Spatial coherence and seasonal predictability of monsoon onset over Indonesia

Vincent Moron*,°[1] Andrew W. Robertson°, Rizaldi Boer+

* CEREGE, UMR 6635 CNRS, Université d'Aix-Marseille, & Institut Universitaire de France, France

° International Research Institute for Climate and Society (IRI), Columbia University, Palisades, New York, USA

+ Laboratory of Climatology, Bogor Agricultural University, Bogor, Indonesia

submitted to *Journal of Climate*

revised version

May 2008

[1] Corresponding author address : Vincent Moron, CEREGE, UMR 6635 CNRS & Université d'Aix-Marseille, Europôle de l'Arbois, BP 80, 13545 Aix en Provence, France. (authors' emails : VM (moron@cerege.fr), AWR (awr@iri.columbia.edu), RZ (rizaldiboer@gmail.com)

Abstract

The seasonal potential predictability of monsoon onset during the August–December season over Indonesia is studied through analysis of the spatial coherence of daily station rainfall and gridded pentad precipitation data from 1979 to 2005. Onset date, defined using a local agronomic definition, exhibits a seasonal northwest-to-southeast progression from northern and central Sumatera (late August) to Timor (mid December). South of the equator, interannual variability of the onset date is shown to consist of a spatially-coherent large-scale component, together with local-scale noise. The high spatial coherence of onset is similar to that of the September–December seasonal total, while post-onset amounts averaged over 15–90 days and September–December amount residuals from large-scale onset show much less spatial coherence, especially across the main islands of monsoonal Indonesia. The cumulative rainfall anomalies exhibit also their largest amplitudes before or near the onset date. This implies that seasonal potential predictability over monsoonal Indonesia during the first part of the austral summer monsoon season is largely associated with monsoon onset, and that there is much less predictability within the rainy season itself. A cross-validated canonical correlation analysis using July sea surface temperatures over Tropical Pacific and Indian Oceans (80°–280°E, 20°S–20°N) as predictors of local-scale onset dates exhibits promising hindcast skill (anomaly correlation of ~0.80 for the spatial average of standardized rain gauges and ~0.70 for standardized gridded pentad precipitation data).

1. Introduction

Rainfall over Indonesia is governed by the austral-Asian monsoon, whose onset progresses from northwest-to-southeast during the austral spring (Aldrian and Susanto, 2003; Naylor *et al.*, 2007). This is also the season when the El Niño - Southern Oscillation (ENSO) exerts its strongest influence on Indonesian rainfall, particularly during the September–December monsoon onset season (Hamada *et al.,* 2002). The impact of ENSO then diminishes during the core of the rainy season in December–February (Haylock and McBride, 2001; Hendon, 2003; Aldrian *et al.*, 2005, 2007; Giannini *et al.*, 2007), suggesting that the timing of monsoon onset may be potentially predictable.

The date of onset of the rainy season is of particular importance for the agriculture sector over Indonesia (Naylor *et al.*, 2002, 2007). It determines the suitable time for planting crops, while delayed onset during El Niño years (Hamada *et al.*, 2002; Boer and Wahab, 2007) can lead to crop failure. For irrigated rice farmers in Java, information on onset timing is also important for developing strategies (Boer and Subbiah, 2005; Naylor et al., 2007) to avoid exposure of the second rice crop to higher drought risk at dry season planting (April–July), particularly for farmers located at the tail-end of the irrigation system. Farmers in Indonesia often suffer from "false rains" in which isolated rainfall events around the expected onset date do not signal the sustained onset of the monsoon. Such false starts occurring in September prompt potato farmers in Pengalengan in West Java to start planting. In the eastern part of Indonesia, such as East Nuna Tenggara, multiple false starts can cause multiple failures, with farmers sometimes planting up to four times in a season.

This paper discusses the seasonal potential predictability of monsoon onset during the August–December season over Indonesia. The approach taken is based on quantifying the spatial coherence of specific rainfall properties: the September–December (SOND hereafter) rainfall total, rainfall onset date, and post-onset rainfall totals following Haylock and McBride (2001) and Moron *et al.* (2006, 2007). The seasonal predictability of large-scale monsoon onset is then estimated based on sea surface temperatures (SST) in July using a cross-validated canonical correlation analysis (CCA). The two precipitation datasets (rain-gauge and CPC merged analysis of precipitation, CMAP) are described in section 2, together with the definition of onset. Results are presented in section 3, with conclusions drawn in section 4.

2. Data and method

a. Global Summary Of the Day (GSOD) station data

Daily rainfall at rain-gauges for the period 1979–2004 was extracted from 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). There are 91 available stations for Indonesia. The station-years having at least 50% of daily data are extracted and the 57 stations having at least 10 available years are selected. Missing entries (< 13%) were filled using a simple stochastic weather generator (Wilks, 1999), considering the wet-to-wet and dry-to-wet persistence and a gamma distribution for wet days, computed on a monthly basis at each station. If a month is completely missing (< 3% of station-months for SOND), this method simulates a climatological daily sequence for that month.

b. CPC Merged Analysis of Precipitation (CMAP)

Gridded pentad CMAP on a 2.5-degree latitude-longitude grid was selected within a window (12°S–6°N, 90–130°E) over the 1979–2005 period, based only on rain gauges and satellite estimates (Xie and Arkin, 1996). Over this window, there are typically 1–2 rain gauges per grid-box including land (P. Xie, personal communication).

c. Definition of onset

Monsoon onset date can be defined in various ways. We used an agronomical definition (e.g. Sivakumar, 1988) based on local rainfall amounts using thresholds to define the onset, requiring a certain amount of rainfall within a specified period of time, with no extended dry spell occurring afterward. This local definition is sensitive to small-scale processes but is used here in order to be relevant to agricultural management, and to prevent any *a priori* inflation of spatial coherence.

Onset date is defined to be the first wet day of the first 5-day sequence receiving at least 40 mm that is not followed by a dry 10-day sequence receiving less than 5 mm within the following 30 days from the onset date. Onset is computed from August 1st because August-September are the driest months over Indonesia (Aldrian and Susanto, 2003; Aldrian *et al.*, 2007). The latter criterion helps to avoid "false starts," which could be defined, for example, as the difference between the first 5-day wet sequence receiving at least 40 mm and the onset as defined above. The

identification of false starts is sensitive to the choice of the post-onset dry-spell length. In fact, the sensitivity of crops to post-onset dry spell varies. In tropical countries, dry spell with length of more than 7 days would have serious impact on crop yields (Niewolt, 1989). Other study found that 21 rice varieties being exposed to dry spell with length of 16 days during vegetative stage will have delayed harvesting time between 2 and 27 days and reduced yield between 10 and 91% (Dikshit *et al,.* 1987). Indeed, false starts defined with a 10-day dry spell in the following 30 days occur in 46% of station-years, ranging from less than 40% in northern and central Sumatera and Kalimantan to a maximum > 50% in western and central Java. These percentages decrease by a factor of 2–3 when the length of the post-onset dry spell is chosen to be 15 days. The mean onset date is also earlier (by one or two weeks in mean) with a post-onset dry spell lasting 15 days rather than 10 days. Nevertheless, this parameter (and the others entering the onset definition) has only a very weak impact on the large-scale and regional-scale interannual variability of onset dates (for example, the spatial averages of CMAP and GSOD onset-date anomalies computed with both parameters are correlated at 0.99 and 0.97 respectively). Increasing the length of the initial wet spell reduces the noise introduced by weather variability, but the threshold of 5 days is used to facilitate comparison between CMAP and GSOD datasets. The National Agency for Meteorology and Geophysics of Indonesia (BMG) defines the monsoon to start when, after September 1, two consecutive 10-day sequences each receive at least 50 mm of rain. While changing the length and/or the amount of rainfall of the initial wet spell modifies the climatological mean onset date, its impact on interannual variability is again found to be much smaller. The onset date is undefined for 2 cases in CMAP and the missing entries are filled with the latest available onset dates for the corresponding grid-points.

d. Spatial coherence estimates

The spatial coherence of interannual precipitation anomalies is estimated empirically in terms of the interannual variance of spatially-averaged standardized anomalies given by the Standardized Anomaly Index (SAI, Katz and Glantz, 1986). Use of the SAI in the context of tropical rainfall is discussed extensively in Moron *et al.* (2006, 2007). The interannual variance of the SAI (var[SAI]) measures the spatial coherence between M stations (or gridpoints) because it depends on the inter-station correlations; it ranges from var[SAI] = 0 when two samples of equal-size, perfectly covariant, are perfectly out-of-phase, var[SAI]=1/M when all the correlations are zero and var[SAI] = 1 when all stations are perfectly correlated.

The SAI is an empirical estimate of the shared in-phase "signal" across the network. The "noise" component can be defined in terms of the (square-rooted spatial average) squared deviations relative to the SAI. This definition of signal and noise is analogous to the distinction between externally-forced and internally-generated variance in ensembles of General Circulation Model (GCM) simulations (i.e. Rowell, 1998), with stations or grid-points playing the role of GCM ensemble members. A signal-to-noise ratio (SNR) can be formed by dividing the SAI by the noise, but this second-order statistic is more sensitive to sampling issues than the SAI used here.

Statistical significance of interannual correlations is assessed against 1000 synthetic timeseries of the same length and spectral density as the observed pair, but random phase (Janicot *et al.*, 1996), with the two-sided 90%, 95% and 99% significance levels indicated in the following by one (*), two (**) and three (***) asterisks respectively.

3. Results

a. Onset date

The mean onset dates determined from CMAP and GSOD, plotted in Fig. 1a, exhibit a NW-SE progression from late August in northern-central Sumatera and Kalimantan to mid December in Timor. The dates agree well between the two datasets, while there is a large inter-station variability over Java (Fig. 1a) that could be related to small-scale topographic features. Onset occurred before November 1, December 1 and January 1 in 67% (65%), 79% (86%), and 94% (95%) of cases respectively in CMAP (GSOD). Mean onset dates computed for subsets of GSOD stations averaged by sub-region (Table 1 and Fig. 1a) are in good agreement with *Naylor et al.* (2007; their Fig. 1). Using their definition (i.e. the first day when accumulated rainfall from August 1st reaches 200 mm) leads to similar median dates to those shown in Table 1, except in northern areas (not shown). Moreover, the interannual variability is highly consistent between both definition with cross-correlations > 0.85*** for all regions displayed in Table 1 except for northern Sumatera (r = 0.52***). Onset date is less relevant in the northern regions because of the differing seasonality of rainfall north of the equator (Aldrian and Susanto, 2003).

The interannual variability of onset date for the 14 stations over western and central Java is shown in Fig. 1b in terms of the individual standardized anomaly timeseries (dotted). The signal that is common to the 14 stations, defined by

147 the SAI (heavy solid), accounts for a moderate fraction (var[SAI]=0.41) of the total variance at the individual
148 stations, indicating substantial inter-station noise. However, the SAI is correlated at 0.80*** (0.83***) with the
149 large-scale SAI (leading PC timeseries) computed from CMAP onset dates over all 128 gridpoints (heavy dashed
150 blue and red curves respectively), suggesting that the signal in onset over western and central Java is related to the
151 large scale despite considerable small-scale noise. This is also seen in the other sub-regions (Table 1). The influence
152 of ENSO is clearly visible in Fig. 1b, with big delays in large-scale onset during the 1982 and 1997 El Niño events.
153 In fact, the correlation between large-scale SAI (leading PC time series) of CMAP is correlated at 0.84*** (0.84***)
154 with the Niño 3.4 SST anomalies in October, corresponding to the mean onset date across the domain. Some
155 skewness is also visible with delayed onsets exhibiting larger amplitudes than early onsets.
156
157 The leading EOFs of CMAP and GSOD onset dates are plotted in Fig. 2. The leading CMAP EOF accounts for 36%
158 of total variance (EOF#2 accounts for 9% of total variance), and consists of a large-scale monopolar pattern with
159 highest loadings over "monsoonal" Indonesia, (i.e. from southern Sumatera to the Timor Sea (Aldrian and Susanto,
160 2003, their Fig. 2). Loadings remain substantial toward the southeast, but fall off rapidly over northern Sumatera, the
161 Malay Peninsula and northern Kalimantan where they are generally close to zero. The loadings of the leading EOF of
162 GSOD onset dates (31% of the variance) are generally similar to those of CMAP, while their PC timeseries are
163 correlated > 0.90***; there is thus a high level of consistency at large scale between these two contrasting datasets.
164 Similarly, the cross-correlations between the SAIs of each region defined in Fig. 1a are always positive and
165 significant at the one-sided 95% level or greater.
166
167 As discussed in Sect. 2d, the station-scale noise can be defined in terms of the (square-rooted spatial average)
168 squared deviations of the stations' rainfall relative to the SAI. The noise variance computed in this fashion for each
169 of the sub-regions (not shown) is fairly uniform in space, though somewhat smaller in southern Kalimantan, southern
170 Sumatra and western and central Java. However, differences in the spatial sampling between sub-regions do not
171 allow for confidence in this second-order statistic.
172
173 *b. Seasonal rainfall total and post-onset amounts*

The temporal correlations between the leading PC of onset (Fig. 1b) and the leading PC of SOND seasonal total exceed -0.90*** for both datasets. The variance explained by the leading EOF of SOND total (51% in CMAP and 32% in GSOD) is even larger than that of onset, presumably due to the seasonal integration of rainfall that filters out some of the local-scale noise inherent in the definition of onset date. In fact, 46% (CMAP) and 67% (GSOD) of onsets occurred between September 1st and December 31st. This suggests that at least some of the spatially-consistent interannual variability of SOND amount is actually conveyed by the anomalous timing of the monsoon onset.

Three approaches are used to test this hypothesis, by estimating the spatial coherence of rainfall amount beyond the onset date. (i) Firstly, the spatial coherence of the rainfall summed over the 15, 30, 60 and 90 days *following* the local onset date is computed. Post-onset rainfall is *a priori* independent of the timing of the onset of the monsoon, although both may be influenced by ENSO and local-scale SST. The disadvantage of this approach is that post-onset amounts refer to different temporal windows depending on the particular year and station location. Nonetheless, 30-day amounts, for example, refer to periods before January 1st in 69% (CMAP) and 88% (GSOD) of cases. (ii) In the second approach, the component of the SOND total accounted for by the large-scale onset, defined as the leading PC of each dataset (Fig. 2), is removed using a least-squares linear regression. The remaining residual is thus associated only with post-onset amounts, and all information linearly related to large-scale onset is removed *a priori*. (iii) The last method is to compare cumulative spatial-average rainfall anomalies computed from August 1st as expressed as percentage of the long-term mean for early and late onset years.

Estimates of var[SAI] for each quantity are given in Table 2. The spatial coherence is high (i.e. large var[SAI]) for both onset date and seasonal total, but falls to near-zero for post-onset rainfall and SOND residuals. There is nonetheless a weak increase of spatial coherence as the length of the post-onset averaging period increases from 15 to 90 days, expected due to the progressive cancellation of meteorological events as the length of considered period grows. The difference between CMAP and GSOD results could come from the area, mainly oceanic, that is not sampled in GSOD and/or smoothing provided by gridbox-pentad averages in CMAP.

Standardized anomaly timeseries of post-onset 90-day amount for each CMAP gridbox are shown in Fig. 3a, together with the SAI. Spatial coherence is generally low in most years, with the exceptions of the 1982 and 1997

large El Niño events. The loadings of the leading EOF of the post-onset 90-day amount are displayed in Fig. 3b. These are weak over monsoonal Indonesia, especially between southern Sumatera to Sulawesi, where those of the leading EOF of onset peak (Fig. 2), and this mode explains less variance (22% in CMAP and 11% in GSOD) than does the leading EOF of CMAP onset date (36% in CMAP and 32% in GSOD). The temporal behavior of the leading PC is nevertheless consistent with that of onset date, i.e. the post-onset season tends to be anomalously dry when onset is anomalously late, and vice versa, at least for CMAP (r between leading PC of onset and of post-onset 90-day amount is -0.87*** in CMAP and -0.19 in GSOD, the post-onset PCs being correlated at 0.37* between the two datasets). Note that the second EOF of post-onset 90-day amount in GSOD (not shown) explains 10% of total variance and is correlated at -0.60*** (respectively 0.51**) with the leading PC of onset date in GSOD (respectively the leading PC of post-onset 90-day in CMAP). The fact that the loadings are rather large over the eastern Indian Ocean and scattered patches of the northern and eastern oceanic margins of the domain (Fig. 3b) could be evidence of a deterministic signal and warrants further study.

The leading EOF of SOND residuals (Fig. 3c) shares some similarities with that of post-onset 90-day rainfall amounts (Fig. 3b), at least for CMAP (25% explained variance); both have relatively high homogeneous loadings over eastern Indonesia, and weak loadings across monsoonal Indonesia. The leading EOF of GSOD (16% explained variance) lacks similarity with its CMAP counterpart, and their PCs are not significantly correlated (r = 0.23). Nearby stations often have quite different loadings, such as over Java (Fig. 3c). By construction, the leading PC of SOND residuals is orthogonal to the leading PC of onset date.

Figure 4 shows the spatial average of the cumulative rainfall anomalies (averaged over the 57 stations across Indonesia in the upper panel and the 14 stations of Western and Central Java in the lower panel) computed from August 1st and expressed as percentage relative to long-term mean for the 6 latest and earliest mean onset dates. A constant modulation of rainfall anomalies would lead to a straight horizontal line at the mean rainfall anomaly. The largest positive (negative) cumulative anomalies occurred in both cases before or around the early (late) onset dates while the curves usually tend to zero thereafter (Fig. 4). The spatially-averaged rainfall anomalies at the end of the rainy season, somewhere in March-April, are still consistent with the phase of the onset date but the amplitude of

these anomalies is weak (Fig. 4). It suggests that the strongest spatially-coherent signal at large-scale (Fig. 4a) and for a particular subset of stations (Fig. 4b) is before or near the onset date while it tends to cancel thereafter.

c. Seasonal predictability of onset

The substantial spatial coherence of onset date suggests seasonal predictability. To provide a measure of the latter, regression models are built using cross-validated CCA between July SST over the Tropical Pacific and Indian Oceans (80°–240°E, 20°N–20°S) as predictors, and GSOD or CMAP onset dates as predictands. Note that the 14% of missing entries in GSOD were firstly filled with a simple linear regression using the closest CMAP grid-point as predictor. The models were built using the Climate Predictability Tool (CPT) software developed at IRI (http://iri.columbia.edu/outreach/software/); the predictor and predictand fields were prefiltered using EOFs, with the number of modes retained determined by maximizing the model's goodness-of-fit under cross-validation, with 5 years withheld at a time. The leading 5 and 2 (1) EOF modes are retained in SST and CMAP (GSOD) and most of the cross-validated skill is associated with the leading CCA mode whose predictand pattern (i.e. SST pattern) is almost identical for CMAP and GSOD.

Homogeneous correlation maps of the leading CCA mode are shown in Figs. 5a and 5b for SST and onset-date respectively. The SST anomaly structure (Fig. 5a) exhibits a classical ENSO pattern, together with high correlations around Indonesia, such that warm ENSO events are associated with delayed onset (Hamada *et al.*, 2002; Hendon, 2003). The corresponding structure in onset dates (Fig. 5b) indicates that the delayed onsets extend right across Indonesia, with high loadings over monsoonal Indonesia, decreasing weakly (strongly) toward eastern (northwestern) Indonesia. The regression model hindcast skill is plotted in Fig. 5c in terms of anomaly correlation, with regional averages given in Table 1 (last column). Skill values are highest over monsoonal Indonesia, exceeding 0.5** from southern Sumatera to southern Kalimantan and Timor, reaching 0.80*** for the SAI computed over all stations (0.70*** for CMAP). The sub-island subsets of stations in Table 1 achieve station-averaged skills ranging from 0.22 (northern Sumatera) to 0.84*** (southern Kalimantan). The spatial variability of skill over Java could be due to random sampling but also to deterministic signals associated with small-scale orographic features and/or orientation relative to low-level winds.

4. Conclusion

The spatial coherence of onset date and post-onset rainfall is analyzed from GSOD rain-gauges and the CMAP dataset. The onset date is defined using an agronomic approach, i.e. the first significant wet spell (here 40 mm in 5 days) without any potentially damaging dry spell (here 10 days receiving less than 5 mm) thereafter (here in the 30 post-onset days). This definition is best-suited for end-users purpose but suffers from the subjective choice of the parameters. Nevertheless, these parameters broadly reflect the needs and risks associated with major crops of Indonesia, such as lowland rice. The long-term mean onset dates, as well as the frequency of false starts are sensitive to these subjective parameters and future applications should carefully consider the impact of these choices on specific crops. However, for our main purpose of analyzing the spatial coherence of anomalous onset dates, the sensitivity to these parameters largely vanishes.

The interannual variability of rainy season onset over monsoonal Indonesia is shown from both gridded pentad CMAP and daily station GSOD rainfall datasets to be characterized by a large-scale coherent signal, together with a moderate amount of local-scale noise (Figs. 1b & 2). Considering small subsets of GSOD stations recovers this signal, despite the complexity of the island topography (Table 1). The interannual anomalies are dominated by delayed onsets (Fig. 1b). Conversely, the spatial coherence of interannual rainfall anomalies beyond the onset date is weak, as revealed by the amount of rainfall in the 15- to 90- days after the onset and the SOND residuals from large-scale onset (Table 2 & Fig. 3a). The leading EOF of post-onset 90-day CMAP amounts exhibits weak and rather inconsistent loadings over the main islands with high loadings restricted to eastern Indian Ocean and scattered patches of the northern and eastern margins (Fig. 3b). However, this signal is strongly consistent in sign with onset date in CMAP (i.e. late onset associated with smaller post-onset amount and vice versa). The leading EOF of SOND residuals from large-scale onset lacks consistency between the GSOD and CMAP datasets, but both nonetheless exhibit large spatially-coherent loadings over eastern Indonesia, but not over the eastern Indian Ocean (Fig. 3c). The spatial average of cumulative rainfall anomalies also exhibit their largest amplitudes before and near the onset date (Fig. 4), while the post-onset cumulative rainfall anomalies tend almost monotonically toward zero. There may thus be some predictability in post-onset seasonal amounts, but most of the spatially-coherent signal in SOND seasonal total, especially across islands, is merely related to the onset.

Our main finding is that most of the large-scale interannual signal of SOND seasonal rainfall total is conveyed by variations in the onset date of the rainy season. This implies that (i) rainfall monitoring at a small set of stations spread across Indonesia should be sufficient to establish interannual anomalies of onset date, and (ii) the scale of the interannual variability of the onset suggests a large scale forcing and potential seasonal predictability. Indeed large-scale onset is found to be highly correlated with an ENSO SST pattern during July (Fig. 5a), i.e. at least one month and half before the mean local-scale onset date. A cross-validated CCA using July SST in tropical Indian and Pacific Oceans (80°–240°E, 20°N–20°S) as a predictor leads to promising skill values for the large-scale onset date (r = 0.80*** for GSOD ; Fig. 5b). Further work is needed to examine the associated circulation changes and to investigate the roles of ENSO and Indian Ocean climate variability (Hendon, 2003).

The spatial variation of hindcast skill (Fig. 5c) and onset EOF loadings (Fig. 2) warrants further study. Both exhibit maxima from southern Sumatera to southern Kalimantan—quite close to the Equator—and decreases gradually southward across Java and Sonde islands and more rapidly northward (Figs 2 & 5c, Table I). The latter decrease could be related to the year-round rainfall there (Aldrian and Susanto, 2003) and onset date should be viewed merely as an increase of the rainfall rather than the transition between a real dry and wet season. In that case, the onset date is sensitive to the subjective choices used to define it and is clearly less robust. This does not apply to monsoonal Indonesia south of 5°S. The highest EOF loadings and SST-related skill over southern Sumatera to southern Kalimantan coincide with the largest inter-quartile range of interannual variability (Table I). This subequatorial band is perhaps the most sensitive to the spatial shift of the ITCZ that probably triggers the onset of the rainy season. The complex orography across Java could also enhance the intra-regional noise even between close stations but we must also keep in mind that the spatial sampling is highest over Java (Fig. 1a). Similarly, the nature of spatial coherence for post-onset rainfall and SOND residuals over eastern Indonesia and the eastern Indian Ocean (post-onset rainfall only), as well as sea-land contrast needs further investigation using better sampled datasets and/or regional model simulations.

The large-scale signal in onset is still strongly present in multi-station small sub-island regions (Table 1), indicating the potential to downscale the large-scale onset signal to the near-local scale. However, it is clear that individual stations exhibit considerable noise (Table 1). Thus, careful consideration needs to be given to the trade-off between

313 potentially more-accurate forecasts at the aggregated scale, versus local specificity for use in climate risk
314 management. The large-scale nature of seasonal predictability of onset should enable improved agricultural planning
315 in the future, together with better identification of false starts to the rainy season via real-time monitoring and short-
316 term forecasts of the large-scale evolving monsoon circulation. Forecasts of the Madden-Julian oscillation may lend
317 an additional source of predictability at intraseasonal lead times (Wheeler and McBride, 2005).
318
319 **Acknowledgments**: We are grateful to P. Xie for information on the rain gauge measurements used in the CMAP,
320 and to two anonymous reviewers whose comments helped us to clarify the paper. This research was supported by
321 grants from the National Oceanic and Atmospheric Administration (NOAA), NA050AR4311004, the US Agency for
322 International Development's Office of Foreign Disaster Assistance, DFD-A-00-03-00005-00, and the US
323 Department of Energy's Climate Change Prediction Program, DE-FG02-02ER63413.
324

References

Aldrian, E., and R. D. Susanto, (2003) Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature, *Int. J. Climatol.*, **23**, 1435-1452.

Aldrian, E., D. Sein, D. Jacob, L. Dümenil Gates, and R. Podzun, (2005) Modeling Indonesian rainfall with a coupled regional model, *Climate Dynamics*, **25**, 1-17.

Aldrian, E., L. Dïumenil Gates, and F.H. Widodo, (2007) Seasonal variability of Indonesian rainfall in ECHAM4 simulations and in the reanalyses: the role of ENSO, *Theor. Appl. Climatol.*, **87**, 41-59.

Boer R., and A.R. Subbiah, (2005) Agriculture drought in Indonesia. In Vijendra S. Boken, Arthur P. Cracknell and Ronald L. Heathcote eds. Monitoring and Predicting Agricultural Drought: A Global Study. Oxford University Press, pp 330-344.

Boer R., and I. Wahab, (2007) Use of seasonal surface temperature for predicting optimum planting window for potato at Pengalengan, West Java, Indonesia. In M.V.K. Sivakumar and J. Hansen eds. Climate Prediction and Agriculture: Advance and challenge. Springer, New York, pp. 135-141.

Dikshit, U.N., D. Parida, and D. Satpathy, (1987) Genetic evaluation and utilization: Drought tolerance. IRRN 12:6-7.

Giannini, A., A.W. Robertson and J.H. Qian, (2007) A role for tropical tropospheric temperature adjustment to ENSO in the seasonality of monsoonal Indonesia precipitation predictability, *J. Geophys. Res. (Atmosphere)*, **112**, D16110, doi:10.1029/2007JD008519.

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

Haylock, M, and J.L. McBride, (2001) Spatial coherence and predictability of Indonesian wet season rainfall, *J. Climate*, **14**, 3882-3887.

Hendon, H.H. (2003) Indonesian rainfall variability: impacts of ENSO and local air-sea interaction, *J. Climate*, **16**, 1776-1790.

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

Katz, R.W., and M.H. Glantz, (1986) Anatomy of a rainfall index, *Mon. Wea. Rev.*, **114**, 764-771.

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

353 Moron, V., A.W. Robertson, M.N. Ward, and P. Camberlin, (2007) Spatial coherence of tropical rainfall at regional
354 scale, *J. Climate*, **20**, 5244-5263.
355 Naylor, R.L., W. Falcon, N. Wada, and D. Rochberg, (2002) Using El-Niño Southern Oscillation climate data to
356 improve food policy planning in Indonesia, *Bulletin of Indonesian Economic Studies*, **38**, 75-91.
357 Naylor, R.L., D.S. Battisti, D.J. Vimont, W.P. Falcon, and M.B. Burke, (2007) Assessing the risks of climate
358 variability and climate change for Indonesian rice agriculture, *Proc. Nat. Acad. Sci.*, **104**, 7752-7757.
359 Niewolt, S., (1989) Estimating of agricultural risks of tropical rainfall, *Agric. For. Meteo.*, **45**, 251-263.
360 Rowell, D.P., (1998) Assessing potential seasonal predictability with an ensemble of multi-decadal GCM
361 simulations, *J. Climate*, **11**, 1073-1093.
362 Sivakumar, M.V.K., (1988) Predicting rainy season potential from the onset of rains in southern Sahelian and
363 Sudanian climatic zones of West Africa, *Agric. For. Meteo.*, **42**, 295-305.
364 Xie, P., and P.A. Arkin, (1996) Analyses of global monthly precipitation using gauge observations, satellite
365 estimates, and numerical model predictions, *J. Climate*, **9**, 840-858.
366 Wheeler, M.C., and J.L. McBride, (2005) Intraseasonal variability of the Australian-Indonesian monsoon region, in
367 "Intraseasonal variability of the Atmosphere-Ocean system", W.K.M. Lau and D.E. Walliser eds, Praxis Publishing
368 Wilks, D.S., (1999) Interannual variability and extreme-value characteristics of several stochastic daily precipitation
369 models, *Agric. For. Meteo.*, **93**, 153-169.
370

Figure captions

Figure 1: (a) Mean onset date computed in CMAP (shading) and GSOD (dot) as the first wet day of a 5-day sequence receiving > 40 mm from August 1st without a dry 10-day sequence receiving < 5 mm in the following 30 days from onset. The name of each subset of stations is displayed. (b) Standardized onset date for western and central Java GSOD stations (dotted lines) with the average, i.e. Standardized Anomaly Index – SAI – (solid black line), together with the CMAP SAI (blue dashed line) and standardized leading PC time series (red dashed line) computed from all 128 CMAP gridpoints. The dashed horizontal lines delineate the 95% confidence interval of a set of 14 white noise time series. Note that one standard deviation corresponds to an averaged deviation of ~ 20 days for Western and Central Java.

Figure 2: Leading empirical orthogonal function (EOF) of CMAP (shading) and GSOD (dot) onset dates, plotted as correlations with the principal component timeseries. The timeseries of onset date at each gridpoint were standardized prior to EOF analysis.

Figure 3: (a) Individual standardized anomalies of rainfall total for the 90-day period after the local onset date at the 128 CMAP gridpoints (dots) with the SAI (solid). The dashed horizontal lines delineate the 95% confidence interval of a set of 128 white noise time series. (b) Leading empirical orthogonal function (EOF) of post-onset 90-day amounts in CMAP (shading) and GSOD (dot). (c) Leading EOF of SOND residuals. Units in (b) and (c) are correlations with the respective principal component timeseries.

Figure 4: Spatial average of cumulative rainfall anomalies (a) for all 57-stations and (b) 14-stations from western and central Java (Fig. 1a) computed from August 1st and expressed as percentage from the long-term mean for the six latest (in red) and earliest (in blue) onsets (computed from the spatial average of onset dates). The dashed line indicates each year and the full bold line indicates the mean of the 6 years. The time series are low-pass filtered with a Butterworth filter (cut-off frequency = 1/30 cycle-per-day). The asterisks indicate the station average onset date.

Figure 5: Homogeneous correlation maps of (a) SST, and (b) onset date from CMAP (shading) and GSOD (circles), of the leading canonical correlation analysis (CCA) mode (c) MOS skill (i.e. correlation between observed and hindcast onset date) associated with the leading CCA mode between July SST and onset dates.

Tables

Table 1: Statistics of GSOD station onset date by sub-region (N is number of stations), computed from the Standardized Anomaly Index (SAI) of each region. The hindcast skill refers to the correlation between the observed and hindcast SAI with a cross-validated CCA using July SSTs as predictors. One, two and three asterisks indicate correlation significant at the two-sided 90%, 95%, 99% level according to a random-phase test (Janicot *et al.*, 1996).

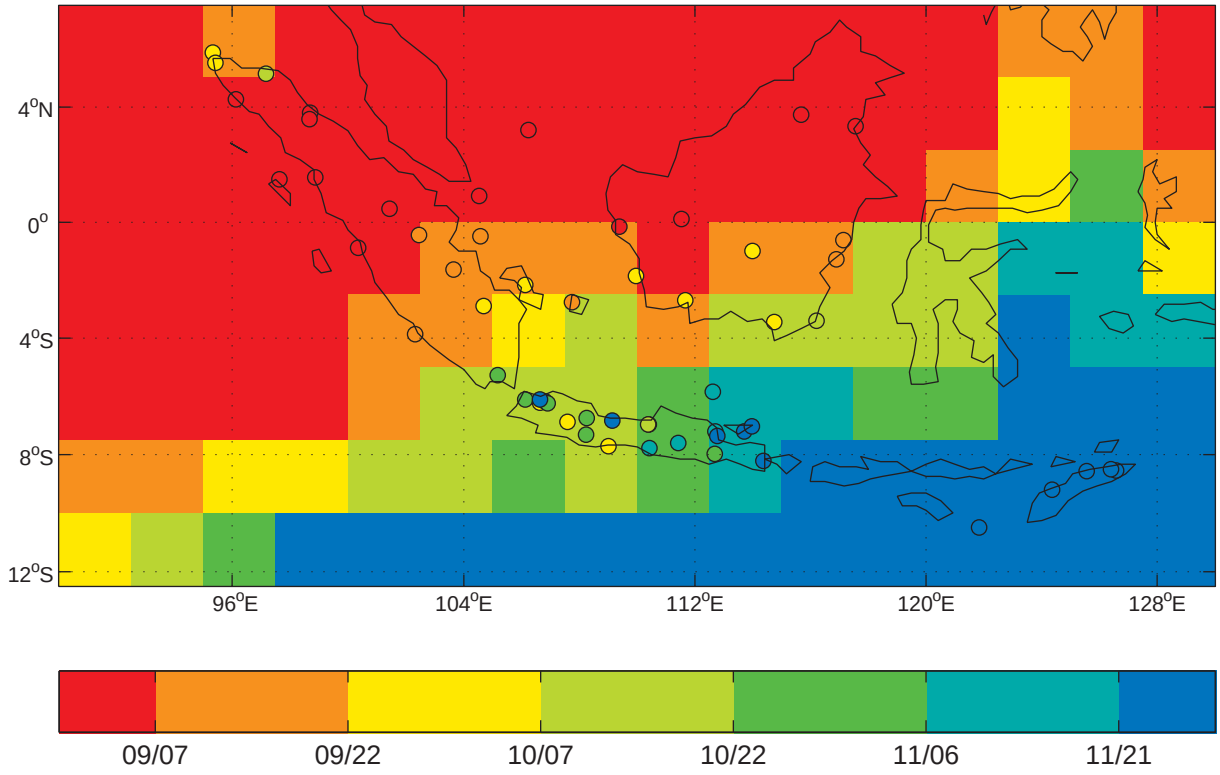
	N	25%, 50% and 75% percentiles of the spatial average	var [SAI]	Correlation with large-scale SAI of CMAP	Correlation with PC#1 of CMAP	Hindcast Skill
Western and Central Java (W of 112°E)	14	Oct 16, Oct 28, Nov 11	0.41	0.80***	0.83***	0.59***
Eastern Java (E of 112°E)	7	Nov 13, Nov 21, Dec 2	0.44	0.74***	0.72***	0.61***
Southern Sumatera (S of 1°S)	6	Sept 5, Sept 19, Oct 17	0.60	0.86***	0.88***	0.74***
Central Sumatera (between 1°S and 2°N)	7	Aug 15, Aug 24, Aug 31	0.43	0.76***	0.74***	0.51**
Northern Sumatera (N of 2°N)	6	Sep 1, Sep 11, Sep 15	0.23	0.46**	0.41**	0.22
Southern Kalimantan (S of 1°S)	6	Sep 17, Sep 22, Oct 25	0.72	0.80***	0.79***	0.84***
Central Kalimantan (N of 1°S)	5	Aug 11, Aug 29, Sep 12	0.48	0.70***	0.69***	0.46**
Eastern Indonesia (E of 120°E and S of 8°S	5	Nov 30, Dec 13, Dec 25	0.57	0.63**	0.60**	0.49**

405 **Table 2** : Interannual variance of the Standardized Anomaly Index (*Var[SAI]*) of the 57 GSOD stations, and 128
406 gridpoints of CMAP for local onset date, and post-onset 15-, 30-, 60-, and 90- day rainfall totals. Var[SAI] ranges
407 between 0 (correlation of –1 between two equal-sized and perfectly covarying samples), 1/*m* (= 0.02 for m=57 and
408 0.008 for m=128) where *m* is the number of locations for spatially independent variations, and 1 (perfect correlation
409 between stations) (Moron *et al.*, 2007).

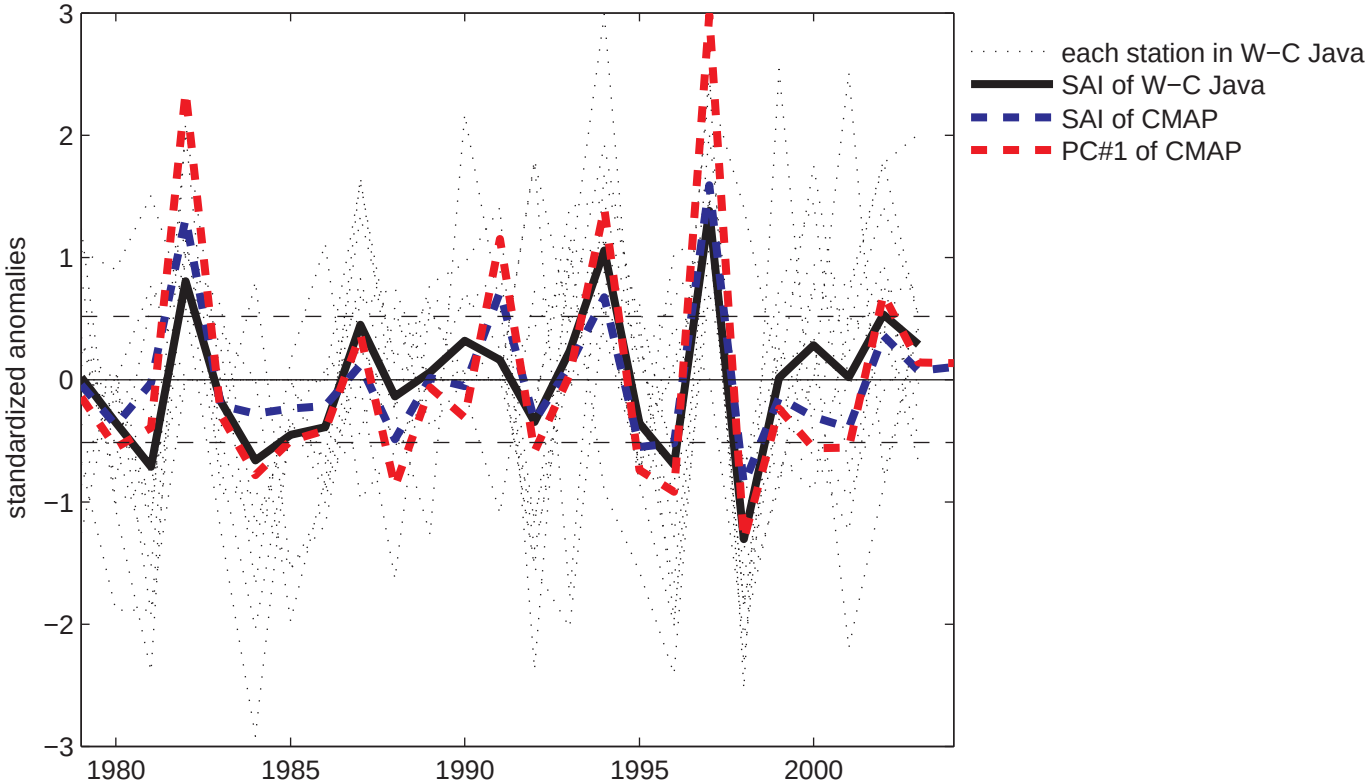
	Var(SAI) GSOD	Var(SAI) CMAP
Onset	0.30	0.31
15-day	0.03	0.05
30-day	0.03	0.08
60-day	0.05	0.11
90-day	0.06	0.14
SOND	0.26	0.46
SOND residuals	0.10	0.16

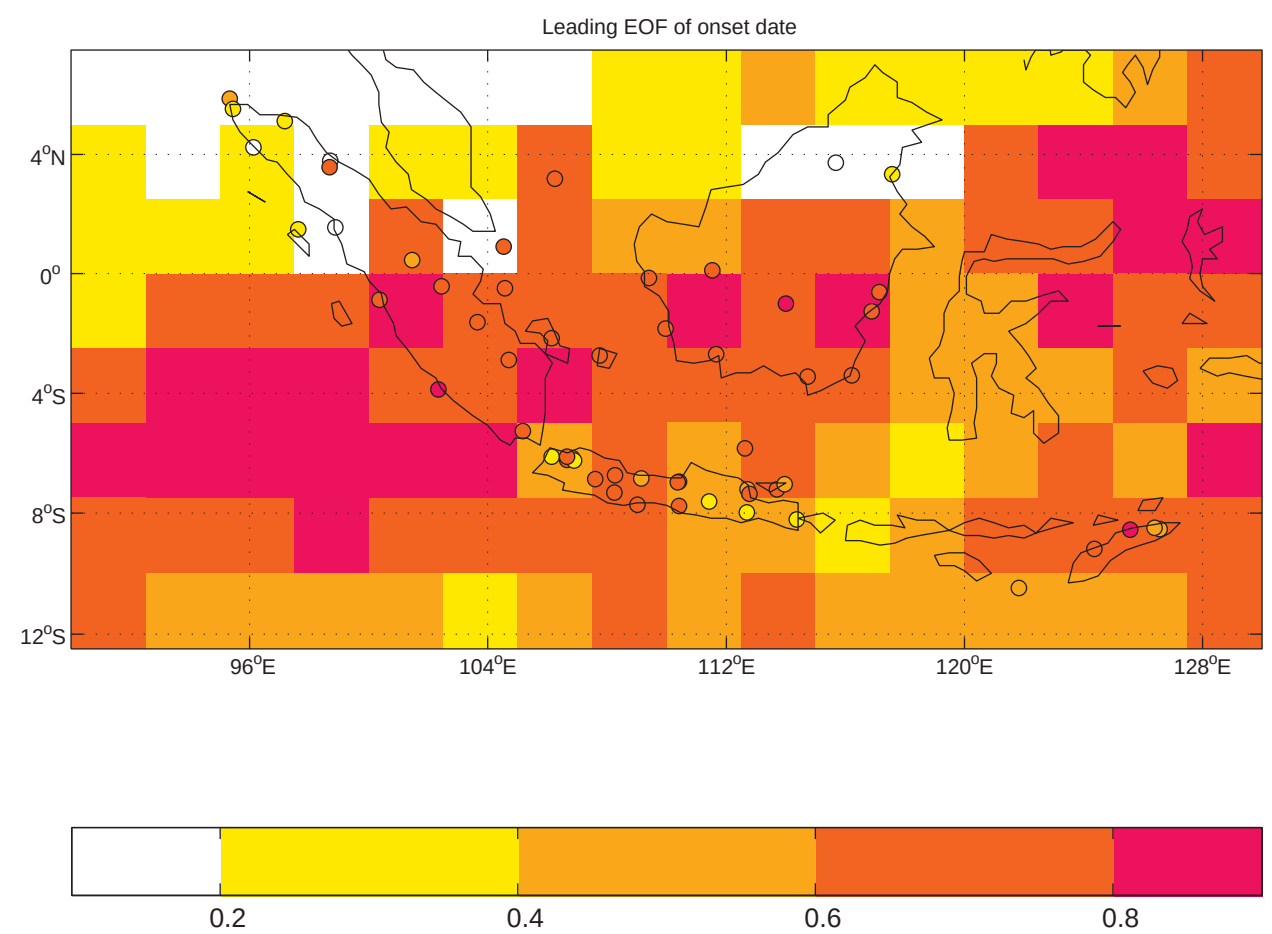
410

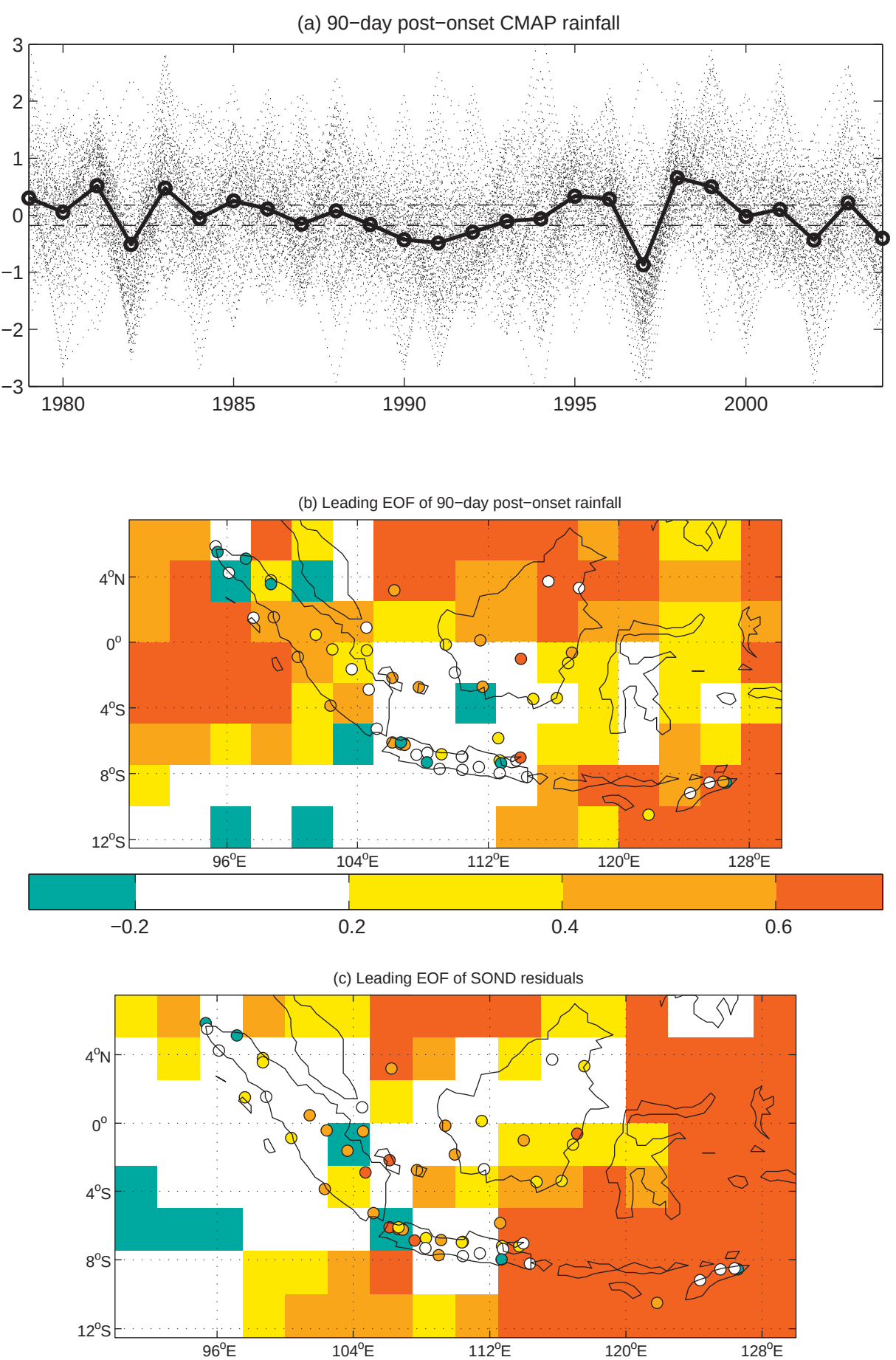
(a) Mean onset date

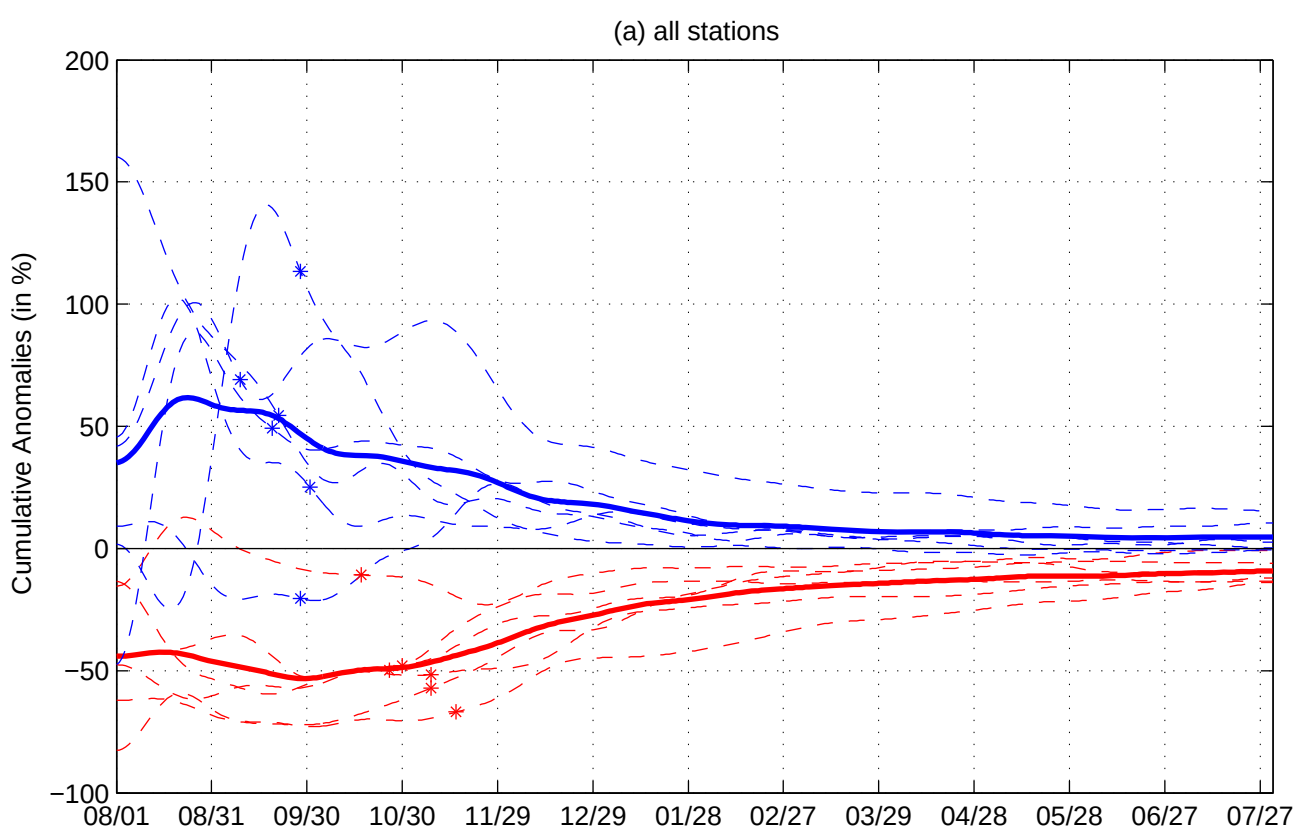


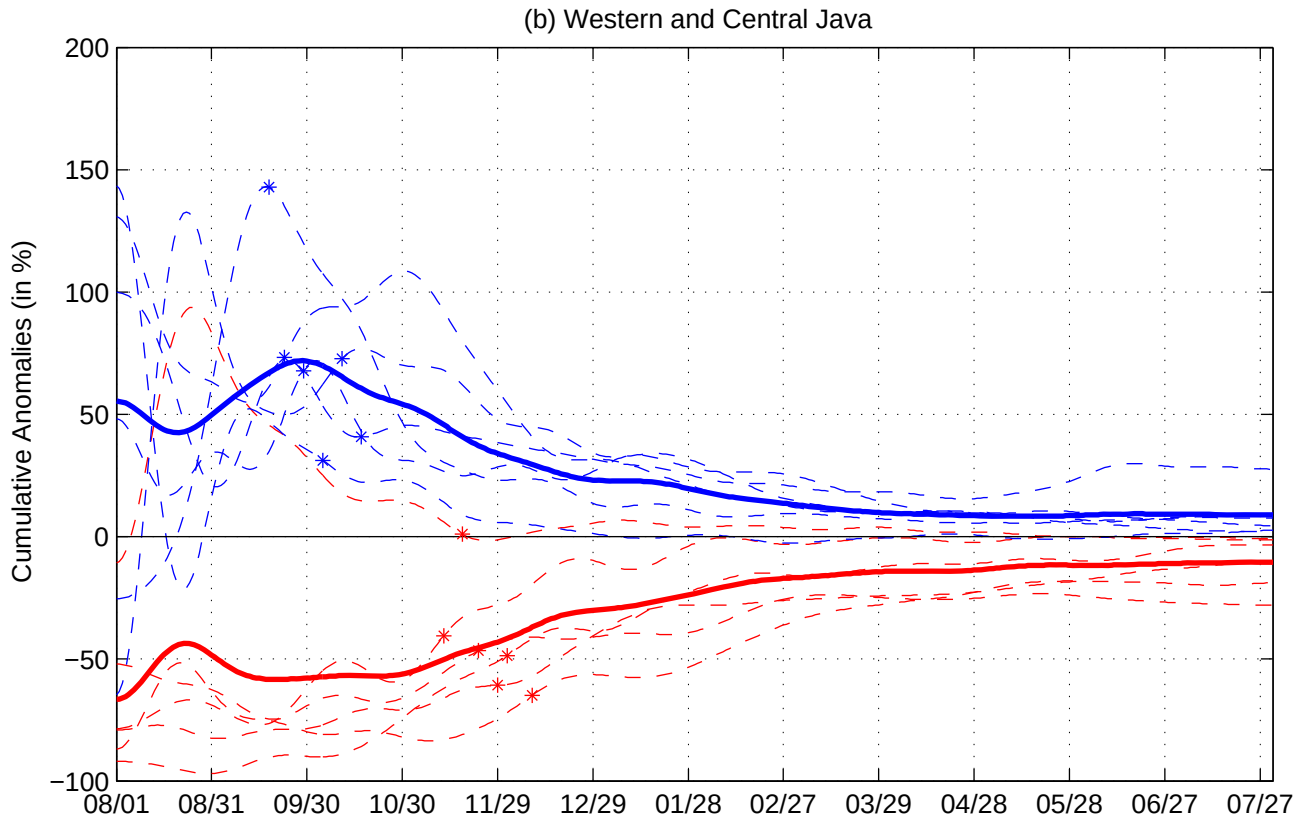
(b) Western & Central Java vs large−scale onset date

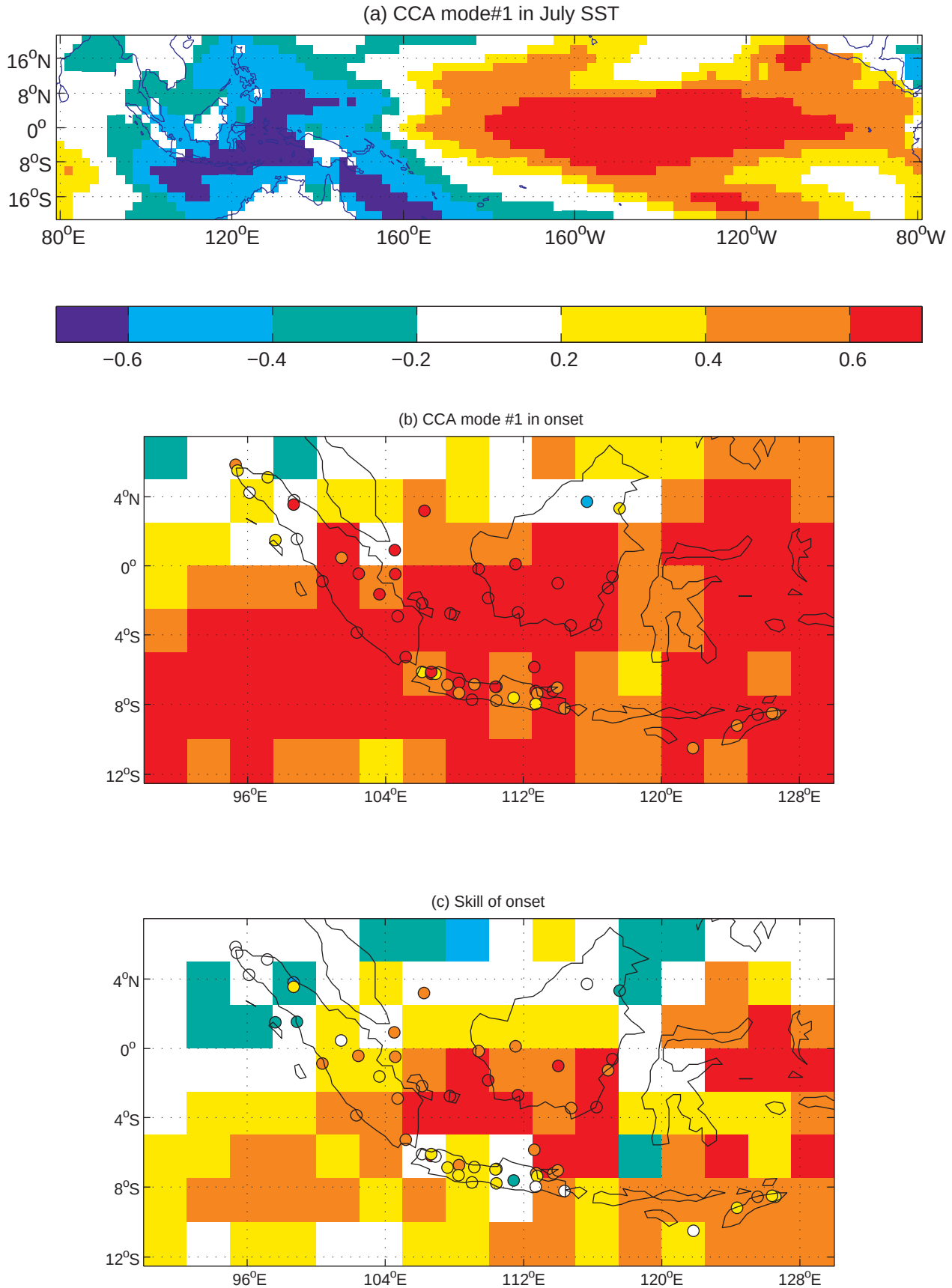



Leading EOF of onset date
4°N
0°
4°S
8°S
12°S
96°E
104°E
112°E
120°E
128°E
0.2
0.4
0.6
0.8


(a) 90−day post−onset CMAP rainfall
3
2
1
0
−1
−2
−3
1980
1985
1990
1995
2000
(b) Leading EOF of 90−day post−onset rainfall
4°N
0°
4°S
8°S
12°S
96°E
104°E
112°E
120°E
128°E
−0.2
0.2
0.4
0.6
(c) Leading EOF of SOND residuals
4°N
0°
4°S
8°S
12°S
96°E
104°E
112°E
120°E
128°E


(a) all stations
Cumulative Anomalies (in %)
200
150
100
50
0
−50
−100
08/01
08/31
09/30
10/30
11/29
12/29
01/28
02/27
03/29
04/28
05/28
06/27
07/27


(b) Western and Central Java
Cumulative Anomalies (in %)
200
150
100
50
0
−50
−100
08/01
08/31
09/30
10/30
11/29
12/29
01/28
02/27
03/29
04/28
05/28
06/27
07/27


(a) CCA mode#1 in July SST
16°N
8°N
0°
8°S
16°S
80°E
120°E
160°E
160°W
120°W
80°W
−0.6
−0.4
−0.2
0.2
0.4
0.6
(b) CCA mode #1 in onset
4°N
0°
4°S
8°S
12°S
96°E
104°E
112°E
120°E
128°E
(c) Skill of onset
4°N
0°
4°S
8°S
12°S
96°E
104°E
112°E
120°E
128°E