

Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia

Andrew W. Robertson,^{a*} Vincent Moron^{a,b} and Yunus Swarinoto^c

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

^b CEREGE, the University of Aix-Marseille, and the Institut Universitaire de France, France

^c Bureau of Meteorology and Geophysics, Indonesia

ABSTRACT: The seasonal predictability of rainfall over a small rice-growing district of Java, Indonesia is investigated in terms of its daily characteristics during the September–December monsoon-onset season. The seasonal statistics considered include rainfall frequency, mean daily intensity, median length of dry spells, as well as the onset date of the rainy season. General circulation model retrospective seasonal forecasts initialized on August 1 are downscaled to a set of 17 station locations using a nonhomogeneous hidden Markov model. Large ensembles of stochastic daily rainfall sequences are generated at each station, from which the seasonal statistics are calculated and compared against observations using deterministic and probabilistic skill metrics. The retrospective forecasts are shown to exhibit moderate skill in terms of rainfall frequency, seasonal rainfall total, and especially monsoon onset date. Some skill is also found for median dry-spell length, while mean wet-day persistence and daily rainfall intensity are not found to be predictable. Copyright © 2008 Royal Meteorological Society

KEY WORDS seasonal predictability; Indonesia; hidden Markov model

Received 3 July 2008; Revised 22 October 2008; Accepted 26 October 2008

1. Introduction

Seasonal climate forecasts are typically issued in terms of three-month averages of rainfall or temperature, as a compromise between maximizing the ratio of predictable climate signal to unpredictable weather noise, while still capturing seasonal evolution (e.g. Goddard *et al.*, 2001). However, such seasonally averaged forecasts are often of limited use to decision makers, where risk management in agriculture, for example, may require information on aspects such as the onset of the rainy season, or the probability of rainfall occurrence, long dry spells, or rainfall extremes within the growing season. In addition, the skillful spatial scale of current general circulation modal (GCM) seasonal predictions is of the order of several hundred kilometers (Gong *et al.*, 2003), again much larger than may be required for effective climate risk management at the scale of a small administrative district. Downscaling is required, within the physical constraints of the regional climate system, and the limitations of available downscaling methodologies.

Recent work suggests that in the tropics, rainfall frequency at the station scale is more seasonally predictable than the seasonal total of rainfall; this is primarily due to the relatively higher spatial coherence of interannual

anomalies of rainfall frequency compared to those of mean daily rainfall intensity (Moron *et al.*, 2006, 2007).

Probabilistic models of “weather within climate” with daily resolution based on stochastic weather generators, hidden Markov models, and K-nearest neighbors approaches have been used to express GCM-based seasonal forecasts in terms of ensembles of stochastic local daily weather sequences that can then, in principle, be used to drive models of crop growth and yield (Hansen and Ines, 2005; Ines and Hansen, 2006; Robertson *et al.*, 2007; Moron *et al.*, 2008d). The nonhomogeneous hidden Markov model (NHMM) has proved to be a promising method for constructing multistation weather generators (Hughes and Guttorp, 1994). Over northeast Brazil, Robertson *et al.* (2004) found that interannual variability in the frequency-of-occurrence of 10-day dry spells could be simulated reasonably, using an NHMM with GCM seasonal-mean large-scale precipitation as a predictor. Similar downscaling results were obtained over Queensland, Australia (Robertson *et al.*, 2006). The NHMM has been applied to two other locations in Australia in downscaling studies (Charles *et al.*, 2003, 2004).

In this paper, retrospective GCM seasonal precipitation forecasts are downscaled to a set of rainfall stations over Indramayu, a small (2140 km²) flat coastal district of West Java, using an NHMM and their skill assessed under cross-validation. We focus on a set of weather statistics of potential relevance to agriculture, namely daily rainfall frequency, mean daily intensity on wet days, mean dry-spell lengths, wet-day persistence, and the

*Correspondence to: Andrew W. Robertson, International Research Institute for Climate and Society (IRI), Columbia University Lamont Campus, 61 Route 9W, Palisades, NY 10960, USA.
E-mail: awr@iri.columbia.edu

monsoon onset date, in addition to the seasonal rainfall total. Deterministic and probabilistic measures of skill are quantified.

Rainfall over Indonesia is governed by the austral-Asian (northwest) monsoon, whose onset progresses from northwest-to-southeast during the austral spring (Aldrian and Susanto, 2003). Many studies have shown that the El Niño – Southern Oscillation (ENSO) exerts its strongest influence on Indonesian rainfall during the September–December monsoon onset season (e.g., Hamada *et al.*, 2002; Juneng and Tangang, 2005). The impact of ENSO then diminishes during the core of the rainy season in December–February (Haylock and McBride, 2001; Aldrian *et al.*, 2005, 2007; Giannini *et al.*, 2007), suggesting that the timing of monsoon onset may be potentially predictable. Moron *et al.* (2008a,b) have recently argued that much of the seasonal predictability in the September–December total rainfall is associated with changes in monsoon onset date.

Indramayu, situated on the north coast of West Java, is an important rice-growing district contributing about one-quarter of Java's rice production. Farmers experience droughts and floods that cause significant losses in rice production. The date of onset of the rainy season is of particular importance, determining the suitable time for planting crops, while delayed onset during El Niño years (Hamada *et al.*, 2002; Naylor *et al.*, 2002; Boer and Wahab, 2007) can lead to crop failure. Earlier onsets occur in La Niña years, but these are generally less pronounced than the delays during El Niño (Moron *et al.*, 2008b); these may, however, be potentially beneficial to farmers by extending the length of the growing season. "False rains", in which isolated rainfall events occur around the expected onset date also present problems for farmers.

This paper is motivated by the needs of the Indonesian Bureau of Meteorology and Geophysics (BMG), which has been working with the agricultural office to develop climate forecasts that are specific to agriculture over Indramayu. The September–December season is selected for its importance to agriculture as well as its relatively

high seasonal predictability of rainfall. The paper is organized as follows. Section 2 describes the rainfall data and GCM simulations, Section 3 describes the hidden Markov model and statistical methods. The results are presented in Section 4, with conclusions given in Section 5.

2. Data

2.1. Observed rainfall data

Daily rainfall observations recorded at 17 station locations over Indramayu during the period 1979–2002, for the September–December (SOND) season were used in this study; these data were provided by BMG. Missing values (<6% of station days) were simply flagged for the NHMM. For the purposes of computing observed rainfall statistics against which to validate the forecasts, the missing values were filled using a simple stochastic weather generator (Wilks, 1999), considering the wet-to-wet and dry-to-wet day persistence and a gamma distribution for rainfall amounts on wet days. All parameters were computed separately for each station and calendar month; if a month is completely missing, this method simulates a climatological daily sequence for that month. The average number of wet days (defined here as receiving 0.1 mm or more of rain) is 20–30 days, with mean intensities (i.e. the mean amount of rainfall on wet days) of about 2–4 mm/day. Their spatial distributions are rather uniform, as shown in Figure 1.

An agronomical definition of monsoon onset (e.g. Sivakumar, 1988) is adopted based on local rainfall amounts. Onset is defined as the first wet day of the first 5-day sequence receiving at least 40 mm, which is not followed by a dry 15-day sequence receiving less than 5 mm within the following 30 days from the onset date. The latter criterion helps to avoid false starts. Onset is computed from the 1st September. Changing the length and/or the amount of rainfall of the initial wet spell modifies the climatological mean onset date, but

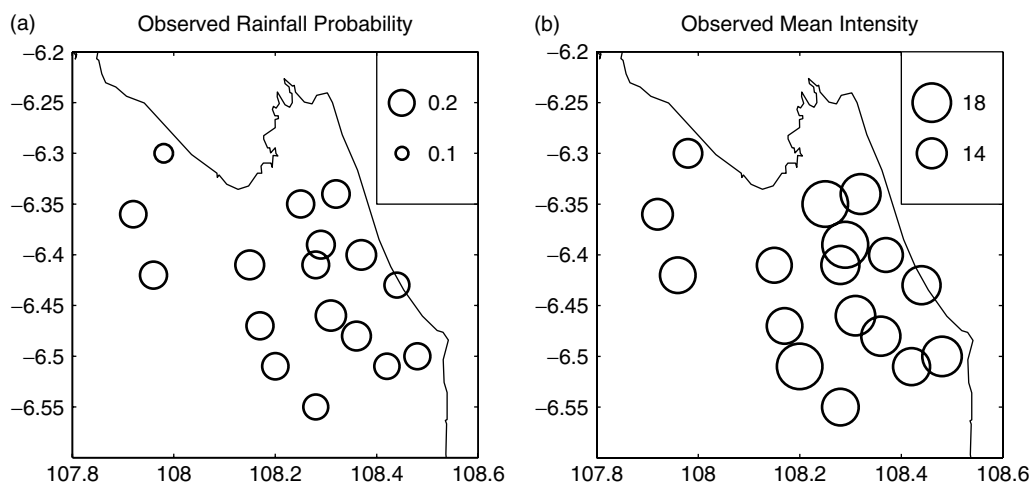


Figure 1. Observed climatological mean station values of (a) rainfall probability, and (b) mean rainfall intensity (mm/d).

the impact on its interannual variability is found to be minimal.

2.2. Seasonal climate forecast model

A set of retrospective seasonal forecasts from the ECHAM4.5 atmospheric GCM driven with constructed analog predictions of sea surface temperature (SST) were initialized on August 1 of each year 1979–2002 (Li *et al.*, 2007). In this “two-tier” system, SST is predicted on a monthly basis from the previous month (here July) using the constructed analog approach (van den Dool, 1994). The ECHAM4.5 atmospheric GCM (Roeckner *et al.*, 1996) is then run at T42 horizontal resolution (approx. 2.8° grid) using the SST predictions at the lower ocean boundary, with the 24 ensemble members initialized from slightly differing initial conditions taken from long simulations with observed SSTs prescribed. There is no initialization of the atmosphere (or land surface conditions) through data assimilation. These retrospective forecasts were made at IRI and obtained through the IRI Data Library.

3. Methods

3.1. Nonhomogeneous hidden Markov model (NHMM)

The NHMM used here follows the approach of Hughes and Guttorp (1994) to model daily rainfall occurrence, while additionally modeling rainfall amounts; it is fully described in Robertson *et al.* (2004, 2006). In brief, the time sequence of daily rainfall measurements on a network of stations is assumed to be generated by a first-order Markov chain of a few discrete hidden (i.e. unobserved) rainfall “states”. For each state, the daily rainfall amount at each station is modeled as a finite mixture of components, consisting of a delta function at zero amount to model dry days, and a combination of two exponentials to describe rainfall amounts on days with nonzero rainfall. The state-transition matrix is treated as a (logistic) function of a multivariate predictor input time series obtained from the GCM retrospective forecasts. Missing data is treated explicitly, with parameter estimates derived from the days that are present (Kirshner, 2005).

3.2. Downscaling experimental design and cross-validation

The GCM retrospective forecasts are downscaled using the NHMM to obtain a large ensemble of stochastic daily rainfall sequences at each of the 17 stations, for the period September 1 to December 31, 1979–2002. Firstly, monthly GCM precipitation fields were obtained for the months August–January over a regional window (80E–180E, 20S–15N) and standardized at each grid-point by subtracting the mean and dividing by the standard deviation (SD). The resulting anomalies were then weighted spatially using a Gaussian ($\sigma_x = 60^\circ$, $\sigma_y = 15^\circ$)

to emphasize gridpoints over Indonesia, and then interpolated linearly to daily values, selecting the September 1 to December 31 period.

The NHMM was trained using the 24-member GCM ensemble mean precipitation under 8-fold cross-validation, omitting 3 consecutive years at a time. A principal components analysis (PCA) of the daily-interpolated GCM ensemble mean precipitation fields was used to define the inputs to the NHMM, retaining the leading 3 PCs (92.4% variance). The daily interpolation was carried out linearly, with the monthly-mean values specified at the mid-points of each month. The correlations of the (seasonal averaged) PCs with the (seasonal and station averaged) station rainfall are 0.59, –0.49, and 0.66 respectively, while the respective correlations with the Nino3.4 index are –0.79, 0.69, and –0.85. For each fold of the cross-validation, the PCs were recomputed on the training subset of 21 years.

To make the rainfall simulations, we proceed as follows for each of the 8-folds of the cross-validation. For each of the 24 ensemble members, the (linearly interpolated) daily GCM precipitation fields for the 3 left-out years were projected onto the leading 3 empirical orthogonal functions (EOFs) computed from the respective 21-year training period. The resulting 24 time series (one per GCM ensemble member) were then used in conjunction with the NHMM trained on the 21-year training period to make 3 NHMM simulations, yielding a total of 72 simulated daily rainfall sequences for each SOND season. Note that the individual GCM ensemble members were used for simulation, rather than the GCM’s ensemble mean, in order to retain the distribution across the GCM ensemble. Skill levels were found to decrease if the individual ensemble members were used in the NHMM training step, in place of the ensemble mean.

4. Results

4.1. NHMM training

The choice of the appropriate number of hidden states k in the NHMM was guided by computing the log-likelihood of models with different choices of k under cross-validation (Figure 2). As is typical, the out-of-sample log-likelihood increases sharply with k initially, and then levels off, with diminishing returns for high values. We chose $k = 4$; the downscaling results were checked for $k = 3–6$ and found to be very similar. In all cases, the NHMM was initialized 30 times from random seeds, selecting the solution with the highest (in-sample) log-likelihood. Note that the log-likelihood is negative because the likelihood – which is the probability of the observed rainfall data, given the model – is less than unity; the model fit is not perfect even for large k because (a) the NHMM is a simple representation of the rainfall process and its relationship with large-scale GCM monthly precipitation, (b) the GCM forecasts contain errors, and (c) the parameters estimated in the NHMM training are maximum likelihood estimates.

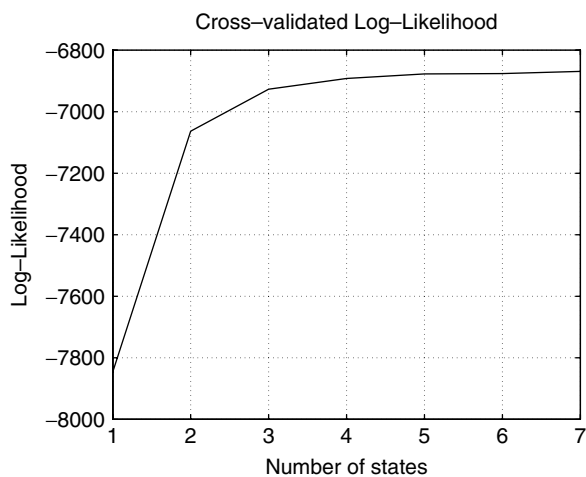


Figure 2. Cross-validated log-likelihood as a function of the number of NHMM states.

4.2. NHMM interpretation

Maps of rainfall properties associated with each of the states are plotted in Figure 3, with the estimated state sequence in time shown in Figure 4. The four rainfall states describe daily rainfall conditions ranging from dry (state 1) to wet (state 4), in terms of rainfall probability at each station (Figure 3(a)–(d)), and the rainfall distribution on wet days, with the latter plotted here in terms of mean rainfall intensity (Figure 3(e)–(h)), calculated from each state's rainfall-distribution parameters. Rainfall probabilities are stratified rather monotonically by the NHMM state, with much smaller differences between stations for a given state. Mean rainfall intensities vary less smoothly between the states, with larger interstation differences, especially for the dry state where there are few wet days over which to estimate the rainfall distribution parameters.

The temporal evolution of rainfall in the dataset can be described by estimating the most-likely sequence of the four NHMM states. This is performed using the Viterbi algorithm (Forney, 1978), which uses the NHMM parameters (estimated here for the whole dataset without cross-validation) together with the rainfall data. Figure 4 provides a graphic illustration of the rainfall variability at the district level, in terms of its seasonality, subseasonal variability, as well as interannual variability. The driest state predominates during September, with spells of the wetter states becoming more prevalent in November–December. The stochastic nature of the model is clear, with a considerable variability of the sequences from year to year, and within each season. The monsoon onset was clearly substantially delayed during the El Niño events of 1982, 1987, 1994, and 1997.

4.3. Forecast ensembles

The downscaling experiment performed in this study yields ensembles of retrospective forecasts, consisting of stochastic daily sequences of rainfall at the 17 rainfall station locations. In order to investigate the characteristics of these daily sequences, we focus on six seasonal summary statistics: seasonal rainfall total, rainfall frequency (days ≥ 0.1 mm), the mean daily intensity on wet days, the average length of dry spells, the mean wet-day persistence, and the monsoon onset date. The distribution of dry-spell lengths is skewed to the right because of the seasonal transition from the dry to the wet season, and the mean dry-spell length is biased by the dry season. We thus choose the median dry-spell length that is more indicative of post-onset conditions, and then take its natural logarithm to further reduce the skew of the distribution. Each summary statistic is computed at each of the 17 station locations.

To assess model performance at the Indramayu district level, we average each summary statistic over the 17

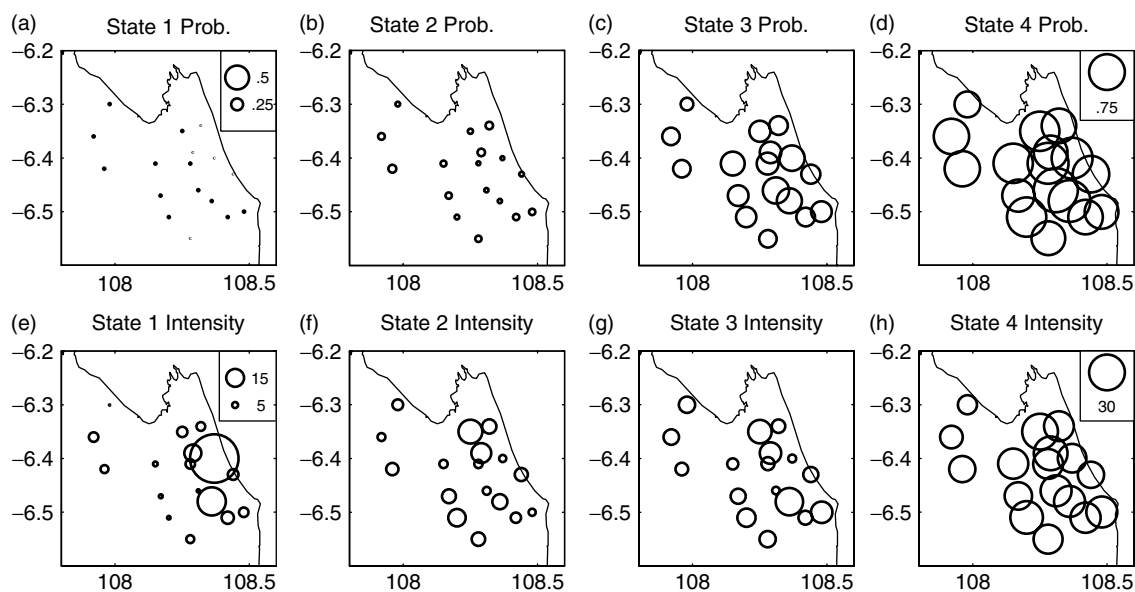


Figure 3. Rainfall probabilities (a)–(d) and mean intensities (e)–(h) associated with the 4-state model. Intensities are in mm/day.

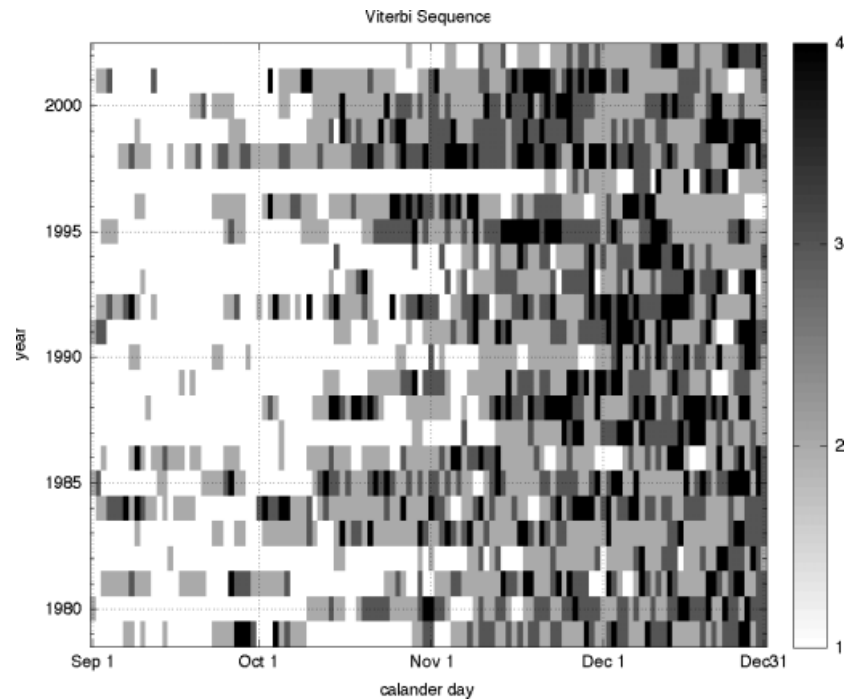


Figure 4. Estimated state sequence of 4-state model. Gray scale denotes the state. The states are ordered from driest (white) to wettest (black) as in Figure 3.

stations using a standardized anomaly index (SAI; Katz and Glantz, 1986). The SAI is computed by standardizing the interannual time series at each station (subtracting the mean and dividing by the STD) and then averaging the standardized anomalies spatially across the stations to form an index; it thus gives each station equal weight in the index.

4.4. Mean biases

Figure 5 shows the climatological (i.e. marginal) distributions of the SAI of each of the six rainfall statistics computed from the observations (panel a; 24 years) and simulations (panel b; 72 simulations \times 24 years); note that the observed distribution is purely interannual, whereas the simulated distribution contains both interannual and intraensemble variability. In order to identify biases in the simulations, the SAI was computed using the station means and STDs computed from the observations in *both* panels. Table I gives the observed and simulation means in physical units, averaged simply across stations, together with the percentage biases in the mean and STD. The STD in Table I was computed at each station and for each ensemble member individually, and then averaged.

Mean biases for seasonal total, rainfall frequency and mean intensity are negligible (about 1% or less), and about 10% for median dry-spell length and wet-day persistence. Onset dates are systematically too early by about 1 week on an average. The interquartile ranges (IQR), given by the boxes in Figure 5, are generally similar between the observed and simulated ensembles, while the tails of the simulated distributions are longer. The forecast distributions are generally less skewed

than their observed counterparts, with the median more centrally located in the IQR.

The bias in the interannual STD is given in Table I, averaged across members and stations. It is very small for seasonal total, rainfall frequency and onset date (<2%). Thus, the simulations generally do not suffer from insufficient interannual variability that is often encountered in simple stochastic weather generators (Katz and Parlange, 1998). However, the interannual STD is somewhat underestimated for mean intensity and wet-day persistence (10–15%), and overestimated for median dry-spell length (19%).

In general there will be some cancellation in these biases between stations, so that the biases at individual stations are larger than those given in Table I. For the mean bias, however, this cancellation was found to be minimal (the station-average of the absolute value of the mean bias at individual stations is always within 2% of the values in Table I). For the interannual STD the cancellation of the bias between stations is larger: the station-average of the absolute bias in STD increases to about 13% for seasonal total and rainfall frequency, 19% for mean intensity, 24% for median dry-spell length, 15% for wet-day persistence, and 6% for onset date.

Ensemble forecasts can be expressed most simply in terms of the ensemble mean, together with estimates of its uncertainty. Figure 6 shows quantile–quantile (Q–Q) plots of the interannual distributions of observed versus the forecast *ensemble mean*, again using the SAI without any bias correction. The 45° straight line would be obtained, approximately, if the two samples (forecast mean and observed data) came from the same distribution. The forecast distributions of seasonal total and

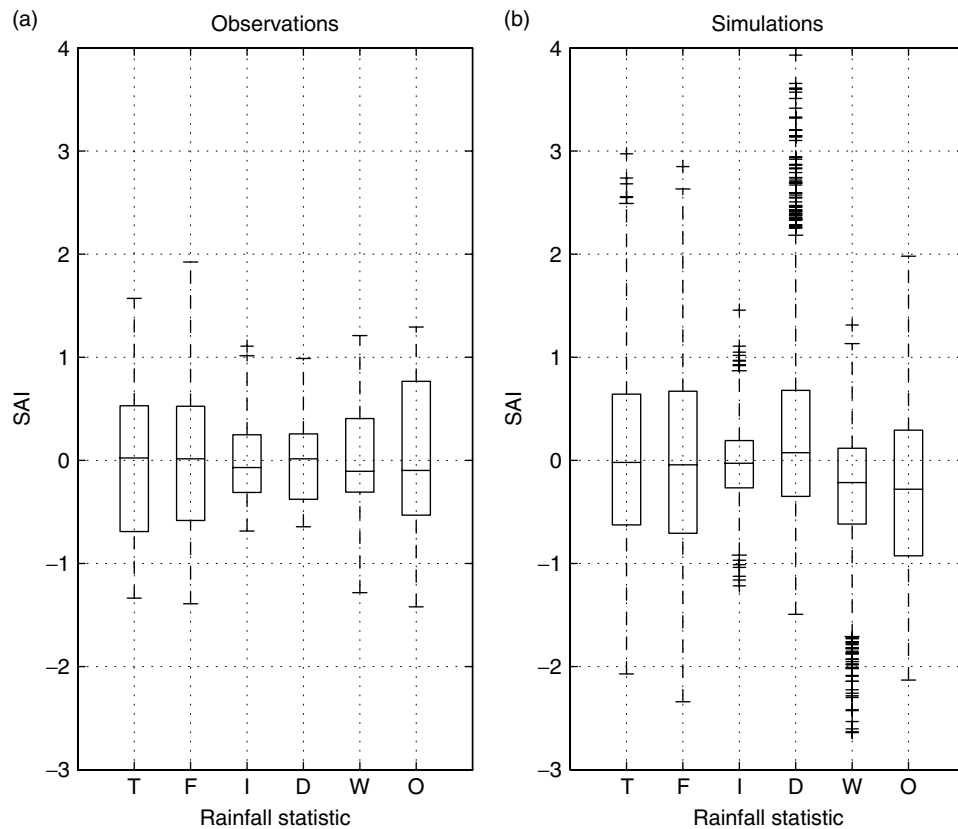


Figure 5. Box plots of standardized anomaly index (SAI) of (a) observations, and (b) raw forecasts across all 24 years. The rainfall statistics are seasonal rainfall total (T), daily rainfall frequency (F), mean daily intensity (I), median length of dry spells (D), mean wet-day persistence (W), and monsoon onset date (O). Boxes denote the median and interquartile range (IQR). Whiskers extend 1.5 IQR from box ends, with outliers denoted “+”. In panel (b) there are 72 simulations for each year.

Table I. Biases in the simulations, averaged over all 17 stations. The bias in the interannual standard deviation (STD) is computed for each simulated ensemble member separately, and then averaged. Percentage biases are expressed as percentages of the observed value.

	Obs mean	Mean bias	Mean bias (%)	STD bias (%)
Seas total (mm/d)	3.68	0.04	1.04	1.45
Frequency	0.21	0	-0.13	-0.31
Mean intensity (mm/d)	17.8	-0.15	-0.78	-14.06
Median dry spell (days)	2.89	0.26	10.35	19.01
Mean wet day persistence probability	0.39	-0.04	-9.06	-9.5
Onset date (days after 9/1)	60.81	-8.3	-13.4	1.55

(especially) rainfall frequency are quite accurate, while late onset-date forecasts tend to be too weak. The forecasts distributions of mean intensity and wet-day persistence ensemble mean are both much too narrow and thus severely lack forecast resolution.

4.5. Spatial coherence

Having assessed overall simulation biases and before turning to measures of forecast skill, we examine the spatial coherence of seasonal anomalies between stations. The amplitude of the SAI for a particular year depends on the size of the correlations between stations, and thus its variance gives a measure of spatial coherence of the field (Moron *et al.*, 2006). For relatively homogeneous

regions such as Indramayu, the spatial coherence provides a measure of potential predictability at the station scale (Moron *et al.*, 2006). The observed IQR of the SAI (Figure 5(a)) are largest for seasonal total, rainfall frequency, and monsoon onset date, while they are smallest for rainfall intensity and dry-spell length, suggesting higher predictability of the former quantities compared to the latter ones. Values of the variance of the SAI (VSAI) and the estimated number of spatial degrees of freedom (DOF; Moron *et al.*, 2006) are given in Table II. Each statistic is computed for each of the 72 simulations separately, with the mean and two-STD intraensemble range on either side reported in the table. As seen in previous studies of tropical rainfall (Moron *et al.*, 2006, 2007),

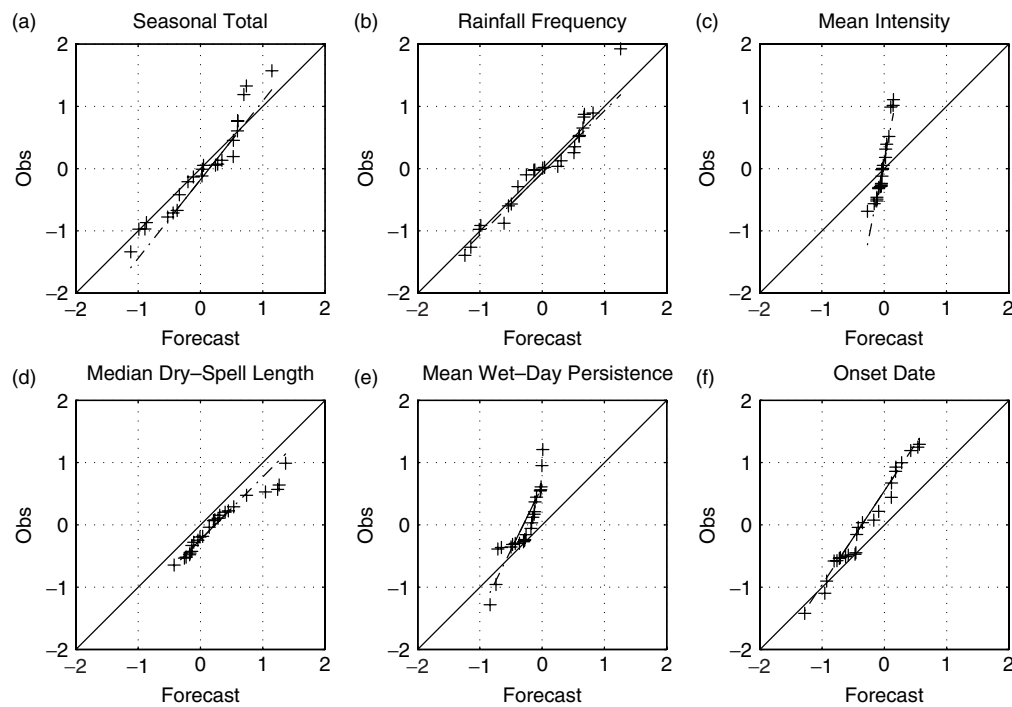


Figure 6. Quantile–quantile plots of the ensemble mean SAI of the forecasts against the observed values.

Table II. Spatial coherence statistics for the observations and simulations, in terms of the variance of the standardized anomaly index (VSAI) and estimated number of degrees of freedom (DOF). For the simulations, the mean and 2-sigma values are given (see text for details).

	VSAI-obs	VSAI-sim	DOF-obs	DOF-sim
Amount	0.6	0.81 ± 0.40	2.55	1.99 ± 0.58
Frequency	0.61	0.95 ± 0.44	2.46	1.34 ± 0.16
Mean intensity	0.25	0.11 ± 0.08	6.34	8.87 ± 1.58
Median dry-spell length	0.19	0.86 ± 0.77	8.17	3.47 ± 1.88
Mean wet day persistence probability	0.32	0.34 ± 0.26	5.93	5.39 ± 2.56
Onset date	0.62	0.68 ± 0.32	2.46	3.05 ± 1.43

spatial coherence of interannual anomalies in the station data is largest (high VSAI and low DOF) for rainfall frequency, closely followed by seasonal total, with mean intensity being much less coherent. Of the other statistics, onset date also exhibits high coherence, as found recently over Indonesia in the study of Moron *et al.* (2008a,b). The spatial coherence of the NHMM simulations generally follows the observed behavior. The simulations have a tendency to overestimate the coherence of seasonal total, rainfall frequency, and median dry-spell length and to underestimate it for rainfall intensity. However, except for the latter, these differences do not consistently fall outside the two-sigma confidence bounds of the simulations, and thus cannot be considered highly statistically significant.

4.6. Ensemble mean skill

Prior to assessing the skill of the forecasts, a simple bias correction was applied at each station to remove the biases in the mean and STD.

Skill is firstly assessed in terms of the forecast ensemble mean. Figure 7 shows forecast reliability and resolution in terms of the SAI of the verification given the forecast $[E(\text{obs}|\text{fct})]$, plotted against the forecast SAI. A bin-width of 0.2 was used to assign the forecasts to categories, for which the observed outcomes were averaged. Figure 7 thus shows the success of the forecasts binned into categories and is plotted in the same format as the Q–Q plots in Figure 6. In all panels the points lie fairly close to the diagonal, indicating reasonably reliable forecasts; i.e. the (bias corrected) forecasts for each bin indicated by a cross tend to be correct on average. On the other hand, there are large differences in forecast resolution between the six rainfall statistics, consistent with Q–Q plots in Figure 6. Rainfall frequency and seasonal total exhibit the most dispersion of the points along the diagonal, indicating that forecasts across the observed range of amplitude are indeed issued. In contrast, the mean intensity and wet-day persistence forecasts are clustered about the climatological mean indicating that the forecasts have no resolution. The dry-spell length

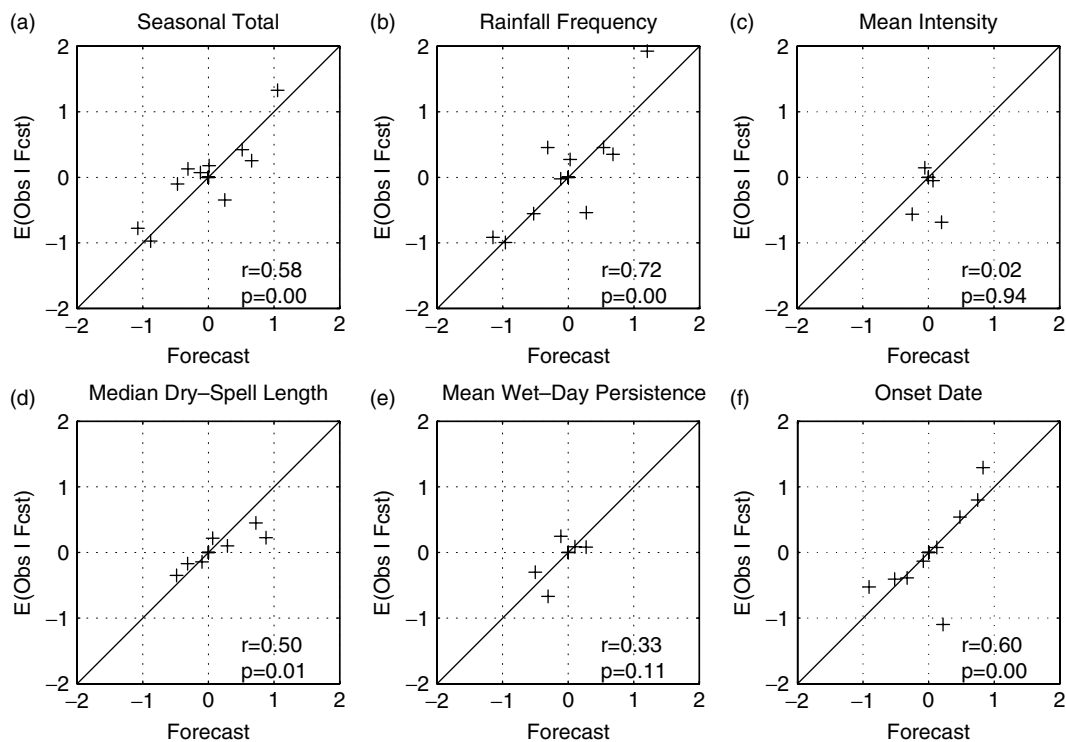


Figure 7. Reliability diagrams for the ensemble mean SAI of the forecasts. The 24 seasonal values of the ensemble mean forecast were binned into 10 classes of width 0.3σ , and the observed outcomes for each class averaged on the ordinate. The diagonal line gives the expected value for perfectly reliable forecasts.

forecasts also show too little forecast resolution. Forecast skill is a combination of the reliability and resolution of the forecasts. Values of the Pearson anomaly correlation and p -value, given in each panel of Figure 7, are generally consistent with these graphs. Thus, the highest anomaly correlation skills are achieved for rainfall frequency, followed by monsoon onset and seasonal total. Median dry-spell length is intermediary, while the forecasts of mean intensity and mean wet-day persistence are not significantly correlated with the verifications.

Figure 8 shows anomaly correlation skills at the individual stations. The stratification between the different rainfall statistics is quite clear in these plots. Inter-station differences may reflect data quality at each station, sampling issues as well as physical inhomogeneities – differences in skill across the small district of Indramayu do not appear systematic, although skill values at inland stations appear to be generally slightly lower.

4.7. Forecast spread

Risk management applications require estimates of forecast uncertainty, for which information contained in the ensemble spread may be applicable (e.g. Palmer, 2002). Figure 9 shows the observed SAI time series for each rainfall statistic, together with box plots depicting the forecast ensembles. The larger interannual variance of the SAI for seasonal total, rainfall frequency and onset-date is immediately apparent, indicative of the potential predictability in these three statistics. The skewness of the simulations of median dry-spell length is also apparent,

which may account for the overestimation of its variance in the simulations (Table II).

Provided an interannual signal is present in the observed SAI, a skillful forecast ensemble should bracket the observed value, such that the probability of the observation given the forecast is as large as possible (Murphy and Winkler, 1987). There is visible evidence that the forecasts of seasonal total, rainfall frequency and onset-date contain skill. Various forecast verification metrics have been developed to quantify the skill of probabilistic forecasts (e.g. Jolliffe and Stephenson, 2003). The continuous ranked probability score (CRPS; Hersbach, 2000), for example, is a squared error metric that measures the distance between the cumulative distribution function (CDF) of the forecast and the verifying observation; the latter ‘‘CDF’’ takes the form of a step function at the value of the observation. Expressed with respect to a baseline given by the CRPS of the climatological forecast distribution, the median CRPS scores (across years) of the six SAI quantities (seasonal totals, rainfall frequency, mean intensity, median dry-spell length, mean wet day persistence, onset date) are -2.48 , 7.74 , -20.24 , -29.58 , -29.20 , and 6.40% , respectively. Negative values denote a forecast worse than climatology, with a perfect forecast given by $+100\%$. Only rainfall frequency and onset date yield skill better than climatology. The CRPS scores were also computed at individual stations. At the station level, the downscaled forecasts were only found to exhibit CRPS scores values better than climatology for onset date; these are plotted in Figure 10.

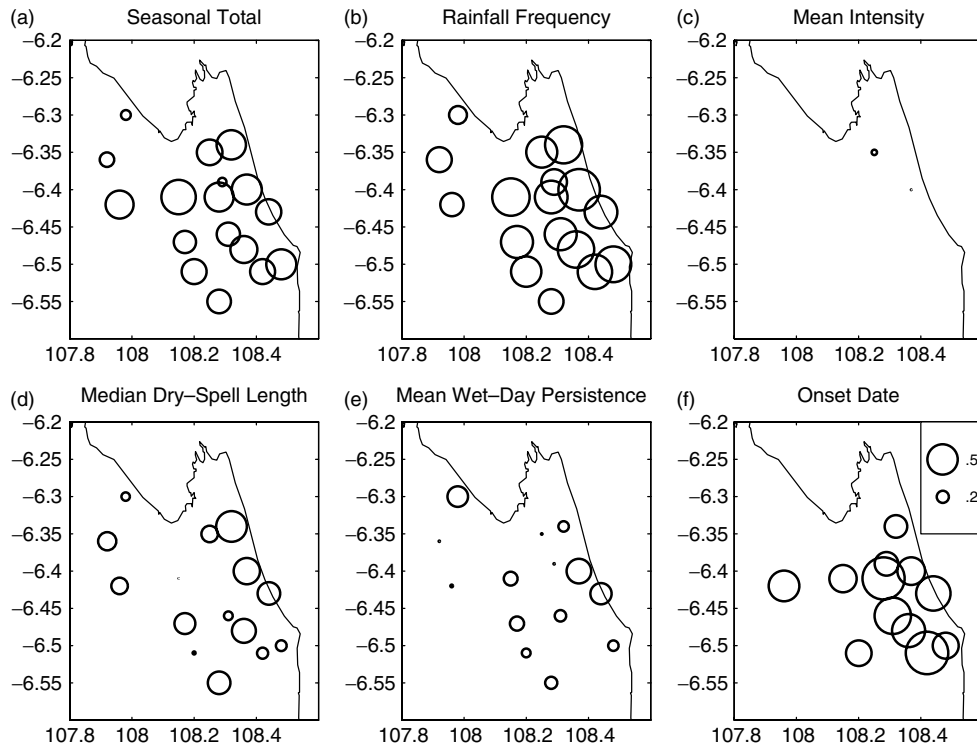


Figure 8. Correlation skills of hindcasts: (a) seasonal rainfall total; (b) rainfall frequency; (c) mean daily intensity; (d) median length of dry spells; (e) mean wet-day persistence; (f) monsoon onset date. Circle diameter is proportional to magnitude of the correlation. Negative correlations omitted.

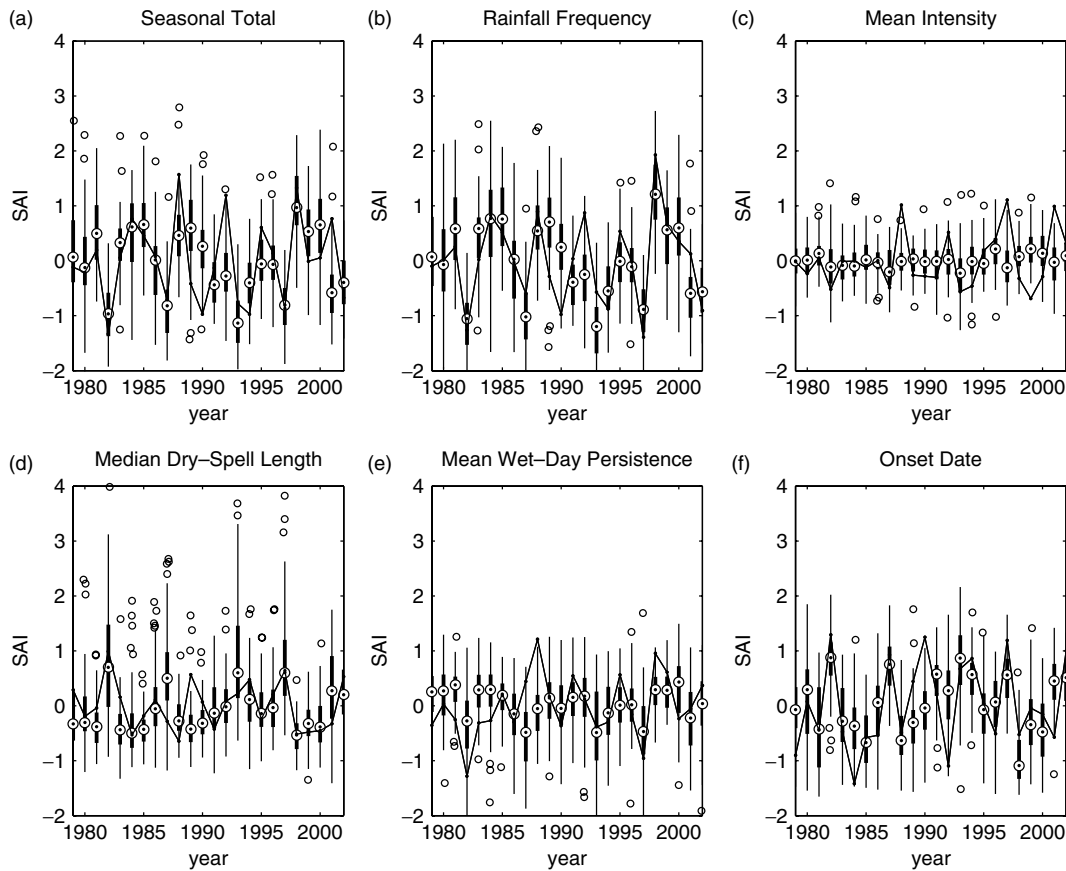


Figure 9. Box plots of standardized anomaly index (SAI) of forecasts, together with observations (solid line). Boxes denote the interquartile range (IQR), about the median (circle with dot). Whiskers extend 1.5 IQR from box ends, with outliers denoted "o". The median CPRS scores of the six SAI quantities are -2.48, 7.74, -20.24, -29.58, -29.20, and 6.40%.

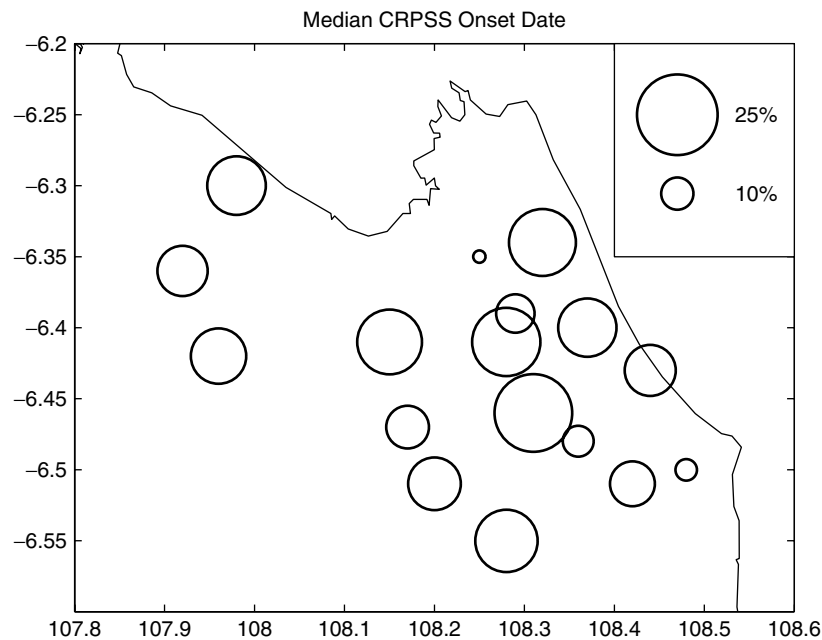


Figure 10. Continuous ranked probability skill (CRPS) scores for monsoon onset date. Circle diameter is proportional to magnitude of the CRPS score (in %).

To be well calibrated, the spread of the forecast distributions should be such that the IQR “prediction interval” boxes in Figure 9 bracket the observation in 50% of years; values below indicate too little spread (boxes too narrow), and values above 50% imply too much spread in the forecast distribution (boxes too wide). For the six SAI quantities in Figure 9, the percentage of observed years within the simulated IQR (i.e. the capture rates) are 58, 62, 37, 71, 46, and 46% respectively. Thus in most cases the forecasts are reasonably well calibrated; there is too little spread for rainfall intensity and too much for dry-spell length.

4.8. Conditional exceedance probabilities

To visualize the reliability of the forecasts, the individual ensemble members can be treated as estimates of quantiles of the forecast distribution (Mason *et al.*, 2007). For example, given only one ensemble member, there should be a 50% probability that the observed value exceeds the forecast, regardless of the value being forecast. Thus, a graph of this “conditional exceedance probability” (CEP) against the forecast rainfall should be a horizontal line with $CEP = 0.5$. Figure 11 shows the CEP curves for each of the 72 ensemble members, calculated across all years using generalized linear regression (Mason *et al.*, 2007). For seasonal total, for example, these are ranked from driest to wettest, from the top to bottom in Figure 11(a).

The CEP curves for mean intensity, median dry-spell length and wet-day persistence all lie close to the climatological probability of exceedance (thin line), showing that these forecasts do not deviate much from climatology; this is consistent with the large negative CPRS scores for these rainfall statistics. On the other hand, the CEP curves for seasonal total, rainfall frequency

and onset date slope less extend over a larger range of SAI and are more evenly spaced. There is still a general tendency for the slopes to be negative, except for onset date, indicating that the forecasts tend to be over confident. However, the distributions are noisy, indicating considerable sampling variability associated with the short 24-year time series.

4.9. A real-time forecast

Figure 12 presents an example of a forecast distribution made for the 2007 SOND season, expressed in terms of probability of exceedance. The figure shows CDFs, smoothed using a kernel density estimator, for the historical observed (solid) and 1979–2002 retrospective forecast (dotted) climatological distributions, and the 2007 forecast distribution (dashed).

The observed and simulated climatological distributions are similar in all cases indicating no serious biases in the retrospective forecasts over the 1979–2002 period, recalling that the bias in the mean and variance has been removed from each SAI (Section 4.6). The 2007 forecast exhibits a dry tendency, with lower probabilities of exceeding a given threshold of seasonal amount, rainfall frequency, and wet-day persistence, and higher probabilities of exceeding a given threshold of median dry-spell length and onset date. The exceedance probabilities of the forecast for rainfall intensity and wet-spell length also deviate from climatology, despite the lack of skill in these quantities.

5. Conclusions

5.1. Summary

We have demonstrated the methodology and evaluated the skill of downscaled rainfall forecasts over Indramayu

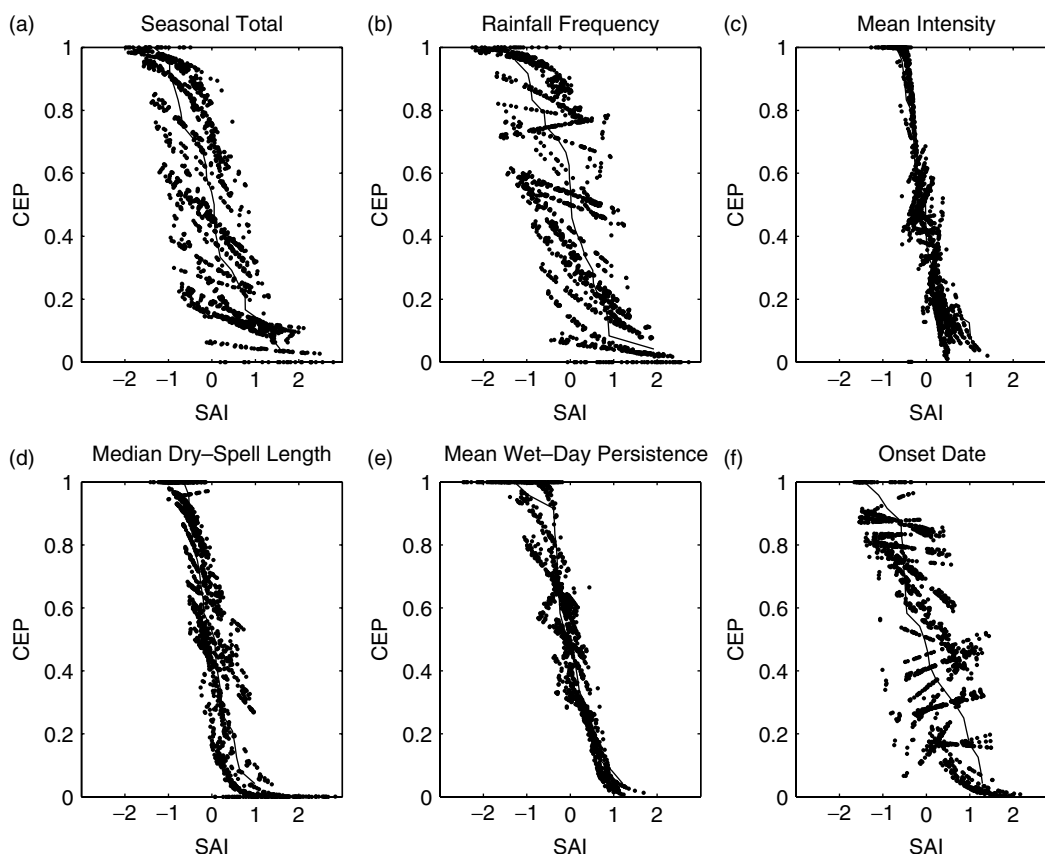


Figure 11. Conditional exceedance probabilities of SAI. Curves denote probability that the observed value in a particular year exceeds the predicted value for that year, for a given ensemble member. Thin continuous line denotes the exceedance probabilities of the observations.

district, West Java, during the September–December monsoon onset season, using a combined GCM–NHMM approach. The quality of the cross-validated retrospective forecasts was assessed for six rainfall summary statistics computed from 72-member daily-rainfall-sequence simulations: seasonal rainfall total, daily rainfall frequency, mean daily intensity, median length of dry spells, wet-day persistence, and monsoon onset date.

Mean biases of the station-averaged rainfall simulations (Table I; Figure 5) are under 1% for seasonal total, rainfall frequency and mean intensity. The simulations overestimate the lengths of dry spells and underestimate the lengths of wet spell by about 8%, and simulated onset date is premature by about a week. Interannual STD are accurate (within 2%) for seasonal total, rainfall frequency and onset date; they are underestimated for rainfall intensity (14%) and wet-day persistence (10%), and overestimated for median dry-spell length (19%).

Various measures of skill of the forecasts were considered. In terms of anomaly correlation of the ensemble mean, the SAI over the stations reaches 0.71 for rainfall frequency, 0.61 for onset date, 0.58 for seasonal total, and 0.50 for median dry-spell length. Neither rainfall intensity nor wet-day persistence exhibit skill. The ensemble mean forecasts exhibit encouraging reliability for all quantities (i.e. the expectation of the observations conditioned on the forecasts is accurate), but with good forecast resolution only for rainfall frequency, seasonal total and onset

date (Figure 7). At the station level (Figure 8), anomaly correlations are most consistently high for rainfall frequency.

Regarding the forecast distributions, the spread of the distribution is generally reasonable: somewhat too broad for seasonal total, rainfall frequency and (particularly) dry-length, and too narrow for wet-spell length, onset date and (markedly) rainfall intensity (Figure 9). Probabilistic skill values using the CRPS score show better-than-climatology values only for SAI of onset date and rainfall frequency. At the station scale, only the monsoon onset date shows positive CRPS scores against a climatological benchmark (Figure 10). CEP of the individual ensemble members (CEPs; Figure 11) indicate the highest reliability for onset date, followed by rainfall frequency and seasonal total. However, clear deficiencies are visible, with a general tendency toward overconfidence of the ensembles (i.e. negative CEP slopes) in all quantities except onset date. The CEP curves are noisy and the short length of the verification series (24 points) is a limiting factor. An example probabilistic forecast was made for 2007, expressed in terms of exceedance probabilities (Figure 12).

5.2. Discussion

The goal of this paper has been to assess the suitability of the NHMM as a downscaling technique to obtain daily rainfall sequences conditioned on seasonal forecasts.

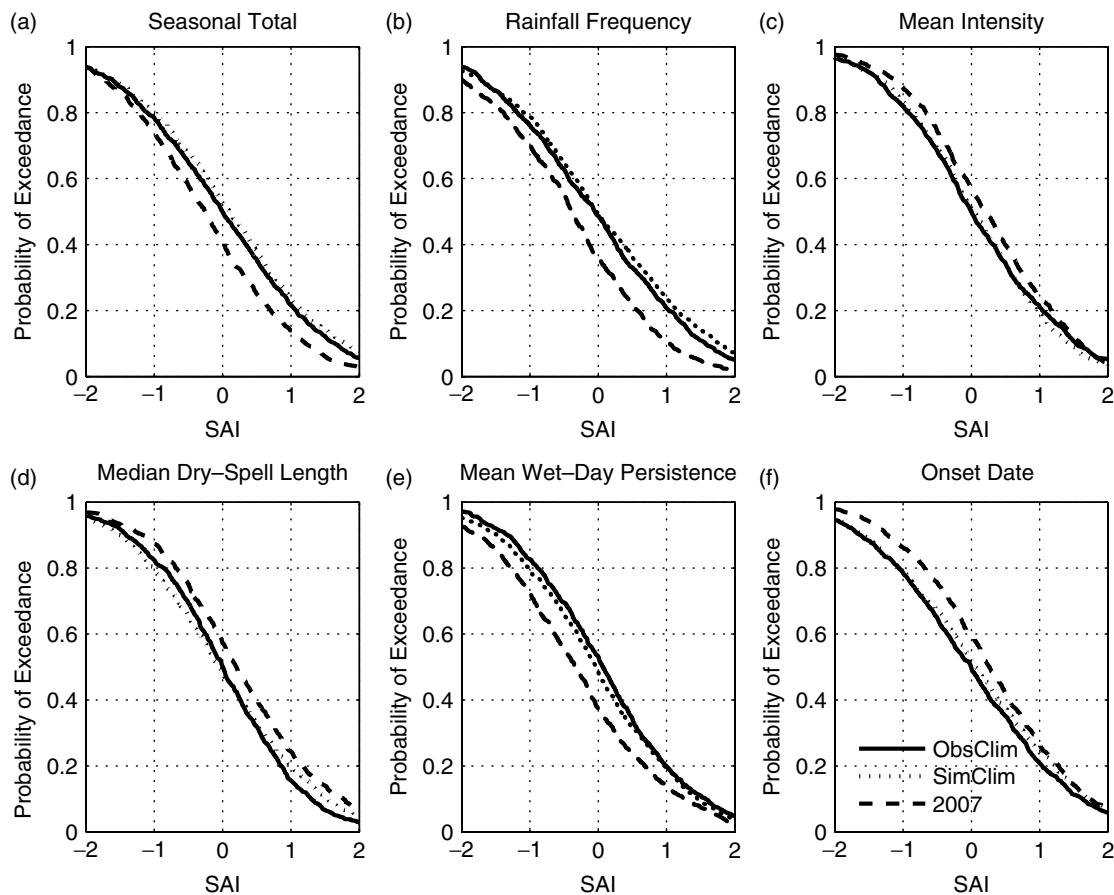


Figure 12. Probability of exceedance for 2007 forecast of SAI. Key: solid—observations, dotted—hindcasts, dashed—2007 forecast.

Taken together, the set of forecast metrics examined here provide an overall picture of forecast quality.

Levels of skill can be differentiated according to the district average versus individual stations, and anomaly correlation versus the CRPS probabilistic score. In terms of the anomaly correlation of ensemble mean forecasts for district average (here SAI), results are consistent with our previous studies of seasonal predictability of tropical rainfall (Moron *et al.*, 2006, 2007, 2008c,d; Robertson *et al.*, 2006, 2007), with highest skill for rainfall frequency and lowest for mean rainfall intensity. Onset date skill predictability is consistent with the analysis of Moron *et al.* (2008a) who found seasonal predictability of seasonal total to be largely associated with onset date over Indonesia. As argued by Moron *et al.* (2006, 2007), interannual anomalies in rainfall frequency are expected to be more spatially coherent than those of mean daily intensity because tropical mesoscale convective clusters can produce big differences in rainfall intensity over short distances; this spatial noise in rainfall intensity is largely unpredictable and can contaminate even seasonal rainfall totals, while being largely filtered out in the rainfall frequency field. The analyses of Moron *et al.* (2008a,b) indicate that monsoon onset date over Indonesia contains a substantial large-scale spatially coherent signal, while post-onset rainfall over Java is dominated by small-scale features and is much less sensitive to ENSO. The anomaly correlation skill seen at

the district level is largely reproduced at the station level as well. Most striking, however, is the lack of CRPS skill except for rainfall frequency and onset date at the district level, and only onset date at the individual stations. The latter result was found robust to details of how the CRPS score was estimated and is encouraging for climate risk management applications where onset date is a critical factor in crop planting.

The cross-validated anomaly correlation skill for the SAI of seasonal total is comparable to raw correlations between the station average rainfall and the GCM principal component predictors. This is encouraging because the summary statistics of the simulations were computed a posteriori from the cross-validated NHMM daily rainfall sequences. While regression models built directly on seasonal statistics can be expected to outperform the NHMM, the latter is motivated by the need for the daily sequences for crop modeling, etc. The NHMM provides an attractive model for downscaling to stochastic daily rainfall sequences because it extends the conventional stochastic daily weather generator to a multistation setting in which the weather-state variable spans all the stations, providing a less-noisy bridge to larger-scale climatic forcing. However, competing methodologies such as resampling methods have been shown to perform with similar skill (Moron *et al.*, 2008d).

The generally low levels of bias in the simulations is encouraging, with accurate levels of interannual variance

for the more skillful quantities, i.e. onset date, seasonal total and rainfall frequency. However, the mean onset date of the simulations is premature by about 1 week. This is probably largely due to biases in the GCM predictors since use of reanalysis-based predictors was largely able to remove this bias (not shown). Further work is required to address this issue, before the GCM–NHMM simulated daily rainfall sequences could be used to drive crop models, for example, which may be sensitive to timing of rainy season onset. Our results are likely to be typical of what might be expected for downscaling over other areas of Indonesia where the large-scale climate signal on onset is pronounced, such as Sumatra, Java, Nusa Tenggara, and Timor.

Acknowledgements

We are grateful to Rizaldi Boer for assistance with the rainfall station data, and to Sergey Kirshner, Simon Mason, and Padhraic Smyth for helpful discussions. The NHMM software (MVNHMM) was developed by S. Kirshner and can be obtained free of charge from <http://www.stat.purdue.edu/~skirshne/MVNHMM/>. 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. The computing for this project was partially provided by a grant from the NCAR CSL program to the IRI.

References

- Aldrian E, Dümenil-Gates L, Widodo FH. 2007. Seasonal variability of Indonesian rainfall in ECHAM4 simulations and in the reanalyses: the role of ENSO. *Theoretical and Applied Climatology* **87**: 41–59.
- Aldrian E, Sein D, Jacob D, Dümenil-Gates L, Podzun R. 2005. Modeling Indonesian rainfall with a coupled regional model. *Climate Dynamics* **25**: 1–17.
- Aldrian E, Susanto RD. 2003. Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *International Journal of Climatology* **23**: 1435–1452.
- Boer R, Wahab I. 2007. Use of seasonal surface temperature for predicting optimum planting window for potato at Pengalengan, West Java, Indonesia. In *Climate Prediction and Agriculture: Advance and Challenge*, Sivakumar MVK, Hansen J (eds). Springer: New York; 135–141.
- Charles SP, Bates BC, Smith IN, Hughes JP. 2004. Statistical downscaling from observed and modelled atmospheric fields. *Hydrological Processes* **18**: 1373–1394.
- Charles SP, Bates BC, Viney NR. 2003. Linking atmospheric circulation to daily rainfall patterns across the Murrumbidgee River Basin. *Water Science Technology* **48**: 233–240.
- Forney GD Jr. 1978. The Viterbi algorithm. *Proceedings of the IEEE* **61**: 268–278.
- Goddard L, Mason SJ, Zebiak SE, Ropelewski CF, Basher R, Cane MA. 2001. Current approaches to seasonal-to-interannual climate predictions. *International Journal of Climatology* **21**: 1111–1152.
- Gong X, Barnston AG, Ward MN. 2003. The effect of spatial aggregation on the skill of seasonal precipitation forecasts. *Journal of Climate* **16**: 3059–3071, DOI: 10.1175/1520-0442(2003)016<3059:TEOSAO>2.0.CO;2.
- Giannini A, Robertson AW, Qian JH. 2007. A role for tropical tropospheric temperature adjustment to ENSO in the seasonality of monsoonal Indonesia precipitation predictability. *Journal of Geophysical Research-Atmospheres* **112**: D16110, DOI:10.1029/2007JD008519.
- 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. *Journal of the Meteorological Society of Japan* **80**: 285–310.
- Hansen JW, Ines AVM. 2005. Stochastic disaggregation of monthly rainfall data for crop simulation studies. *Agricultural and Forest Meteorology* **131**: 233–246.
- Haylock M, McBride J. 2001. Spatial coherence and predictability of Indonesian wet season rainfall. *Journal of Climate* **14**: 3882–3887.
- Hersbach H. 2000. Decomposition of the Continuous Ranked Probability Score for ensemble prediction systems. *Weather and Forecasting* **15**: 559–570.
- Hughes JP, Guttorp P. 1994. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research* **30**: 1535–1546.
- Ines AVM, Hansen JW. 2006. Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology* **138**: 44–53.
- Jolliffe IT, Stephenson DB (eds). 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons: Chichester, ISBN 0-471-49759-2.
- Juneng L, Tangang FT. 2005. Evolution of ENSO-related rainfall anomalies in Southeast Asia region and its relationship with atmosphere-ocean variations in Indo-Pacific sector. *Climate Dynamics* **25**: 337–350.
- Katz RW, Glantz MH. 1986. Anatomy of a rainfall index. *Monthly Weather Review* **114**: 764–771.
- Katz RW, Parlange MB. 1998. Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate* **11**: 591–601.
- Kirshner, S. 2005. Modeling of multivariate time series using hidden Markov models. Ph.D. thesis, University of California, Irvine.
- Li S, Goddard L, DeWitt DG. 2007. Predictive skill of AGCM seasonal climate forecasts subject to different SST prediction methodologies. *Journal of Climate* **21**: 2169–2186.
- Mason SJ, Galpin JS, Goddard L, Graham NE, Rajartnam B. 2007. Conditional exceedance probabilities. *Monthly Weather Review* **135**: 363–372.
- Moron V, Robertson AW, Boer R. 2008a. Spatial coherence and seasonal predictability of monsoon onset over Indonesia. *Journal of Climate* (in press).
- Moron V, Robertson AW, Qian J-H. 2008b. Local versus large-scale characteristics of monsoon onset and post-onset rainfall over Indonesia. *Climate Dynamics* sub judice (in review).
- Moron V, Robertson AW, Ward MN. 2008c. Weather types and rainfall over Senegal. Part I: observational analysis. *Journal of Climate* **21**: 266–287.
- Moron V, Robertson AW, Ward MN. 2008d. Weather types and rainfall in Senegal. Part II: Downscaling of GCM Simulations. *Journal of Climate* **21**: 288–307.
- Moron V, Robertson AW, Ward MN. 2006. Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Monthly Weather Review* **134**: 3468–3482.
- Moron V, Robertson AW, Ward MN, Camberlin P. 2007. Spatial coherence of tropical rainfall at regional scale. *Journal of Climate* **20**: 5244–5263.
- Murphy AH, Winkler RL. 1987. A general framework for forecast verification. *Monthly Weather Review* **115**: 1330–1338.
- Naylor RL, Falcon W, Wada N, Rochberg D. 2002. Using El-Niño Southern Oscillation climate data to improve food policy planning in Indonesia. *Bulletin of Indonesian Economic Studies* **38**: 75–91.
- Palmer TN. 2002. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society* **128**: 747–774.
- Robertson AW, Ines AVM, Hansen JW. 2007. Downscaling of seasonal precipitation for crop simulation. *Journal of Applied Meteorology and Climatology* **46**: 677–693.
- Robertson AW, Kirshner S, Smyth P. 2004. Downscaling of daily rainfall occurrence over Northeast Brazil using a Hidden Markov Model. *Journal of Climate* **17**: 4407–4424.
- Robertson AW, Kirshner S, Smyth P, Charles SP, Bates BC. 2006. Subseasonal-to-interdecadal variability of the Australian monsoon

- over North Queensland. *Quarterly Journal of the Royal Meteorological Society* **132**: 519–542.
- Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dumenil L, Esch M, Giorgetta M, Schlese U, Schulzweida U. 1996. The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate. Technical Report 218. Max Planck Institute for Meteorology: Hamburg, Germany; 90.
- Sivakumar MVK. 1988. Predicting rainy season potential from the onset of rains in southern Sahelian and Sudanian climatic zones of West Africa. *Agricultural and Forest Meteorology* **42**: 295–305.
- van den Dool HM. 1994. Searching for analogues, how long must we wait? *Tellus* **46A**: 314–324.
- Wilks DS. 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology* **93**: 153–169.