that (ii) the behavior is distinctly different between land and sea, with sustained anomalies over ocean persisting well into the rainy season, while island anomalies decay rapidly after mid to late November.

#### **4. Regional-scale weather types and local-scale rainfall**

##### *a. Definition of weather types*

In order to understand better the differences in seasonality of rainfall anomalies seen in Fig. 7, we next analyze the evolution of atmospheric daily circulation using cluster analysis. Cluster analysis has been used to isolate modes of daily circulation variability in mid latitudes (Michelangeli et al 1995; Robertson and Mechoso 2003), and recently the tropics as well. In the latter case, Moron et al (2008) applied K-means to unfiltered daily circulation fields over West Africa, providing a integrated view of the complex regional circulation variability, without the need for explicit time filtering. Their weather types included typical phases of African easterly waves, persistent monsoon surges, and different stages of the seasonal cycle; the temporal sequencing of the

clusters described a systematic monsoonal evolution, together with variability at subseasonal and interannual time scales.

Here we use cluster analysis of daily unfiltered circulation fields to elucidate interannual variability in the monsoon and its onset, and to interpret the rainfall anomalies in our observed datasets and RegCM simulations; note that the rainfall is *not* used in defining the weather types. Daily wind fields at 850 hPa within a regional window (90°–130°E and 12°S–6°N) were extracted from the NCEP/NCAR reanalysis, and then analyzed using a  $k$ -means cluster analysis (Diday and Simon 1976). In this partitioning algorithm, a prescribed number ( $k$ ) of clusters is specified and daily observations are agglomerated around centroids initially chosen from random seeds. The  $k$ -means clustering iteratively minimizes the sum, over all clusters, of the within-cluster spread. The final clusters localize relatively high concentrations of data in the atmospheric phase space. The period analyzed encompasses the onset as well as the peak of the annual rainfall (i.e. August 1 to February 28), with the 850-hPa level chosen to represent the low-level monsoon flow. The zonal and meridional wind components were firstly standardized to zero mean and unit variance at each grid point, and the leading 9 PCs accounting for 75% of the variance were then used as input to  $k$ -means.

Figure 8a shows the classifiability index (CI) defined by Michelangeli et al (1995), to help identify the appropriate number of clusters by quantifying the sensitivity of the classification to the choice of initial seeds. This index is based on the similarity between the cluster centroids obtained from 500 random initializations of  $k$ -means for each value of  $k$ ; CI equals 1 if the classification is insensitive to the choice of the random seeds and 0 if the final solution is fully dependent on it, and thus meaningless. Such perfect robustness is achieved in our case only for  $k=2$ , with a large drop for  $k=3$  and  $k=4$ ; the CI rises to an almost flat plateau for higher values of

*k.* The solution with  $k=5$  clusters appears to be a good compromise between robust clusters and a useful categorization of the low-level atmospheric circulation variability. These five clusters are referred to as “weather types” (WT) in the following.

*b. Wind and rainfall signature of weather types*

The mean seasonal cycle of each WT is shown in Fig. 8b. There is a clear domination by WT#1 then WT#2 during August–mid-October; other WTs almost never occur before late-October. This period should be considered the “pre-onset” season for monsoonal Indonesia, while rainfall already occurs north of and around the Equator. Marked changes take place from mid-October with a progressive disappearance of pre-onset WTs and the increased occurrence of WT#3, #4 or #5. These three “monsoonal” WTs can occur until the end of February, although the prevalence of WT#3 is lower from early January (Fig 6b). The sequential progression from WT#1 to WT#5 characterizes the mean seasonal evolution in the analysis. Indeed, the daily probability transition matrix between WTs (Vautard et al 1990) shows that the one-day persistence (always  $> 0.79$ ) is the only transition that occurs more likely than by chance; thus there are no systematic transitions on the daily time scale between WT#1 and WT#2 before the core of the rainy season, nor amongst WT #3, #4 and #5 during the monsoon.

Composites of 850 hPa wind for the days assigned to each weather types are plotted in Fig. 9, together with composites of CMAP and GSOD rainfall for those same sets of days, recalling that the precipitation data sets were not used in defining the weather types. WT#1 and WT#2 typify the pre-monsoon season, with dry conditions and easterlies south of the Equator; there is a shift toward wetter conditions over southern Sumatra in WT#2. These WTs are thus labelled “pre-monsoonal” and “transitional” respectively. The WTs 3–5 represent monsoonal weather types. The wind fields are clearly different for WT#3 and WT#5 with widespread westerlies between

4°N–4°S in WT#3 (Fig. 9c), and south of 4°S in WT#5 (Fig. 9e), characterizing the evolution of the Australasian monsoon. Rainfall > 5 mm/day is widespread except at the margins. It is notable that WT#4 (Fig. 9d) depicts a weaker monsoon than WT#3 and WT#5, mostly south of the equator; the mean winds are very weak from eastern Indian Ocean toward Timor, and rainfall also is generally lower than for WT#3 and WT#5. The WT#4 (Fig. 9d) closely resembles the “easterly wind regime” identified by Ichikawa and Yasunari (2006) from smoothed 850 hPa zonal winds across Borneo during periods of inactive convection over the Maritime continent; conversely, WT#5 (Fig. 9e) is similar to their strong westerly regime, corresponding to active convection and regional-scale upper divergence. The monsoonal WTs 3–5 also resemble different phases of the MJO, which is usually defined using bandpass filtered data (Sperber 2003; Wheeler and Hendon 2004). A composite analysis (not shown) keyed to the 8 MJO’s phases defined in Wheeler and Hendon (2004) indicated that WT#3 and WT#5 are more prevalent during MJO phases of active regional-scale convection, while WT#4 is more prevalent during inactive phases. In the following, WT#3, #4 and #5 are referred to “westerly”, “quiescent” and “strong-westerly” monsoonal WTs.

*c. Weather types: interannual variability*

The weather types are constructed from unfiltered data and thus represent both the seasonal monsoonal evolution, as well as variability on longer and shorter time scales. The interannual variability in the frequency of occurrence of each WT is shown in Fig. 10a. There is an out-of-phase temporal relationship between the transitional and quiescent WTs on the one hand, and the westerly and strong-westerly ones on the other; this is especially clear for the warm ENSO events of 1982, 1991, 1997, 2002 and 2006 (Fig. 10a). In 1982 and 1997 in particular, the “transitional” and “quiescent” WTs together account for more than 75% of days over the whole season. Antecedent warm Niño 3.4 SST anomalies in July are associated with less of the “westerly” (and

secondly “strong-westerly”) WTs and more of the “quiescent” (and secondly “pre-monsoonal”) WTs than usual (Table 1). The frequency of the “transitional” WT seems more closely associated with the Indian Ocean Dipole (IOD), with increased frequency when the IOD is positive, i.e. when western Indian Ocean is anomalously warm relatively to the east (Table 1). Note that in July, the positive correlation between IOD and Niño 3.4 is not yet evident ( $r$  between IOD and Niño 3.4 = 0.70\*\*\* in October) while the negative one between IOD and western Pacific is already observed (Table 1).

Differences in the seasonal evolution of each WT during early and late onset years (as defined in Sect 3d) are examined in Figs. 10b and 10c. Comparing with climatological seasonality of WT frequency (Fig. 8b), early onset years (Fig. 10b) exhibit slightly higher prevalence of the “transitional” WT during August, and of the “pre-monsoonal” WT in October, at the expense of the “transitional” one. The largest differences during early-onset years occur from November onward, with increased prevalence of the “westerly” WT (and to some extent the “strong-westerly” one) at the expense of the “quiescent” regime. During the late onset years (Fig. 10c), the seasonal cycle is highly perturbed throughout the season. The “pre-monsoonal” WT dominates until mid-September, thus delaying the onset across Indonesia north of about 2°S (Fig. 9a). The “transitional” WT then predominates until mid-November, delaying further the onset south of the equator (Fig. 9b). The ensuing monsoon is then dominated by the “quiescent” WT, leading to recurrent negative rainfall anomalies across the seas, but a blurred signal across the islands (Figs. 9 and 10). Both westerly WTs are highly suppressed.

In order to examine in more detail the spatial structure of interannual anomalies in terms of weather types, we now construct anomaly composites of precipitation for the three monsoonal WTs (i.e. WTs 3–5) using the high-resolution RegCM-25 and GSOD (Fig. 11) datasets. Note

that the same NCEP/NCAR reanalysis wind fields were used in the RegCM's lateral boundary conditions as in the cluster analysis. These anomaly composites are constructed by subtracting the mean seasonal cycle of each weather type, defined by the mean seasonal cycle of total rainfall at each grid point, weighted by the mean seasonal cycle of the frequency of the respective weather type (Fig. 8b). Note that the 72 GSOD stations having at least 25% of data are retained. Missing entries are not considered here. Composite rainfall in Figure 11 is expressed as percentage anomalies relative to the mean seasonal cycle of each WT; a positive rainfall anomaly in these maps means that conditions are wetter when the WT is more frequent than normal, and vice versa, taking into account the mean season cycle. The three monsoonal WTs are strikingly different from each other in RegCM-25 with generally positive anomalies across the seas in the "westerly" (Fig. 11a) and especially the "strong-westerly" (Fig. 11c) types, while the "quiescent" WT (Fig. 11b) is anomalously dry across the seas and also along the northern coast of Java. The composite rainfall anomalies are noisier and weaker over islands in all three WTs, with the exception of NE Borneo, and parts of Sulawesi. The GSOD network tends to agree with RegCM-25 although very few stations are available on the southern shore of Java; the agreement is particularly good over southern 2/3 of Borneo (Fig. 11).

The CMORPH rainfall WT composites (not shown) exhibit similar qualitative broad-scale patterns in sea-island contrasts, compared to RegCM-25, despite the short length of the dataset and a different available period. However, the CMORPH anomalies are more coherent and extensive over the islands than in the regional model, suggesting that either the latter is too noisy, or that CMORPH underestimates small-scale rainfall features associated with topography.

The contrasting rainfall anomalies in the monsoonal WTs along the southern coast of Java and southwestern coast of Sumatra in RegCM-25 coincides with the topography. It is intriguing that

the “quiescent” WT (Fig. 9d), is characterized by positive rainfall anomalies over and south of the mountains (Fig. 11b), while the stronger monsoonal winds of the westerly and strong-westerly WTs (Figs. 9c,e) are associated with negative rainfall anomalies over the topography and along the southwestern and southern coasts (Figs. 11a,c). This suggests an interaction between the large-scale wind field, the topography and the diurnal cycle, in which the diurnal cycle may be stronger during the “quiescent” WT because of the weak regional-scale monsoonal flow. To examine this possibility in detail, Fig. 12 shows RegCM-25 composites of raw values and anomalies of near surface winds (at 10 m height) and rainfall for monsoonal WTs over Java, during the late afternoon hours (4–7 p.m. LT) near the peak of the diurnal cycle of rainfall (Qian 2008). In terms of the raw model fields (i.e. without the mean seasonal cycle subtracted), the three WTs show rainfall maxima across Java, while surrounding seas are clearly drier, especially for “westerly” and “quiescent” WTs (Fig. 12c,e). In terms of anomalies from the mean seasonal cycle, the model’s monsoon westerlies are strengthened over the oceans in the “westerly” (Fig. 12b) and “strong-westerly” (Fig. 12f) WTs, and weakened in the “quiescent” WT (Fig. 12d), consistent with the reanalysis winds in Fig. 9. Furthermore, the “quiescent” WT exhibits a strengthened on-shore flow, corresponding to strengthened sea breezes in the afternoon, accompanied by positive rainfall anomalies over the mountains and southern shore. This contrasts with the “westerly” and “strong-westerly” WTs in which the model exhibits weakened on-shore flow and negative rainfall anomalies. Thus, the diurnal cycle is indeed modulated by these WTs in the manner hypothesized, with land-sea and mountain-valley breezes less disturbed by the monsoon in the “quiescent” WT during which large-scale wind speeds are small.

## **5. Summary and discussion**

### *a. Summary*

This paper has analyzed the characteristics of monsoon onset date and post-onset rainfall over Indonesia in three observed rainfall datasets (CMAP, CMORPH, GSOD), as well as regional climate model simulations with 60-km and 25-km spatial resolution (RegCM-60, RegCM-25), driven by reanalysis lateral boundary conditions. The spatial distribution of mean onset date progresses from northwest to southeast in all the data sets, in good agreement with previous studies (Janowiak and Arkin 2003; Naylor et al 2007) (Fig. 5). The spatial coherence of interannual variations in onset date, analyzed using EOF analyses, is characterized by a band of high large-scale coherence extending from southwest to northeast in CMAP data (Fig. 6a); this band is also found in the regional models, but is broken up over land in the RegCM-25, where loadings are often noisy and much smaller (Fig. 6b). Nonetheless, the leading PC of onset-date obtained from observed and model datasets are very highly correlated, with more marked delays than advances (Fig. 6c); significant correlations with July SST over the central-equatorial and western Pacific are demonstrated (Fig. 6d).

Domain-averaged August–February evolutions of running 30-day rainfall anomalies confirm the more-pronounced signal in late onsets compared to early ones (Fig. 7a). In the regional model, it is found that the spatial coherence of anomalous advances and (especially) delays in onset are much more pronounced over ocean than they are over island (Figs. 7b–d).

A weather-typing analysis is then applied to daily 850-hPa reanalysis wind fields over Indonesia, using *k*-means cluster analysis. The 5-cluster solution exhibits a sequential progression from WT#1 to WT#5, describing the August–February seasonal evolution (Figs 8,9): “pre-monsoonal” WT#1 and “transitional” WT#2 typify the pre-monsoon season, with dry conditions and easterlies over southern Indonesia, and a shift toward wetter conditions over southern Sumatra in the transitional WT; marked changes take place from mid-October with a progression toward the

monsoonal WTs 3–5, accompanied by widespread westerlies (“westerly” WT#3 and “strong-westerly” WT#5), except for the “quiescent” WT#4 which depicts a notably weaker monsoon (Fig. 9).

Interannual variability is interpreted via the WTs in terms of changes in their frequency of occurrence, with an out-of-phase temporal relationship between the prevalence of the “transitional” and “quiescent” WTs on the one hand, and “westerly” and “strong-westerly” WTs on the other, especially large during the warm ENSO events (Fig. 10a). The seasonal cycle is highly perturbed throughout the season during late-onset years, with the rainy season dominated by the “quiescent” WT (Fig. 10c). In this weak-circulation weather type, rainfall anomalies defined with respect to its mean seasonal evolution are negative over ocean with complex structures over land (Figs. 11, 12); in the 25-km RegCM run, this weather-type is anomalously wet along the mountainous southern and western coast of Java and Sumatra, and is associated with an enhanced diurnal cycle (Figs. 11, 12).

#### *b. Discussion*

Our results illustrate how in both CMAP and GSOD, observed interannual rainfall anomalies weaken after onset, for both early- and late-onset years, consistent with previous studies (Hendon 2003; Giannini et al 2007; Moron et al 2009a). This seasonal weakening of interannual anomalies is reproduced in the RegCM for land points, but is much less pronounced over ocean points. (Fig. 7b–d). While it is difficult to draw firm conclusions from the model runs, these suggest that it is the noisiness of the convection over the complex island topography that destroys the spatial coherence in seasonal post-onset rainfall anomalies, in addition to possible seasonal changes in large-scale ENSO teleconnections (Hendon 2003; Giannini et al 2007). This land-sea contrast may be overstated in the RegCM, which tends to exaggerate land-sea contrasts and the diurnal

land-sea breeze circulation (Qian 2008). Contrasting land-sea behavior was also suggested by the observational study of Hamid et al (2001), who found that while the number of convective storms decreased everywhere during the warm 1997–98 ENSO event, relative to neutral conditions in 1998–99, lightning activity increased over islands, suggesting more intense storms there.

The weather types identified here share similarities with the wind regimes of Ichikawa and Yasunari (2006) and with different phases of the MJO. Most notable is the monsoonal “quiescent” weather type (i.e. WT #4) which tends to coincide with the inactive phase of the MJO (Wheeler and Hendon 2004). Although the inactive phase is associated with large-scale subsidence, and thus anomalous dry conditions generally across Indonesia, this dry footprint becomes fragmented across islands, especially over the orography and on southern and western shores. Ichikawa and Yasunari (2006) indicated also that their corresponding “easterly wind regime”, similar to our “quiescent” weather type, is associated with deeper storms over Borneo and more conditionally unstable conditions.

This “quiescent” weather type is more frequent when the large-scale onset is anomalously late, often in warm ENSO years, and can thus be associated with a large-scale weakening of the monsoon. The RegCM-25 simulations provide evidence that relatively quiescent large-scale winds lead to an enhanced diurnal land-sea breeze, generating increased rainfall over islands, particularly mountains, despite anomalous large-scale subsidence, while the latter promotes negative rainfall anomalies across the seas. While further work is required to examine WT associations with the diurnal cycle, this phenomenon may help to explain the distribution of January–June station rainfall anomalies over Java in warm-ENSO years plotted in Giannini et al (2007, their Fig. 9b), in which positive anomalies occur in the south with negative anomalies on the north shore (compare “quiescent” WT RegCM-25 rainfall anomalies in Fig. 11b here).

## **Acknowledgements**

We are grateful to P. Xie for information on the rain gauge measurements used in the CMAP and to L. Ostwald (IRI) for his decisive help in the extraction and allowance of extra CPU time needed for this research. We thank also both reviewers and the editor (B. Kirtman) who help us to clarify the main goal of our paper. This research was supported by grants from the National Oceanic and Atmospheric Administration (NOAA), NA050AR4311004, the US Agency for International Development's Office of Foreign Disaster Assistance, DFD-A-00-03-00005-00, and the US Department of Energy's Climate Change Prediction Program, DE-FG02-02ER63413.

## References

- Aldrian E, Dwi Susanto R (2003) Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *Int J Climatol* 23: 1435-1452.
- Aldrian E, Dümenil Gates L, Jacob D, Podzun R, Gunawan D (2004) Long-term simulation of Indonesian rainfall with the MPI regional model. *Clim Dyn* 22: 795-814.
- Aldrian E, Sein D, Jacob D, Dümenil Gates L, Podzun R (2005) Modeling Indonesian rainfall with a coupled regional model. *Clim Dyn* 25: 1-17.
- Aldrian E, Dümenil Gates L, Widodo FH (2007) Seasonal variability of Indonesian rainfall in ECHAM4 simulations and in the reanalyses: the role of ENSO. *Theor Appl Climatol* 87: 41-59.
- Chang CP, Wang Z, Ju J, Li T (2004) On the relationship between Western Maritime continent monsoon rainfall and ENSO during Northern Winter. *J Clim* 17: 665-672.
- Chang CP, Wang Z, McBride J, Liu CH (2005) Annual cycle of southeast Asia-Maritime continent rainfall and the asymmetric Monsoon Transition. *J Clim* 18: 287-301.
- Diday E, Simon JC (1976) Clustering analysis. In *Digital Pattern Recognition*, Fu KS (ed.), *Communication and Cybernetics*, vol. 10. Springer-Verlag: 47-94.
- Ebert E, Janowiak JE, Kidd C (2007) Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull Amer Meteor Soc* 47-64.
- Giannini A, Robertson AW, Qian JH (2007) A role for tropical tropospheric temperature adjustment to El Niño Southern Oscillation in the seasonality of monsoonal Indonesia precipitation predictability. *J Geophys Res*: 112: D16110, doi:10.1029/2007JD008519.
- Hackert EC, Hastenrath S (1986) Mechanism of Java rainfall anomalies. *Mon Wea Rev* 114: 745-757.
- Hadi TW, Horinouchi T, Tsuda T, Hashiguchi H, Fukao S (2002) Sea-Breeze circulation over Jakarta, Indonesia: A climatology based on boundary layer radar observations. *Mon Wea Rev* 130: 2153-2166.

Hamada JI, Yamanaka MD, Matsumoto J, Fukao S, Winarso PA, Sribimawati T (2002) Spatial and temporal variations of the rainy season over Indonesia and their link to ENSO. *J Meteo Soc Japan* 80, 285-310.

Hamid YH, Kawasaki ZI, Mardiana R (2001) Impact of the 1997-98 El Niño event on lightning activity over Indonesia. *Geophys Res Letters* 28: 147-150.

Harger JRE (1995) ENSO variations and drought occurrence in Indonesia and the Philippines, *Atmos Env* 29: 1943-1955.

Haylock M, McBride J (2001) Spatial coherence and Predictability of Indonesian wet season rainfall. *J Clim* 14: 3882-3887.

Hendon HH (2003) Indonesian rainfall variability: impacts of ENSO and local Air-Sea interaction. *J Clim* 16: 1775-1790.

Ichikawa H, Yasunari T (2006) Time-space characteristics of diurnal rainfall over Borneo and surrounding oceans as observed by TRMM-PR. *J Clim* 19: 1238-1260.

Janicot S, Moron V, Fontaine B (1996) Sahel droughts and ENSO. *Geophys Res Letters* 23: 515-518

Janowiak JE, Xie P (2003) A Global-scale examination of monsoon-related precipitation. *J Clim* 16: 4121-4133.

Joyce RE, Janowiak JE, Arkin PA, Xie P (2004) CMORPH: a method that produces global precipitation estimates from passive and infrared data at high spatial and temporal resolution. *J. Hydrometeor* 5: 487-503.

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen G, Zhu Y, Leetmaa A, Reynolds R, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Jenne R, Joseph D (1996) the NCEP/NCAR 40 year reanalysis project, *Bull Amer Meteo Soc* 77: 437-471.

Kiladis GN, Diaz HF (1989) Global climatic anomalies associated with extremes of the southern oscillation. *J Clim* 2: 1069-1090.

Klein SA, Soden BJ, Lau NG (1999) Remote Sea Surface Temperature Variations during ENSO: Evidence for a Tropical Atmospheric Bridge. *J Clim* 12: 917-932.

Kodama YM, Tokuda M, Murata F (2006) Convective activity over the Indonesia maritime continent during CPEA-I as evaluated by lightning activity and Q1 and Q2 profiles. *J Meteo Soc Japan* 84A: 133-149.

McBride JL, Haylock MR, Nicholls N (2003) Relationships between the Maritime continent heat source and the El Niño Southern Oscillation phenomenon. *J Clim* 16: 2905-2914.

Marteau R, Moron V, Philippon N (2009) Spatial coherence of boreal monsoon onset across central and western Sahel. *J Clim*: in press.

Michelangeli PA, Legras B, Vautard R (1995) Weather regimes: recurrence and quasi-stationnarity. *J Atmos Sci* 52: 1237-1256.

Mori S, Hamada JI, Tauhid YI, Yamanaka MD, Okamoto N, Murata F, Sakurai N, Hashiguchi H, Sribimawati T (2004) Diurnal land-sea rainfall peak migration over Sumatra Island, Indonesian maritime continent, observed by TRMM Satellite and intensive rawinsonde soundings. *Mon Wea Rev* 132: 2021-2039.

Moron V, Robertson AW, Ward MN (2006) Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Mon Wea Rev* 134: 3468-3482.

Moron V, Robertson AW, Ward MN, Camberlin P (2007) Spatial coherence of tropical rainfall at regional scale, *J Clim* 20: 5244-5263.

Moron V, Robertson AW, Ward MN, and Ndiaye O (2008) Weather types and rainfall over Senegal. Part I : observational analysis. *J Clim* 21: 266-287.

Moron V, Robertson AW, and Boer R (2009a) Spatial coherence and seasonal predictability of monsoon onset over Indonesia, *J Clim*: in press.

Moron V, Lucero A, Hilario F, Lyon B, Robertson AW, DeWitt D (2009b) Spatio-temporal variability and predictability of summer monsoon onset over the Philippines, *Clim Dyn*: in press.

Naylor RL, Falcon WP, Rochberg D, Wada N (2001) Using El-Niño Southern Oscillation climate data to predict rice production in Indonesia. *Clim Change* 50: 255-265.

Naylor RL, Falcon W, Wada N, Rochberg D (2002) Using El-Niño Southern Oscillation climate data to improve food policy planning in Indonesia. *Bull Indonesian Eco Studies* 38: 75-91.

Naylor RL, Battisti DS, Vimont DJ, Falcon WP, Burke MB (2007) Assessing the risks of climate variability and climate change for Indonesian rice agriculture, *Proc Natl Acad Sci* 104: 7752-7757.

Omotosho JB (1992) Long-range prediction of the onset and end of the rainy season in the West African Sahel. *Int J Climatol* 12: 369-382.

Pal JS, and co-authors (2007) Regional climate modelling for the developing world: the ICTP RegCM3 and RegCNET. *Bull Amer Meteor Soc*: 88: 1395-1409.

Qian, JH (2008) Why precipitation is mostly concentrated over Islands in the Maritime continent? *J Atmos Sci* 65: 1428-1441.

Robertson AW, Mechoso CR (2003) Circulation regimes and low-frequency oscillations in South Pacific sector. *Mon Wea Rev* 131: 1566-1576.

Robertson AW, Moron V, Swarinoto Y (2008) Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia. *Int J Climatol*, in press.

Ropelewski CF, Halpert MS (1987) Global and regional scale precipitations associated with El Niño/Southern Oscillation. *Mon Wea Rev*: 115: 985-996.

Ropelewski CF, Halpert MS (1996) Quantifying Southern Oscillation - Precipitation relationships. *J Clim* 9: 1043-1059.

Sakurai N, Murata F, Yamanaka MD, Mori S, Hamada JI, Hashiguchi H, Tauhid YI, Sribimawati T, Suhardi B (2005) Diurnal cycle of cloud system migration over Sumatra Island. *J Meteo Soc Japan* 83: 835-850.

Shibagaki Y, Shimonai T, Kozu T, Mori S, Fujiyoshi Y, Hashiguchi H, Yamamoto MK, Fukao S, Yamanaka M (2006) Multiscale aspects of convective systems associated with an intraseasonal oscillation over the Indonesian Maritime Continent. *Mon Wea Rev* 134: 1682-1696.

Sivakumar MVK (1988) Predicting rainy season potential from the onset of rains in southern sahelian and sudanian climatic zones of West Africa. *Agric For Meteo* 42: 295-305.

Sperber KR (2003) Propagation and the vertical structure of the Madden-Julian oscillation. *Mon Wea Rev* 131: 3018-3037.

Stern RD, Dennett MD, Garbutt DJ (1981) The start of the rains in West Africa. *Int J Climatol* 1: 59-68.

Vautard R, Mo K, Ghil M (1990) Statistical significance test for transition matrices of atmospheric Markov Chains. *J Atmos Sci* 47: 1926-1931.

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

Wheeler MC, McBride J (2005) Australian-Indonesian monsoon, in Lau WKM and Waliser DE (eds) *Intraseasonal variability in the Atmosphere-Ocean climate system*, Springer, 125-173.

Wilks DS (1999) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agric For Meteo* 93: 153-169.

Wu P, Hamada JI, Mori S, Tauhid YI, Yamanaka MD (2003) Diurnal variation of precipitable water over a mountainous area of Sumatra island. *J Appl Meteo* 42: 1107-1115.

Xie P, Arkin PA (1997) Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bull Amer Meteo Soc* 78: 2539-2558.

Xie P, Arkin PA (1998) Global monthly precipitations estimates from satellite-observed outgoing longwave radiation. *J Clim* 11: 137-164.

## Figure captions

**Figure 1:** Location of the CMAP grid-boxes (dots) and of the stations having daily observations in GSOD dataset (circles), with the percentages of missing entries (indicated as proportional circles) for Indonesia. Red dots indicate the location of Jatiwangi and Tegal.

**Figure 2:** Daily rainfall records at Jatiwangi (left) and Tegal (right) on Java island in 1997–98 (top) and 2003–04 (bottom). The vertical dashed lines indicate the first day of each month. The full and dashed arrows indicate monsoon onset and false-starts respectively (see text). The rainfall totals for the rainy season are shown in the upper right corner of each panel. There are less than 10 missing days in each case.

**Figure 3:** Topography (in m) in (a) observation (at ~2.5 km resolution), (b) RegCM-60 and in (c) RegCM-25.

**Figure 4:** Mean rainfall (in mm) for 5-day wet sequences (i.e. at least 1 mm) during August–February in (a) CMAP (shading) and GSOD (circles), (b) CMORPH, (c) RegCM-60, and (d) RegCM-25.

**Figure 5:** Mean calendar date of monsoon onset in (a) CMAP (shading) and GSOD (circles), (b) CMORPH, (c) RegCM-60 and (d) RegCM-25.

**Figure 6:** Leading EOF of onset date in (a) CMAP (shading) and GSOD (circles), and (b) RegCM-25, together with (c) associated PC time series, and (d) SST indices. The colors in (a) and (b) express the EOF loadings as correlations between the standardized PC time series and the individual station or grid-point. Panel (c) shows the standardized leading PC of GSOD (full black line), CMAP (full red line and circle), RegCM-60 (full green line and circle) and RegCM-25 (full blue line and circle). Panel (d) shows two SST anomaly indices (Central Pacific (CP): 170°W–120°W, 4°S–4°N in red, and Western Pacific (WP): 100°–130°E, 8°S–0° in black, for July (full line) and October (dashed line) (°C).

**Figure 7:** (a) Evolution of running 30-day spatially-averaged rainfall mean anomaly in CMAP (dashed line with circle) and GSOD (full line with square) for early onset years (in black) and late onset years (in red) relatively to the long-term mean. The early (respectively late) onset years are defined as the lower (respectively higher) quartiles of the average of the standardized leading principal component of onset dates (i.e. 1980, 1988, 1995, 1996, 1998, 1999) (respectively i.e. 1982, 1987, 1991, 1994, 1997, 2002). The filled markers indicate a spatial average that is significantly different from the long-term mean at the two-sided 90% level of significance according to a Student's T test. (b) same as (a) for the early and late onset years in RegCM-25. (c) Same as (b) for islands south of the equator only. (d) Same as (b) for seas south of the equator only.

**Figure 8:** (a) Classifiability index of the k means solution with k=2 to k=10 clusters. 500 random seeds are used each time. (b) Smoothed seasonal mean cycle of the 5 weather types (frequency is computed on running 30-day periods). The calendar dates in abscissa are centre of 30-day periods.

**Figure 9:** Composite maps of rainfall (in mm/day) in CMAP (shadings) and GSOD (filled circle) together with reanalysis winds at 850 hPa for each weather type. The composites were constructed by averaging the respective raw fields over the days assigned to each weather type, without subtracting out the climatological mean. In the text, the five WTs are referred to respectively as “pre-monsoonal”, “transitional”, “westerly”, “quiescent”, and “strong-westerly”.

**Figure 10:** (a) Interannual variability of the frequency of each weather type. (b) running 30-day mean frequency during (b) early and (c) late onset years. The calendar dates in abscissa are centre of 30-day periods. Large colored (small black) dots indicate statistically significant positive (negative) anomalies in WT frequency, compared to neutral years at the two-sided 90% level, according to a Chi-2 test ( $H_0$ : WT frequency during early/late onset years equals that in neutral years).

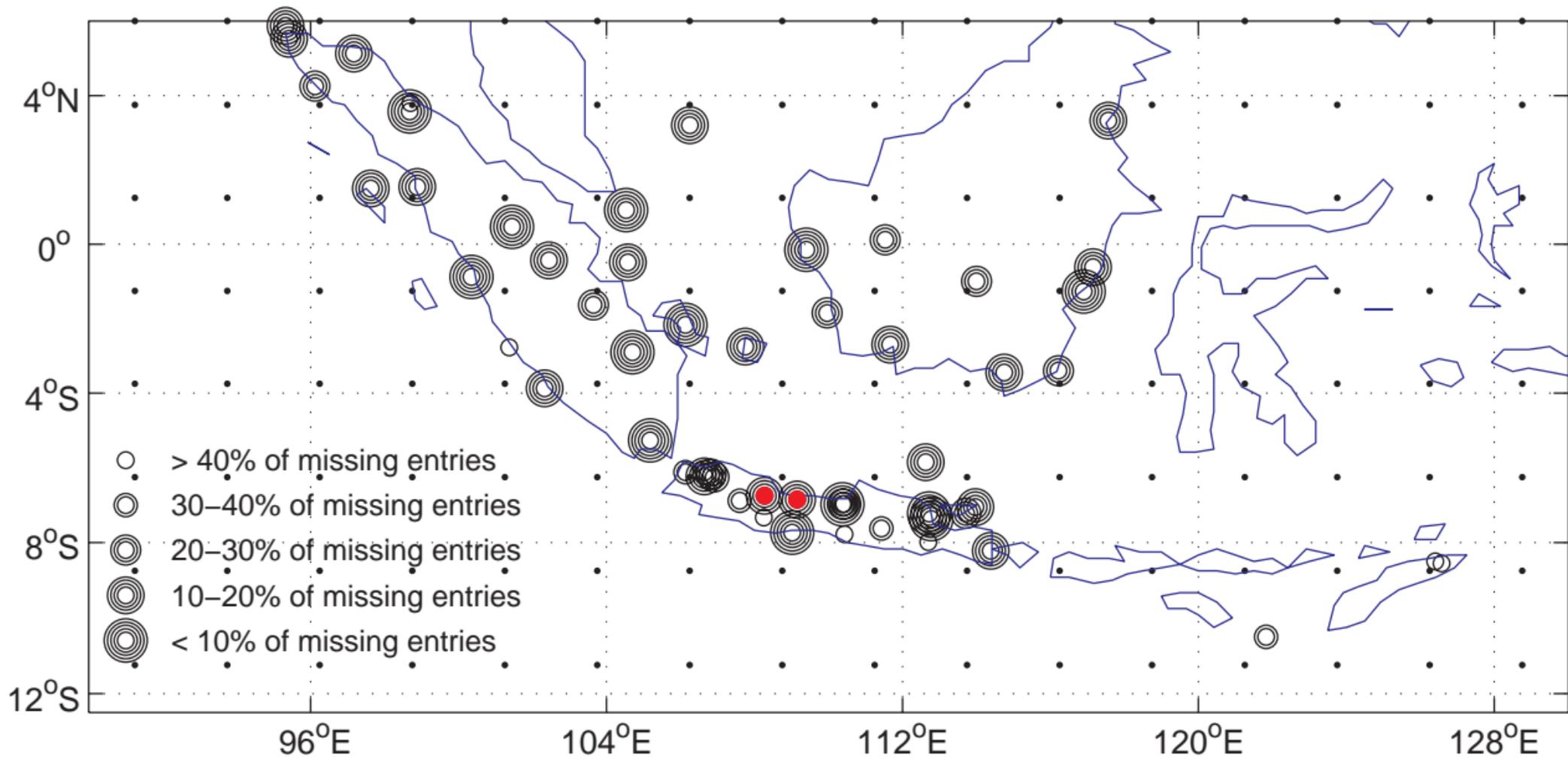
**Figure 11:** Mean rainfall anomalies in GSOD (circles) and RegCM-25 (shading) for (a) WT #3, (b) WT #4, (c) WT #5, relatively to the mean seasonal cycle. The mean seasonal cycle of each station is weighted by the temporal frequency of each weather type to get the mean rainfall associated with the seasonal cycle of each weather type and this is subtracted from the mean rainfall of each weather type for obtaining the anomaly expressed as deviations (in percentage) from the mean.

**Figure 12:** Raw mean rainfall and near surface wind at 4-7 pm local time for (a) WT#3, (c) WT#4 and (e) WT#5 and mean rainfall and near surface wind anomalies relative to the seasonal cycle at 4–7 pm local time for (b) WT#3, (d) WT#4 and (f) WT#5. Rainfall Anomalies are computed and expressed as in Fig. 11. Wind anomalies are computed in the same way except that they are expressed in m/s relatively to the seasonal mean.

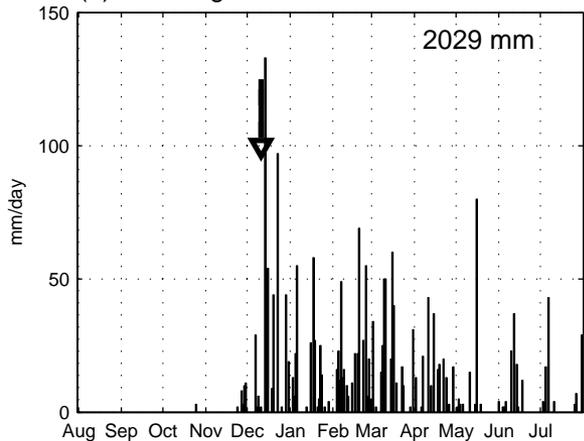
**Table**

	NINO	IOD	WP	ONSET
WT#1	43**	-8	-22	22
WT#2	13	45***	-28*	27
WT#3	-67***	-48***	54***	-72***
WT#4	61***	37**	-51***	79***
WT#5	-39**	-16	39*	-56**
NINO	100	22	-38**	72***
IOD	22	100	-81***	63***
WP	-38**	-81***	100	-72***

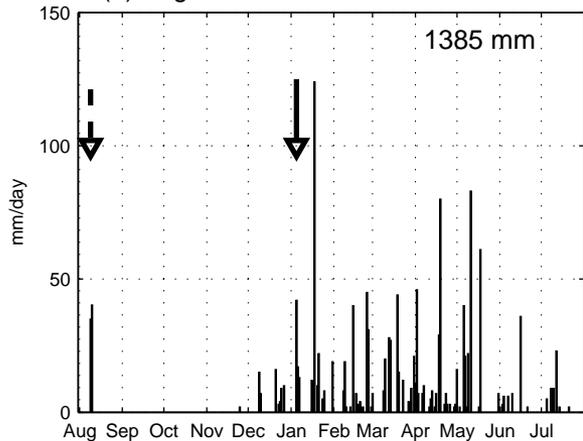
**Table 1** : Cross-correlation (x 100) between the average of standardized leading principal component of onset date (= ONSET), the SST averaged over Niño 3.4 (170°E-240°E, 5°S-5°N), Indian Ocean Dipole Index (= difference between 50°-70°E, 10°S-10°N minus 90°-110°E, 10°S-Eq.) and western Pacific (100°-130°E, 8°S-0°) in July, i.e. 1-4 months before the median local-scale onset across Indonesia, and the seasonal frequency (in ASONDJF) of each weather types. One, two, and three asterisks show the correlations significant at the 90%, 95% and 99% two-sided level according to a random-phase test (Janicot et al 1996).



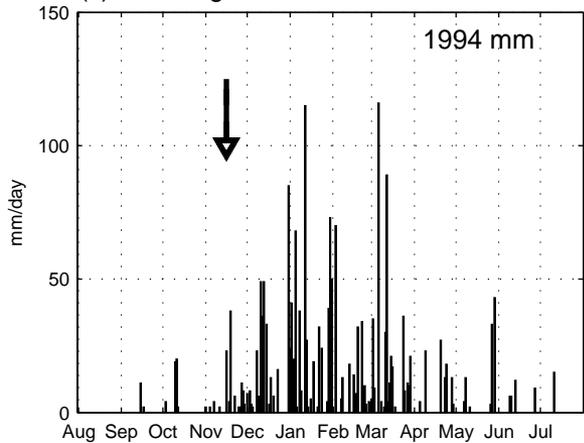
(a) Jatiwangi 108.27°E–6.75°S in 1997–98



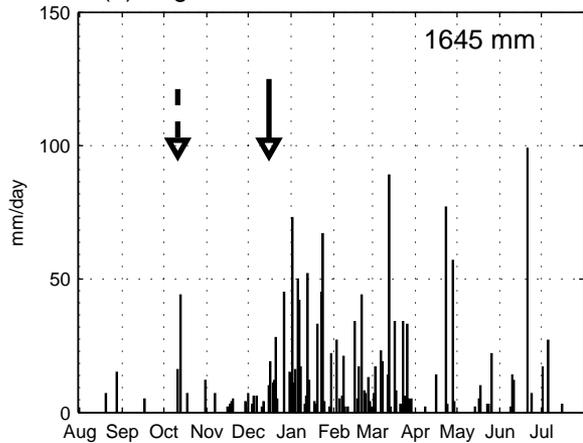
(b) Tegal 109.15°E–6.85°S in 1997–98



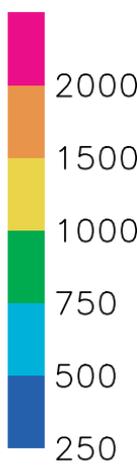
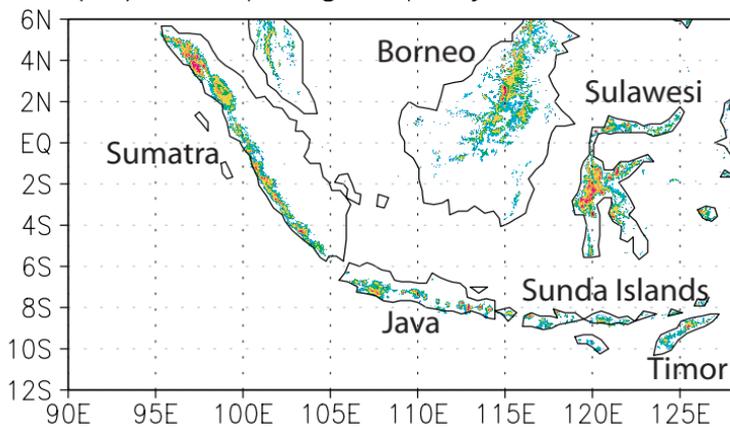
(c) Jatiwangi 108.27°E–6.75°S in 2003–04



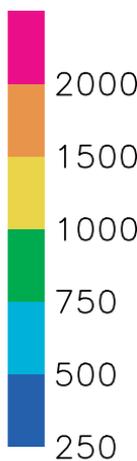
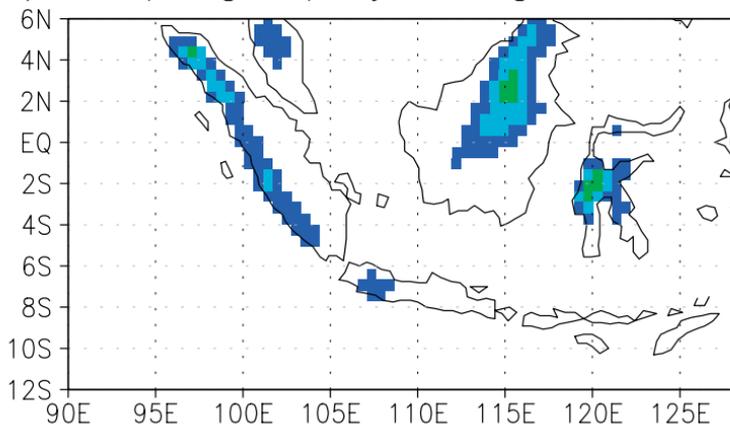
(d) Tegal 109.15°E–6.85°S in 2003–04



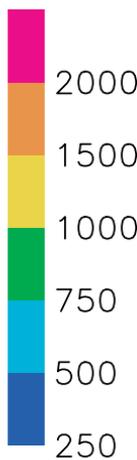
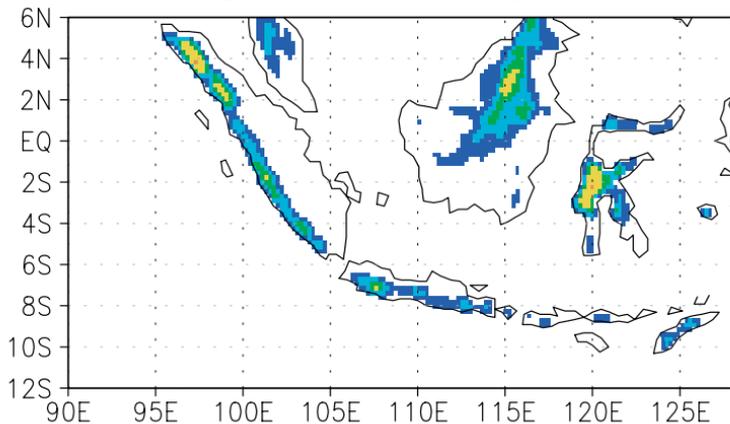
(a) Topography 2.5 km



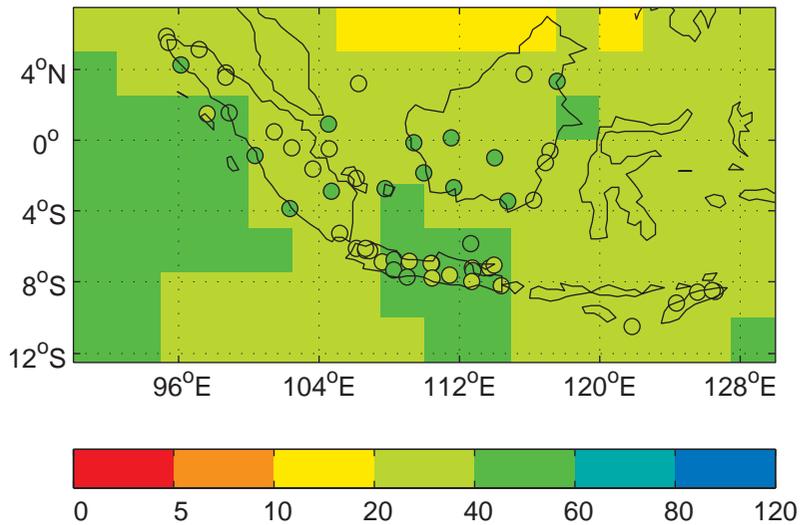
(b) Topography RegCM 60 km



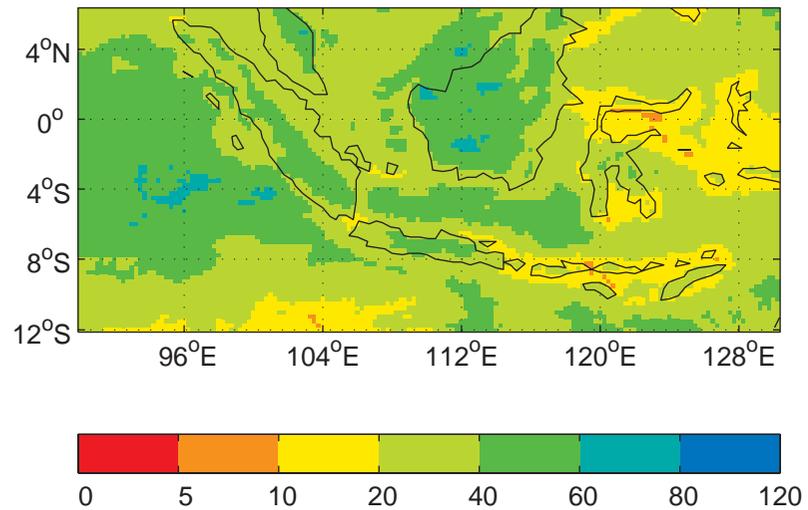
(c) Topography RegCM 25 km



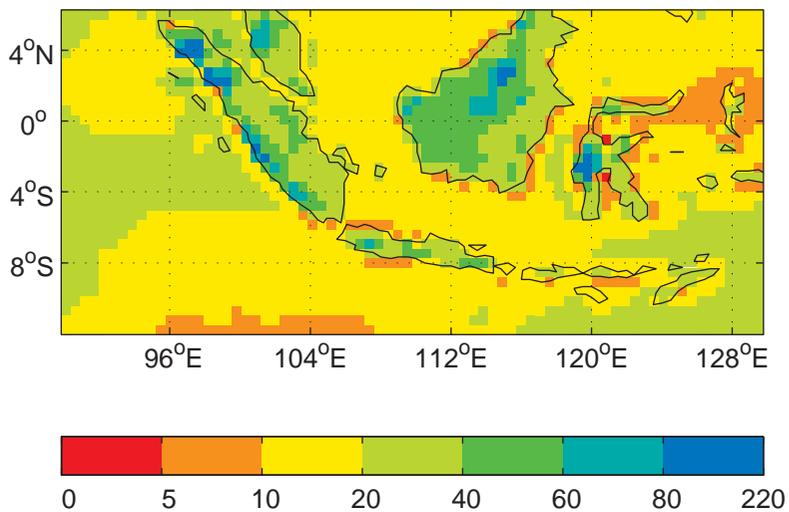
(a) CMAP & GSOD



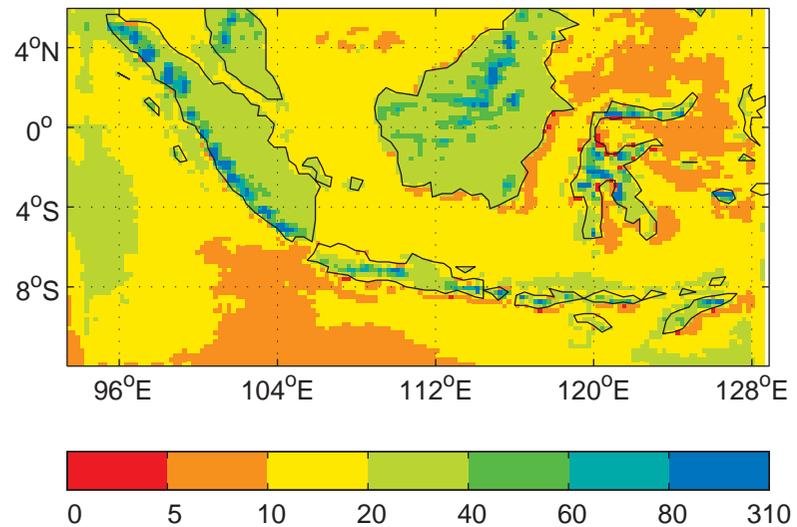
(b) CMORPH



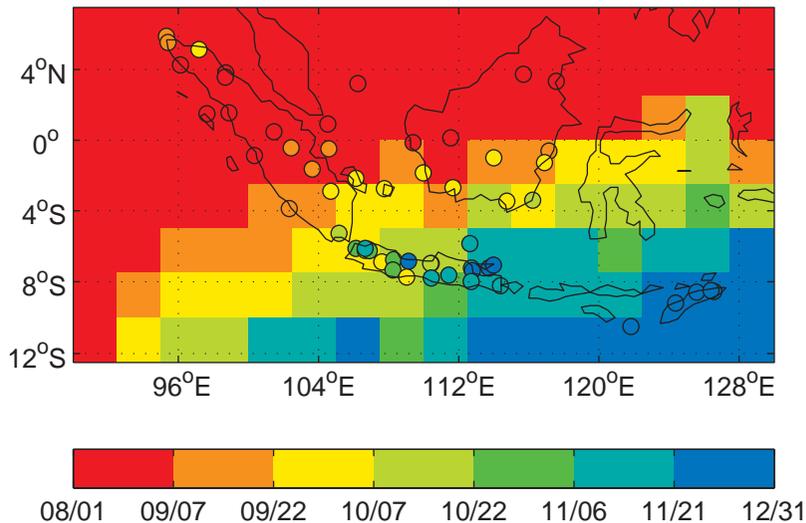
(c) RegCM 60 km



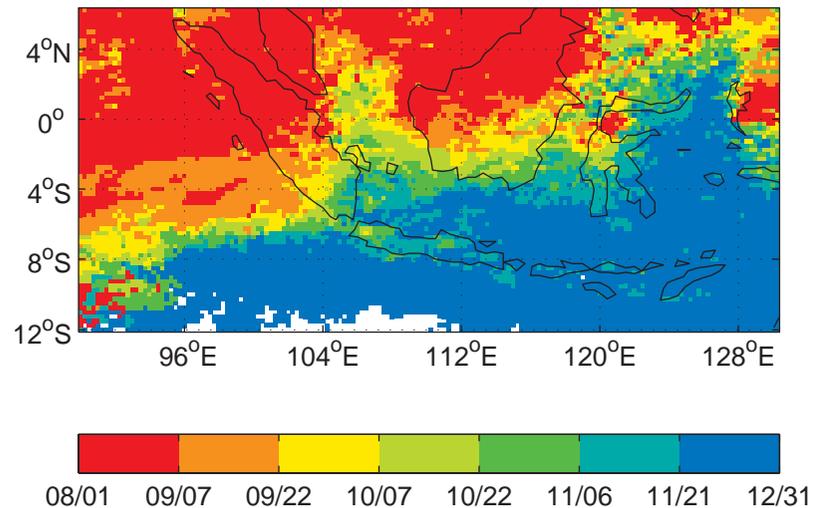
(d) RegCM 25 km



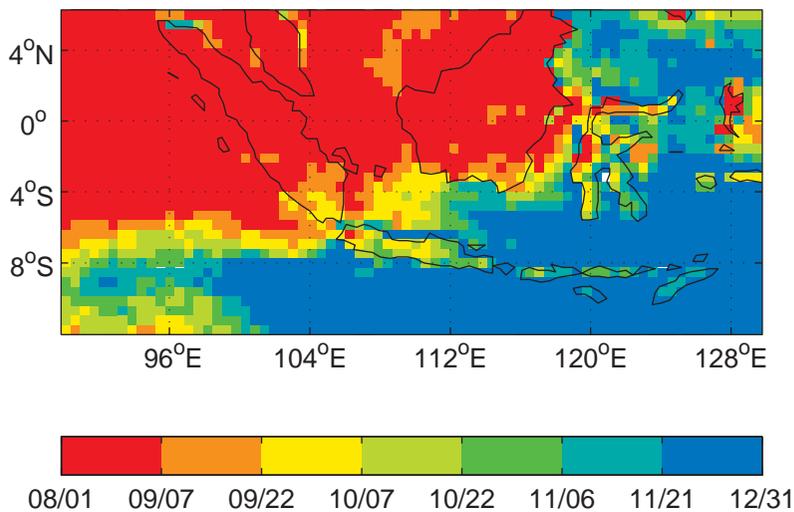
(a) CMAP & GSOD



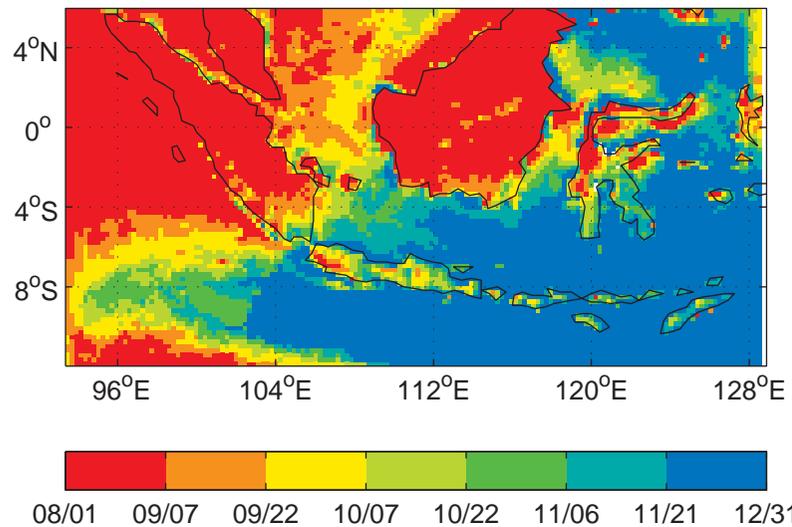
(b) CMORPH



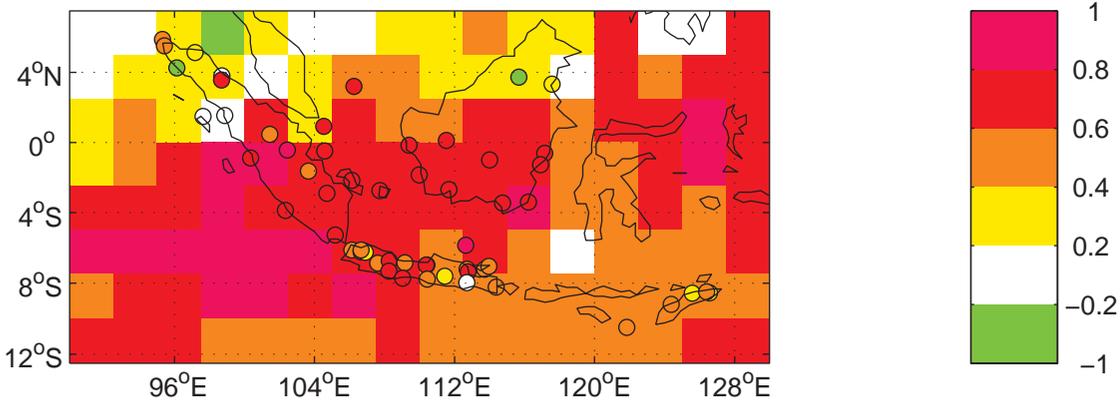
(c) RegCM 60 km



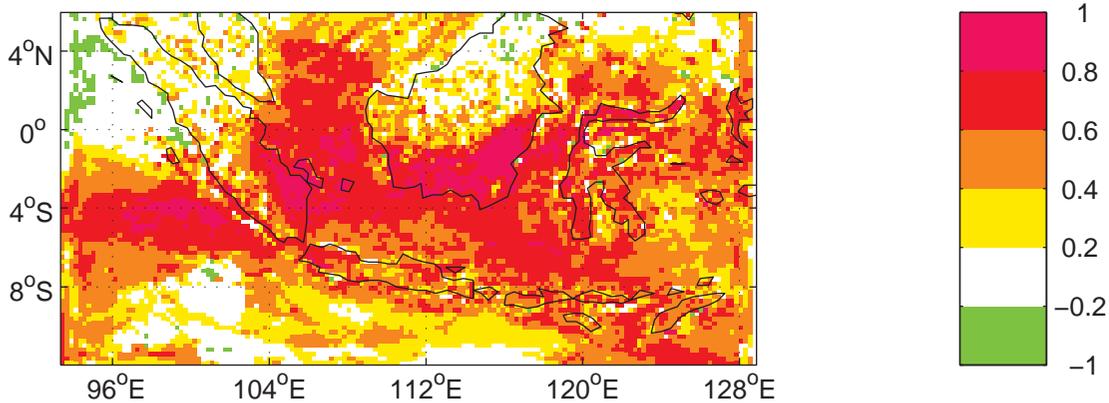
(d) RegCM 25 km



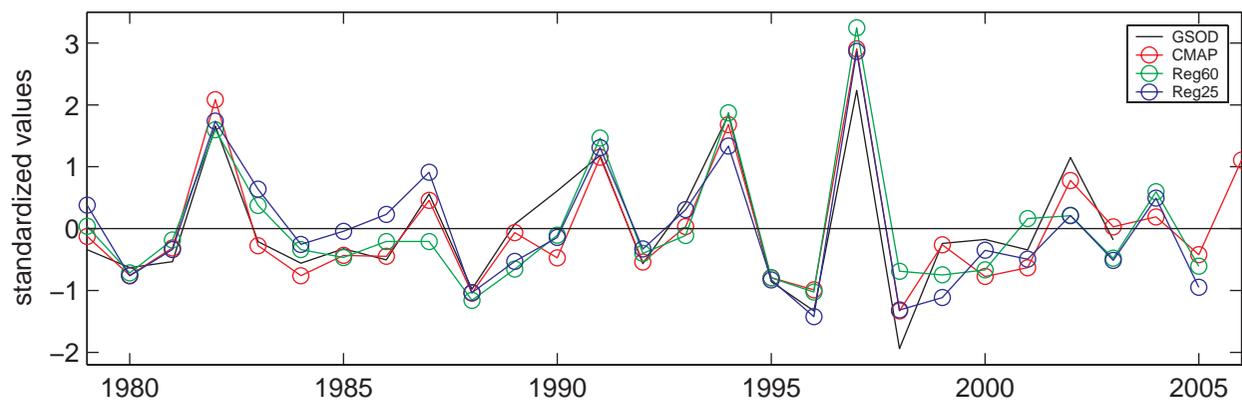
(a) CMAP & GSOD



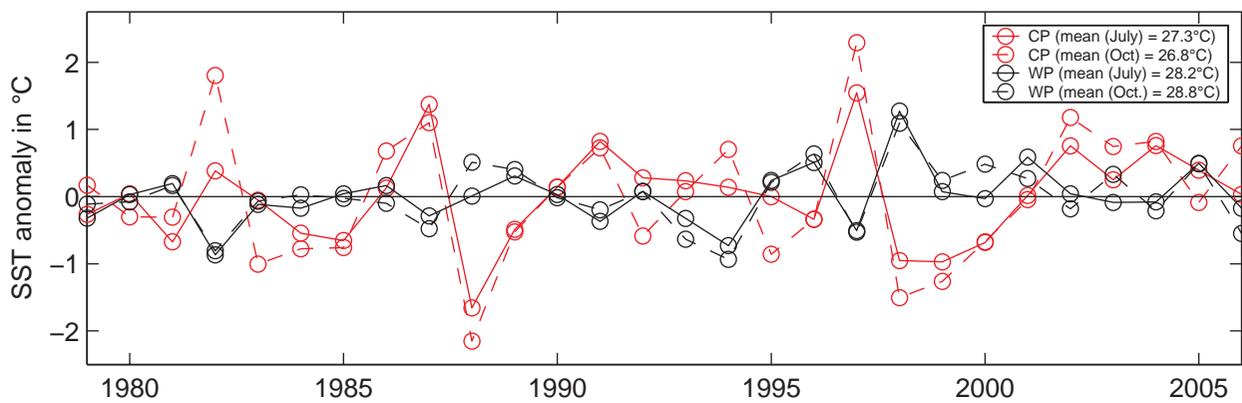
(b) Reg 25 km



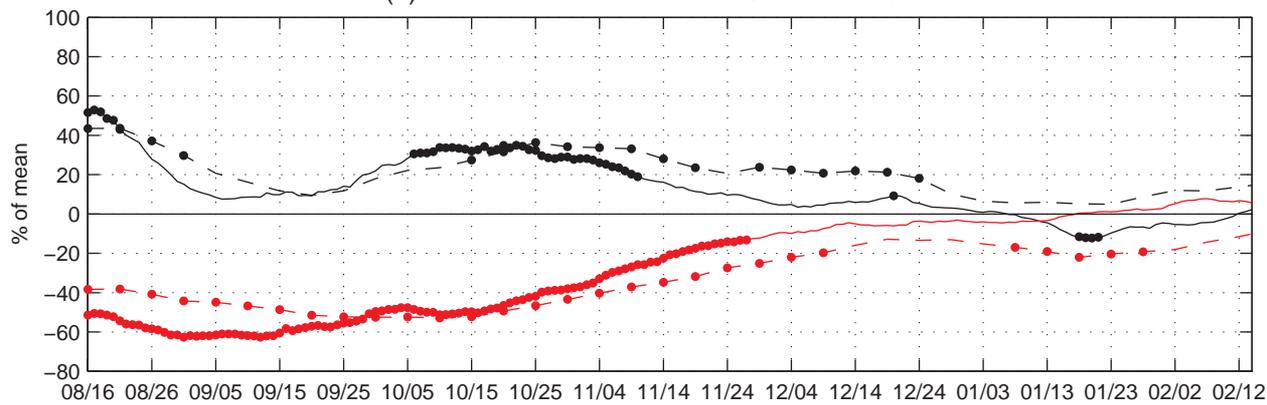
(c) Principal Component



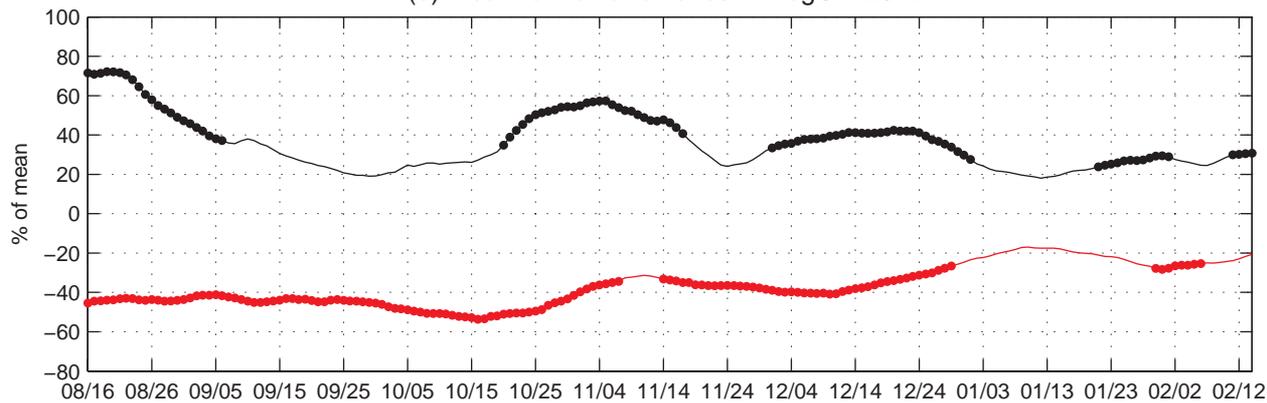
(d) SST in Central and Western Pacific



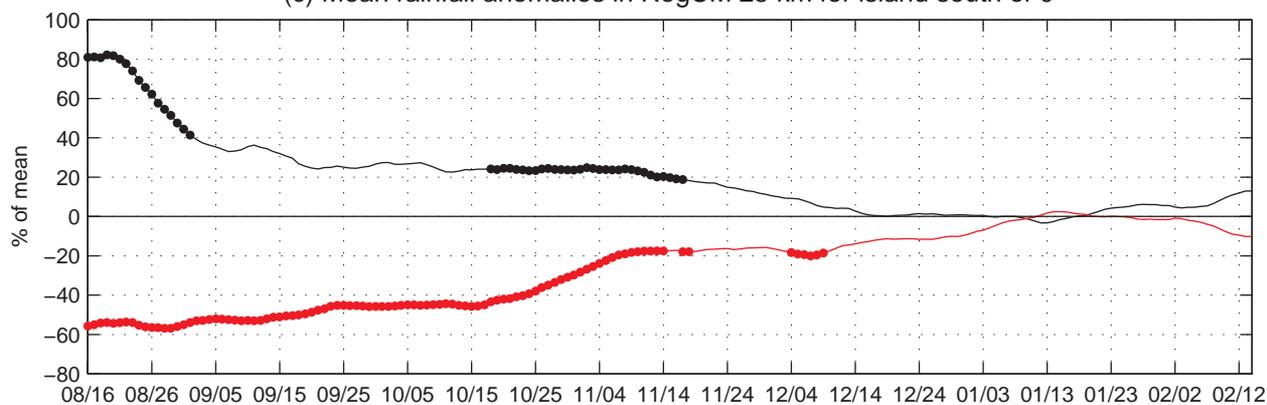
(a) Mean rainfall anomalies in CMAP and GSOD



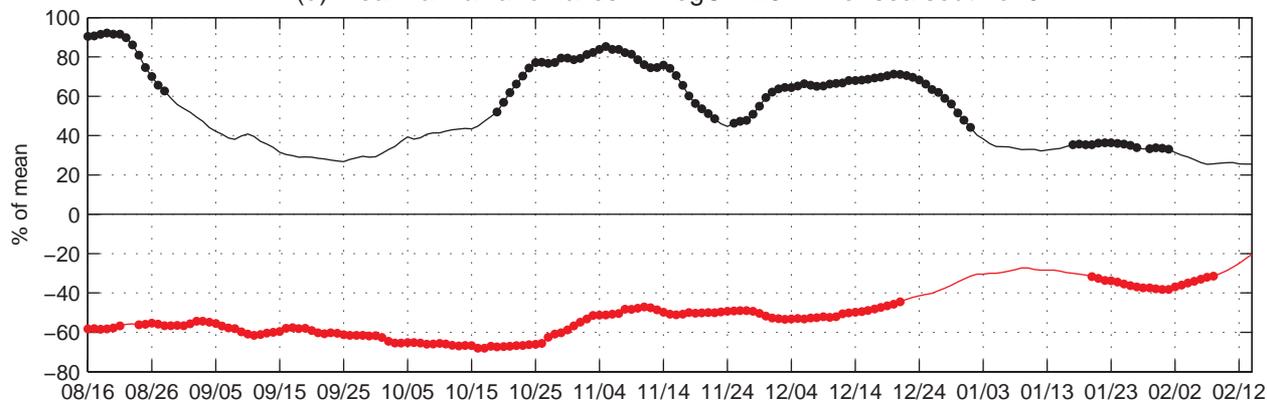
(b) Mean rainfall anomalies in RegCM 25 km



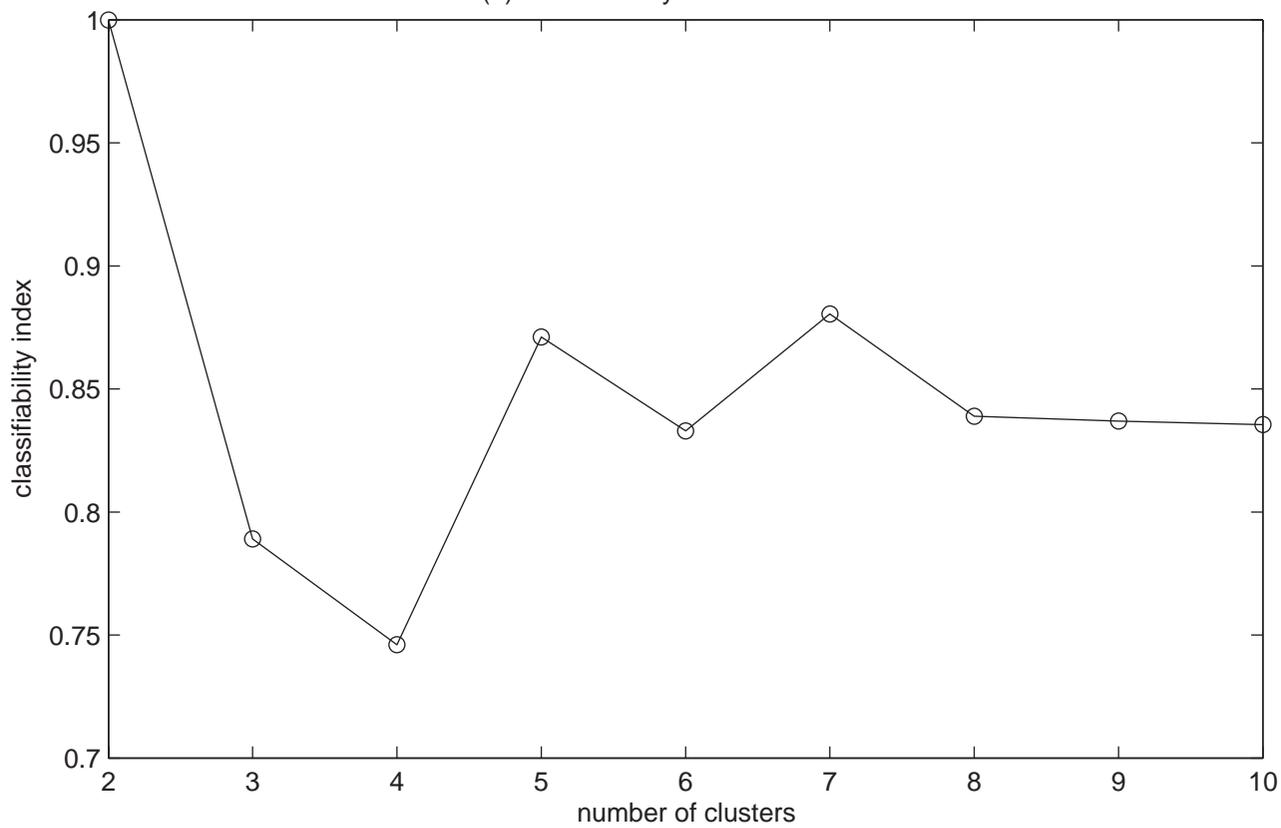
(c) Mean rainfall anomalies in RegCM 25 km for island south of 0°



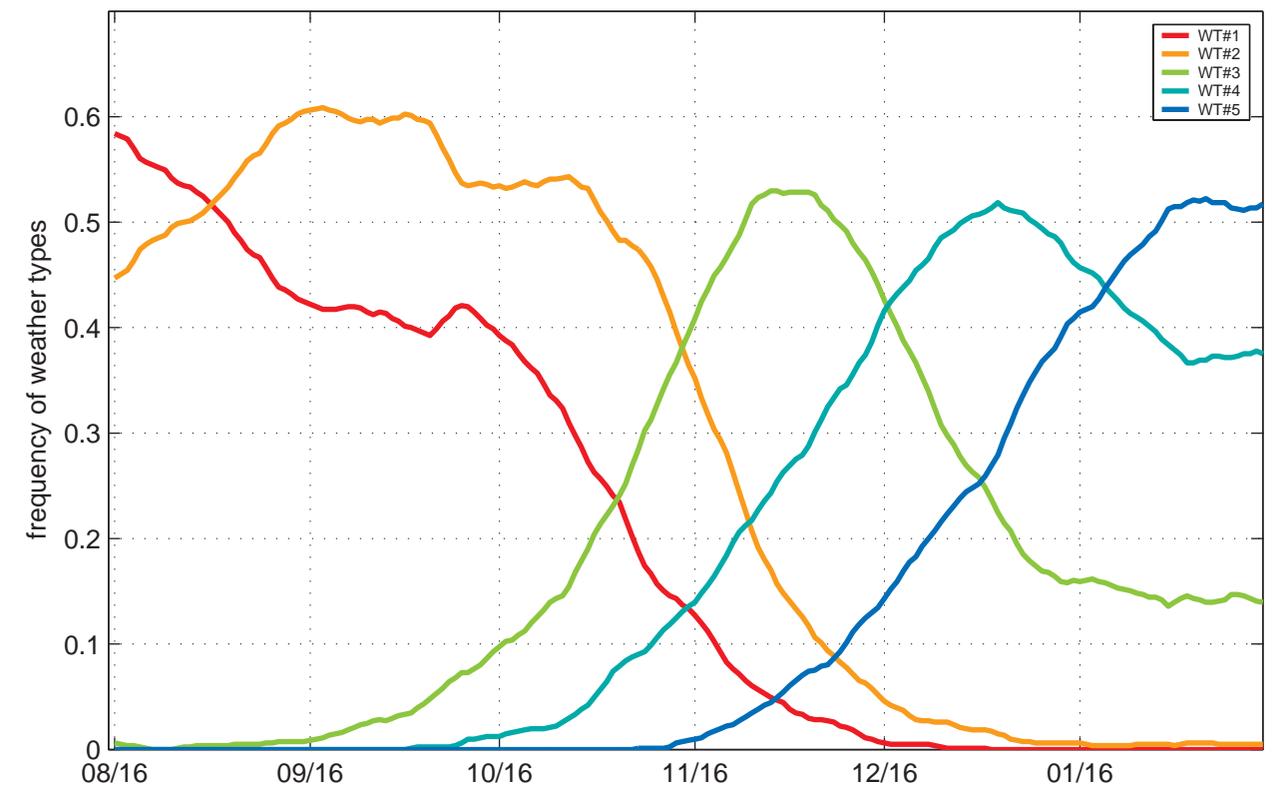
(d) Mean rainfall anomalies in RegCM 25 km for sea south of 0°



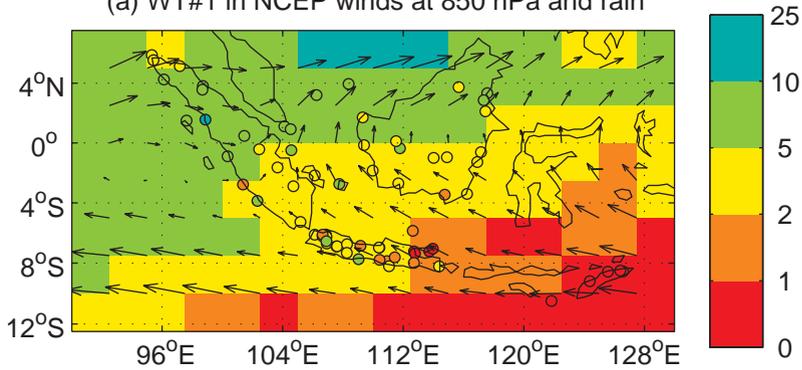
(a) Classifiability Index of k-means



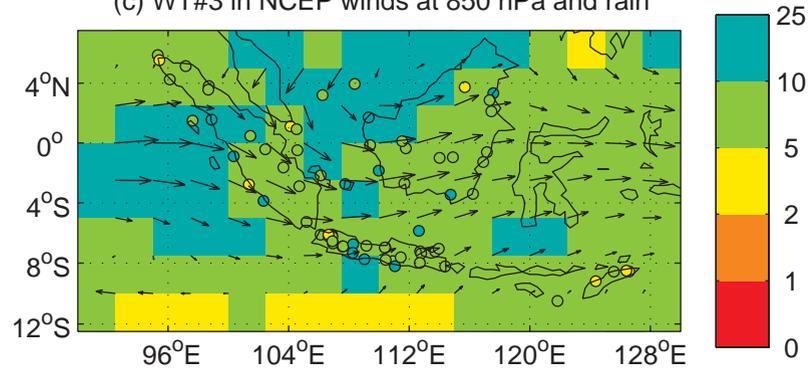
(b) Climatology



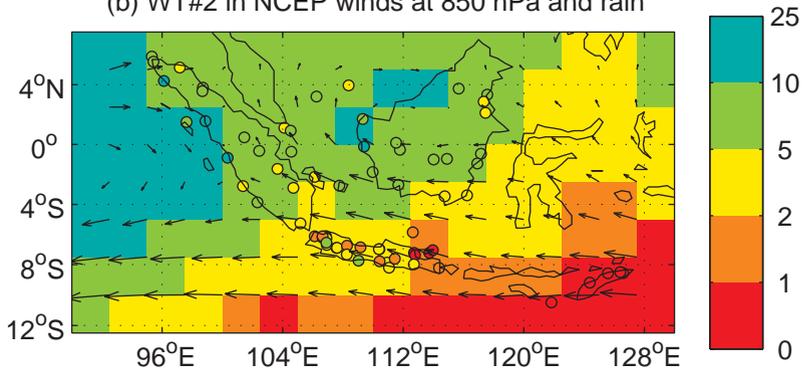
(a) WT#1 in NCEP winds at 850 hPa and rain



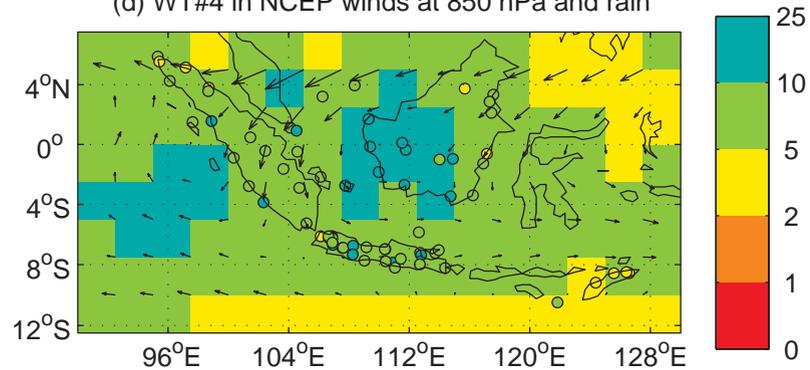
(c) WT#3 in NCEP winds at 850 hPa and rain



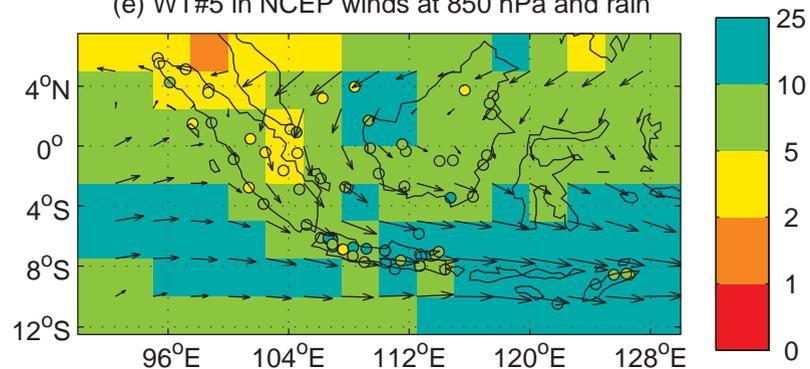
(b) WT#2 in NCEP winds at 850 hPa and rain



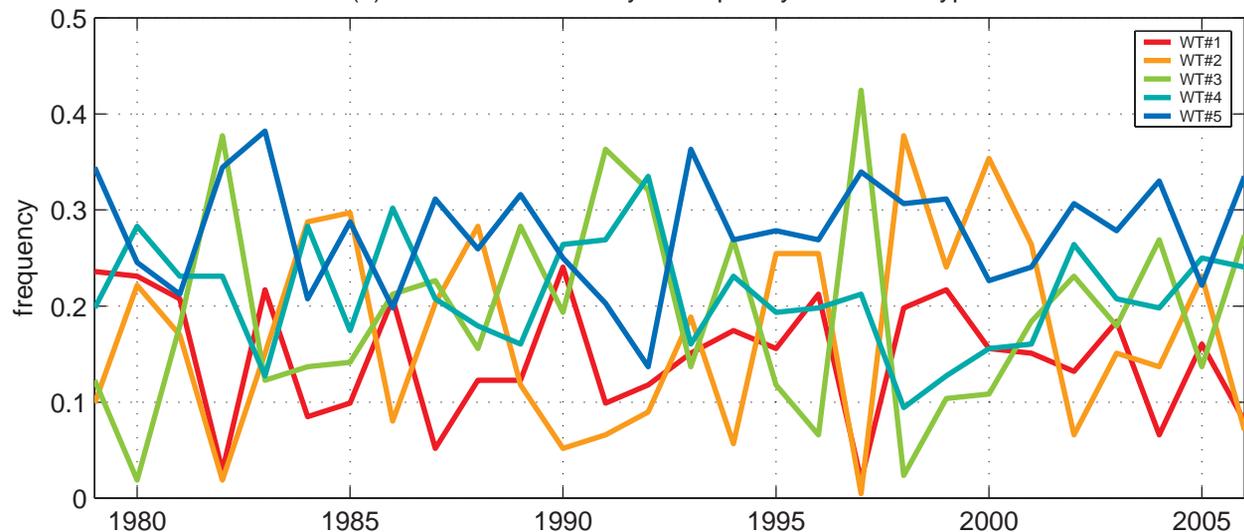
(d) WT#4 in NCEP winds at 850 hPa and rain



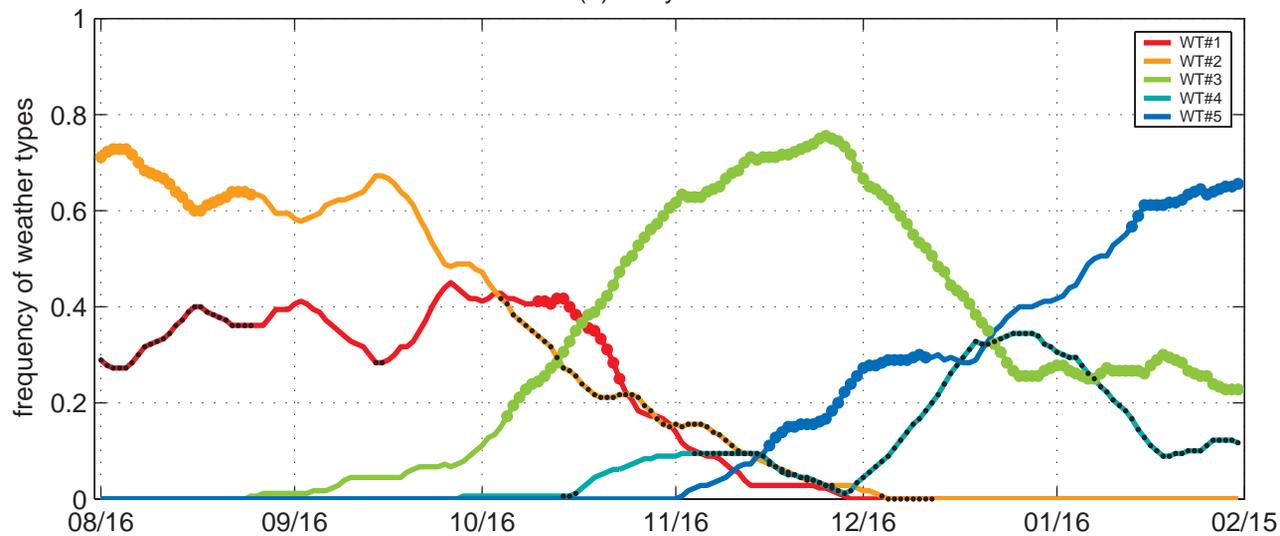
(e) WT#5 in NCEP winds at 850 hPa and rain



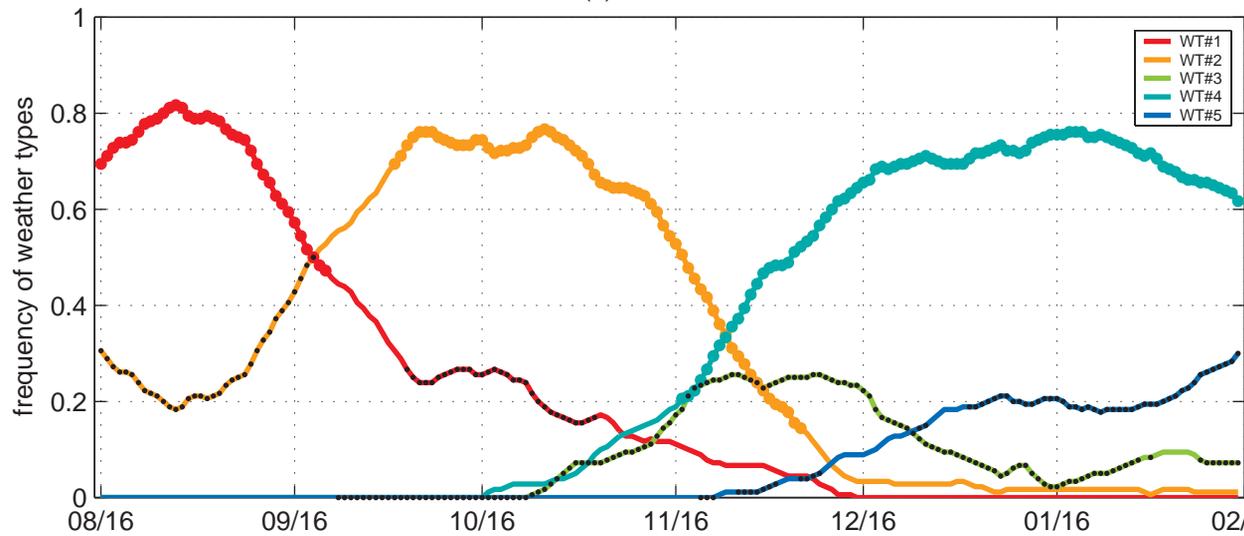
(a) Interannual variability of frequency of weather types



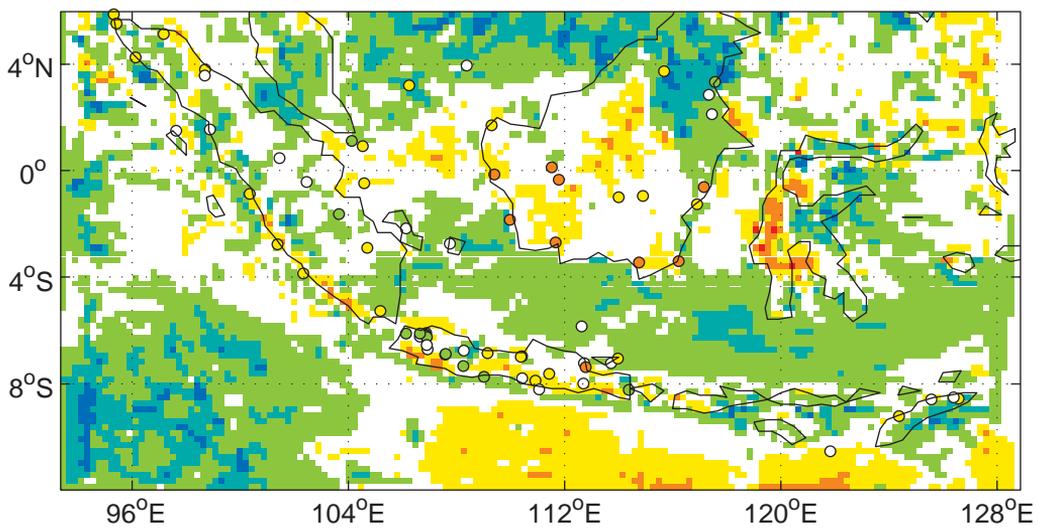
(b) Early onsets



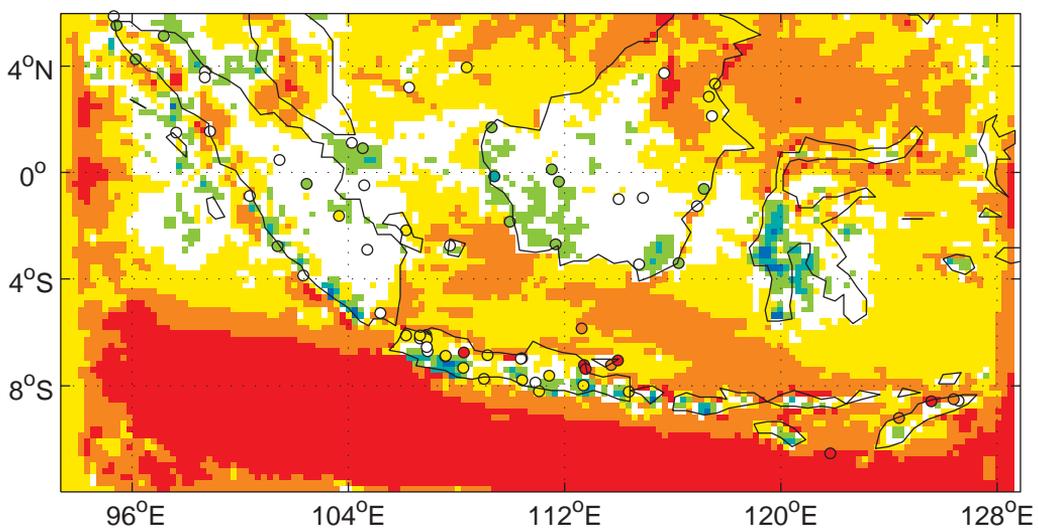
(c) Late onsets



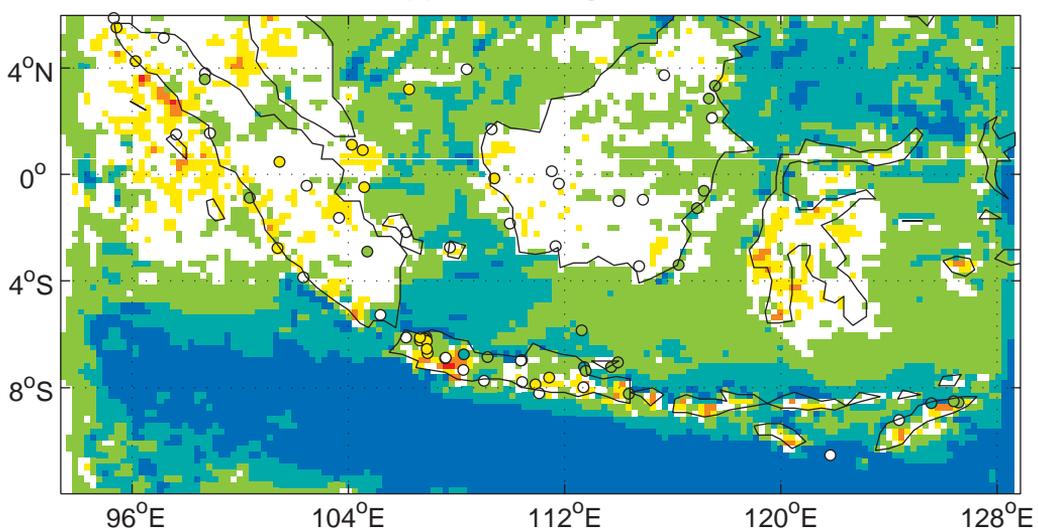
(a) WT#3 in Reg 25 km



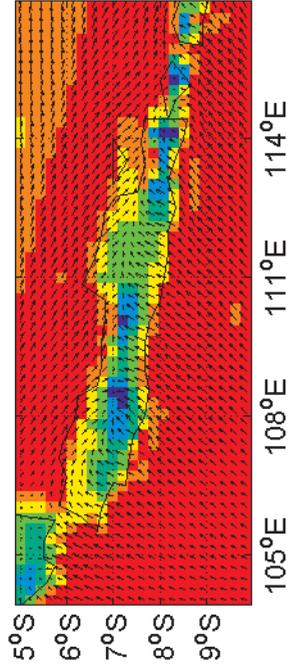
(b) WT#4 in Reg 25 km



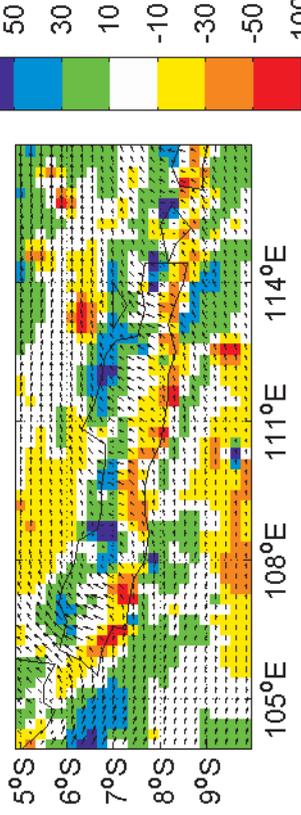
(c) WT#5 in Reg 25 km



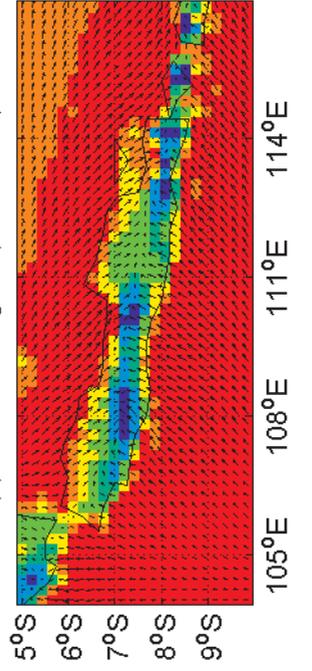
(a) WT#3 in Reg-25 (16-19 LST)



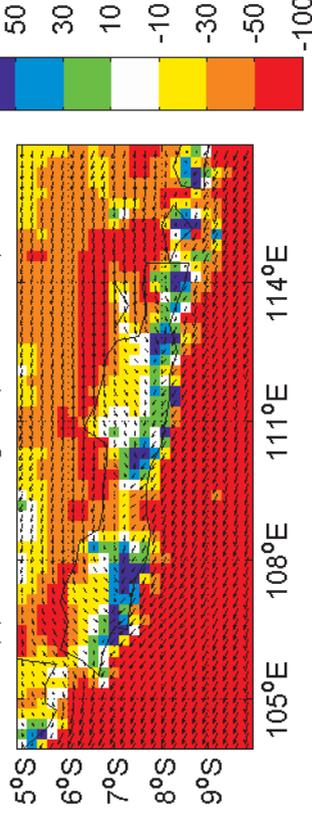
(b) WT#3 in Reg-25 (16-19 LST)



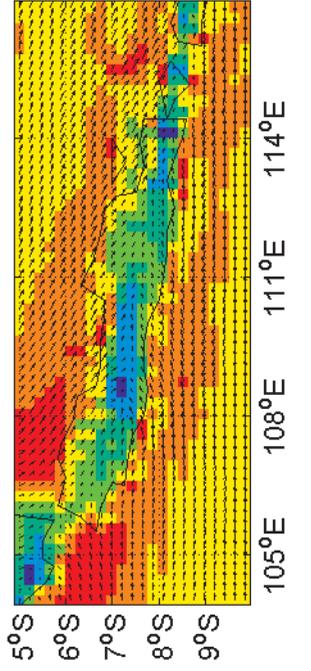
(c) WT#4 in Reg-25 (16-19 LST)



(d) WT#4 in Reg-25 (16-19 LST)



(e) WT#5 in Reg-25 (16-19 LST)



(f) WT#5 in Reg-25 (16-19 LST)

