

## Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland

and similar papers at [core.ac.uk](http://core.ac.uk)

bro

provided by Columbia Unive

<sup>1</sup>*International Research Institute for Climate and Society,  
The Earth Institute at Columbia University, Palisades, New York, USA*

<sup>2</sup>*Donald Bren School of Information and Computer Science,  
University of California, Irvine, California, USA*

<sup>3</sup>*CSIRO Land and Water, Perth, Australia*

(Received 25 April 2005; revised 12 September 2005)

### SUMMARY

Daily rainfall occurrence and amount at 11 stations over North Queensland are examined for summers 1958–1998, using a Hidden Markov Model (HMM). Daily rainfall variability is described in terms of the occurrence of five discrete ‘weather states’, identified by the HMM. Three states are characterized respectively by very wet, moderately wet, and dry conditions at most stations; two states have enhanced rainfall along the coast and dry conditions inland. Each HMM rainfall state is associated with a distinct atmospheric circulation regime. The two wet states are accompanied by monsoonal circulation patterns with large-scale ascent, low-level inflow from the north-west, and a phase reversal with height; the dry state is characterized by circulation anomalies of the opposite sense. Two of the states show significant associations with midlatitude synoptic waves.

Variability of the monsoon on time-scales from subseasonal to interdecadal is interpreted in terms of changes in the frequency of occurrence of the five HMM rainfall states. Large subseasonal variability is identified in terms of active and break phases, and a highly variable monsoon onset date. The occurrence of the very wet and dry states is somewhat modulated by the Madden–Julian oscillation. On interannual time-scales, there are clear relationships with the El Niño–Southern Oscillation and Indian Ocean sea surface temperatures (SSTs). Interdecadal monsoonal variability is characterized by stronger monsoons during the 1970s, and weaker monsoons plus an increased prevalence of drier states in the later part of the record.

Stochastic simulations of daily rainfall occurrence and amount at the 11 stations are generated by introducing predictors based on large-scale precipitation from (a) reanalysis data, (b) an atmospheric general circulation model (GCM) run with observed SST forcing and (c) antecedent June–August Pacific SST anomalies. The reanalysis large-scale precipitation yields relatively accurate station-level simulations of the interannual variability of daily rainfall amount and occurrence, with rainfall intensity less well simulated. At some stations, interannual variations in 10-day dry-spell frequency are also simulated reasonably well. The interannual quality of the simulations is markedly degraded when the GCM simulations are used as inputs, while antecedent Pacific SST inputs yield an anomaly correlation skill comparable to that of the GCM.

**KEYWORDS:** Daily rainfall Hidden Markov model

### 1. INTRODUCTION

The Australian monsoon brings summer rainfall to North Queensland, and has been described extensively (Troup 1961; McBride 1987; Manton and McBride 1992; Suppiah 1992). A monsoonal upper-level anticyclone becomes established over northern Australia as the region of intense convection over Indonesia moves south of the equator during January and February. At low levels, the dry south-east trades are displaced poleward as equatorial westerlies bring moisture into a region of cyclonic vorticity—the monsoon trough—over northern Australia.

At the planetary scale, the summertime circulation exhibits a baroclinic Rossby wave structure consistent with the interactive Rodwell–Hoskins mechanism of monsoons (Rodwell and Hoskins 1996, 2001; Chou and Neelin 2003). Chen (2003) has shown that there is a near-Sverdrup vorticity balance between stretching at low levels (ascent) over the western Pacific warm pool, and low-level poleward advection of planetary vorticity. With the low-level cyclone situated over Australia, the equatorward

\* Corresponding author: Correspondence address: IRI - Monell 230, 61 Route 9W, Palisades, NY 10964, USA.  
e-mail: awr@iri.columbia.edu

low-level motion on its western flank is consistent with descent and the relatively cooler waters of the eastern Indian Ocean.

At the regional scale, North Queensland comes under the influence of both the planetary-scale region of monsoonal ascent to the east, as well as the southward-displaced convergence zone over Indonesia. Thus, summertime rainfall can be seen as a combination of north–south seasonal displacements of the Indonesian convergence zone, together with an essentially east–west monsoonal circulation. Moisture-laden on-shore winds cross the warm Coral Sea producing large amounts of rainfall along the eastern flanks of the coastal ranges.

Although this dynamical picture of the monsoon emerges in the January–February long-term mean, the monsoon is not a steady phenomenon and exhibits considerable subseasonal and interannual-to-interdecadal variability. Its onset is highly variable, ranging by one estimate from 23 November to 27 January at Darwin (12°S, 130°E), with a mean onset date of 24 December over the period 1952–82 (Holland 1986). Several authors have studied subseasonal variability of the Australian monsoon in terms of the onset date and active/break periods within the monsoon season (Troup 1961; Holland 1986; Hendon and Liebmann 1990a; Drosowsky 1996), and the subject has recently been reviewed comprehensively by Wheeler and McBride (2005). The Madden–Julian oscillation (MJO; Madden and Julian 1971) is known to exert an impact on the Australian monsoon (McBride 1987; Hendon and Liebmann 1990b), though the extent of this impact is debated. Hendon and Liebmann (1990a) found an important MJO impact on monsoon onset, defined at Darwin in terms of the first occurrence of wet westerly winds at 850 hPa. However, Drosowsky (1996) found no clear relationship between westerly winds and rainfall on subseasonal time-scales, no dominant time-scale in the length of active periods, nor in the recurrence intervals between them. Cold surges from both the southern hemisphere storm track as well as (more controversially) from the South China Sea are both believed to play important roles in subseasonal variability of the Australian monsoon (Suppiah 1992).

On interannual time-scales, the relationship between El Niño/Southern Oscillation (ENSO) and Australian rainfall has been the subject of numerous studies, although correlations between the Southern Oscillation Index (SOI) and summer rainfall over Queensland are relatively weak (McBride and Nicholls 1983; Nicholls 1989; Syktus *et al.* 2003). On longer time-scales, Pittock (1975) demonstrated a dry phase between 1913 and 1945, followed by a wet phase, which appeared to end around 1978 and has been followed by renewed drier conditions since. The relationship between SOI and summer rainfall exhibits decadal variations over the past century (Suppiah 2004).

The aim of this paper is to examine summertime variability of *daily* rainfall recorded on a network of stations over North Queensland and to relate this variability to the atmospheric circulation on time-scales of daily to interdecadal. The concept of planetary circulation regimes (Legras and Ghil 1985), sometimes called weather regimes (Reinhold and Pierrehumbert 1982), has been introduced in attempting to connect the observations of persistent and recurring midlatitude flow patterns with large-scale atmospheric dynamics. These circulation regimes have intrinsic time-scales of several days to a week, and exert a control on local weather (e.g. Robertson and Ghil 1999). Synoptic scales are key to the subseasonal variability of the Australian monsoon, and the circulation regime paradigm provides a basis for connecting local weather with the seasonal-scale monsoon and its interannual-to-interdecadal variability. Any changes over time in the frequency of occurrence, or structure, of circulation regimes will result in longer time-scale rainfall variability. This paradigm of climate variability provides a counterpart to wave-like decompositions of atmospheric variability, allowing the

connection to be made with oscillatory phenomena (Ghil and Robertson 2002), such as the MJO.

Circulation regimes have most often been defined in terms of clustering, fuzzy (Mo and Ghil 1987) or hierarchical (Cheng and Wallace 1993), maxima in the probability density function (PDF) of the large-scale low-frequency flow (Molteni *et al.* 1990; Kimoto and Ghil 1993a,b), by quasi-stationarity (Ghil and Childress 1987, section 6.4; Vautard 1990), and more recently using a probabilistic Gaussian mixture model (Smyth *et al.* 1999). However, from the perspective of local rainfall, it is advantageous to define the regimes in terms of the local weather that they give rise to. Such regimes may provide the basis to predict aspects of ‘weather-within-climate’ at seasonal lead times.

The Hidden Markov Model (HMM) provides a probabilistic framework for factoring the joint distribution of daily rainfall at a network of stations, by introducing a small set of underlying discrete rainfall states. These states can facilitate interpretation in terms of accompanying circulation anomalies, while the model can be used to generate large numbers of stochastic surrogate sequences of daily rainfall (Hughes and Guttorp 1994). The HMM has recently been shown to be applicable to tropical rainfall occurrence, over north-east Brazil, where the dominant rainfall states were found to be associated with circulation regimes characterized by meridional displacements of the intertropical convergence zone (ITCZ) (Robertson *et al.* 2004a).

In this paper, we apply the HMM to 11 station records of daily rainfall over North Queensland (1958–98) during the October–April summer season, considering both rainfall occurrence and amount. The rainfall dataset and its climatology are described in section 2. The HMM for rainfall occurrence and amount is introduced in section 3. We then apply the HMM in section 4, and use the rainfall states to interpret the subseasonal time-scale variability of the monsoon in terms of circulation regimes, by constructing composites of atmospheric circulation data. Interannual and interdecadal variability of rainfall-state occurrence are then discussed, along with the accompanying rainfall amount changes and relationships with sea surface temperatures (SSTs). We then construct ‘predictive’ non-homogeneous HMMs in section 5 to address seasonal predictability and downscaling from reanalysis and GCM-generated large-scale precipitation fields, as well as predictors based on antecedent Pacific SST anomalies. The paper concludes with a discussion in section 6 and conclusions in section 7.

## 2. OBSERVED RAINFALL DATA

We use daily rainfall amounts at 11 stations over North Queensland, for the 197-day 1 October–15 April season, 1958–1998; leap years end on 14 April. These data were obtained from the Patched Point Dataset (PPD) (Jeffrey *et al.* 2001). The PPD combines observed Australian Bureau of Meteorology (BoM) daily rainfall records with high quality and rigorously tested data infilling and deaccumulation of missing or accumulated rainfall. Four of the 11 stations have more than 10% of missing days infilled: station 2 (18.0%), station 4 (41.6%), station 8 (20.9%), and station 9 (10.3%). Of these, station 8 might be viewed with some caution, since it has considerable infilling and is situated in a region of orography. However, these stations are not found to be outliers in our analyses.

Figure 1 shows the locations of the 11 stations together with the October–April climatological daily probability of rainfall occurrence (defined as days with  $\geq 1$  mm day<sup>-1</sup>) and rainfall amount on days with rain. We refer to the region spanned by this network of stations as North Queensland, throughout the paper. Climatological rainfall occurrence

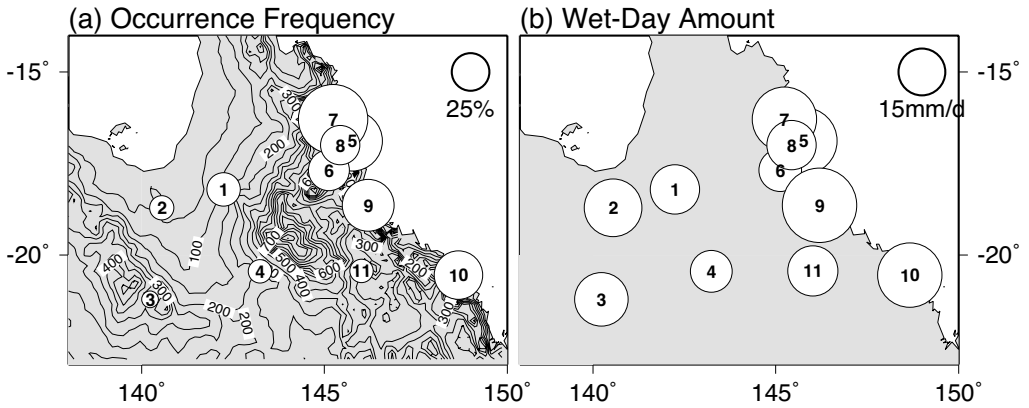


Figure 1. Rainfall station locations with topographic (50-metre) contours. Circle radius denotes (a) the 1 October–15 April climatological daily rainfall probability 1958–1998, and (b) the corresponding mean wet-day amount. The stations are: (1) Croydon (18.2°S, 142.2°E), (2) Julia Creek (18.7°S, 140.5°E), (3) Cloncurry (21.2°S, 140.2°E), (4) Richmond (20.4°S, 143.2°E), (5) Cairns (16.9°S, 145.8°E), (6) Mount Garnett (17.7°S, 145.1°E), (7) Mossman South (16.3°S, 145.2°E), (8) Mareeba (17.0°S, 145.4°E), (9) Ingham (18.6°S, 146.2°E), (10) Proserpine (20.5°S, 148.7°E), (11) Charters Towers (20.4°S, 146.0°E).

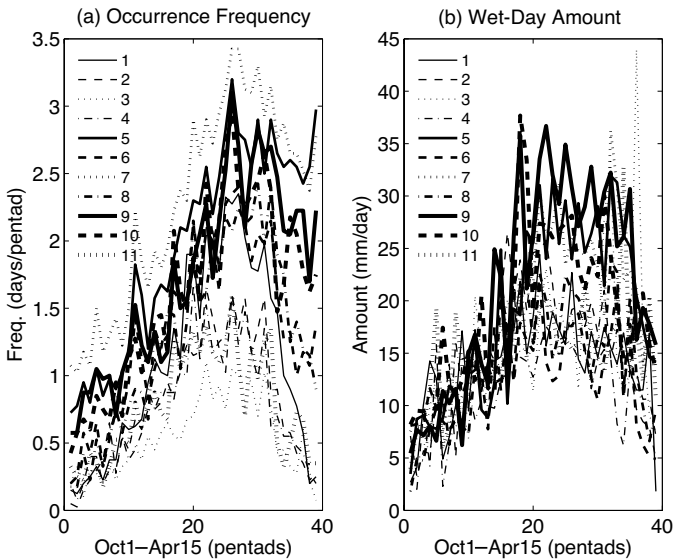


Figure 2. The mean seasonal cycle of (a) rainfall occurrence (days/pentad), and (b) wet-day amount ( $\text{mm day}^{-1}$ ) at each station, computed using pentad block means.

is largest near the east coast, along the coastal escarpment, while rainfall is infrequent inland. Geographical contrasts in wet-day amount are smaller than for occurrence, with the largest intensities at station 9.

The mean seasonal variation in occurrence and wet-day amount is depicted in Fig. 2, in terms of 40-year pentad averages. Both measures increase strongly through the October–November transition season. This seasonal increase continues into the December–February core monsoon season toward the coast, tending to level off at the inland stations. The wet season is largely encompassed by the 1 October–15 April

dataset, although some stations continue to experience significant rainfall well into the austral autumn.

### 3. THE HIDDEN MARKOV MODEL (HMM)

The model for rainfall occurrence largely follows that described in Robertson *et al.* (2004a), with further details given in Robertson *et al.* (2003), and is based on the work of Hughes and Guttorp (1994). The main departure from Robertson *et al.* (2003, 2004a) is that we model rainfall amount by incorporating a mixture model for amount into the HMM. This consists of a delta function to model dry days, and a mixture of exponentials to describe rainfall amount on wet days (i.e. rainfall intensity). Fitting the mixture and exponential parameters is accomplished as an integral part of the HMM, via the expectation-maximization (EM) algorithm (Dempster *et al.* 1977).

Let  $\mathbf{R}_t = (R_t^1, \dots, R_t^M)$  be a vector of rainfall amounts for a network of  $M$  stations on day  $t$ , and let  $\mathbf{R}_{1:T}$  denote a time sequence of such vectors  $\mathbf{R}_1, \dots, \mathbf{R}_T$ . The sequence of observed rainfall measurements  $\mathbf{R}_{1:T}$  is assumed to be generated by a Markov chain of hidden (unobserved) weather states  $\mathbf{S}_{1:T} = (S_1, \dots, S_T)$ , where  $S_t$  takes values from 1 to  $K$ . Making a first-order Markov assumption, we factor the joint distribution over the hidden state sequence as:

$$p(S_1, \dots, S_T) = p(S_1) \prod_{t=2}^T p(S_t|S_{t-1}) \quad (1)$$

where  $p(S_t|S_{t-1})$  is modelled as a  $K \times K$  stochastic matrix of state transition probabilities  $\mathbf{\Gamma} = \{\gamma_{ij}\}$ ,  $1 \leq i, j \leq K$ , and  $p(S_1)$  is a set of initial state probabilities.

The second major assumption in the HMM is that the instantaneous rainfall  $\mathbf{R}_t$  for a particular day  $t$  is assumed to depend only on the state on day  $t$ , so that all rainfall autocorrelation is treated at the state level. Thus,

$$p(\mathbf{R}_{1:T}|\mathbf{S}_{1:T}) = \prod_{t=1}^T p(\mathbf{R}_t|S_t) \quad (2)$$

and for the joint distribution we have

$$p(\mathbf{R}_{1:T}, \mathbf{S}_{1:T}) = \left\{ p(S_1) \prod_{t=2}^T p(S_t|S_{t-1}) \right\} \left\{ \prod_{t=1}^T p(\mathbf{R}_t|S_t) \right\}. \quad (3)$$

We further assume that the  $M$  station components of the vector of rainfall amounts at time  $t$  are conditionally independent of each other given the hidden state  $S_t$ ; spatial dependence is captured implicitly via the state variable,

$$p(\mathbf{R}_t|S_t) = \prod_{m=1}^M p(R_t^m|S_t). \quad (4)$$

The probability models for individual stations,  $p(R_t^m|S_t)$ , are modelled as finite mixtures with  $C + 1$  components, consisting of (a) a delta function modelling zero precipitation, and (b) a mixture of  $C$  exponential components to model the non-zero amount distribution, i.e.

$$p(R_t^m = r|S_t = i) = \begin{cases} p_{im0} & r = 0, \\ \sum_{c=1}^C p_{imc} \lambda_{imc} e^{-\lambda_{imc} r} & r > 0, \end{cases} \quad (5)$$

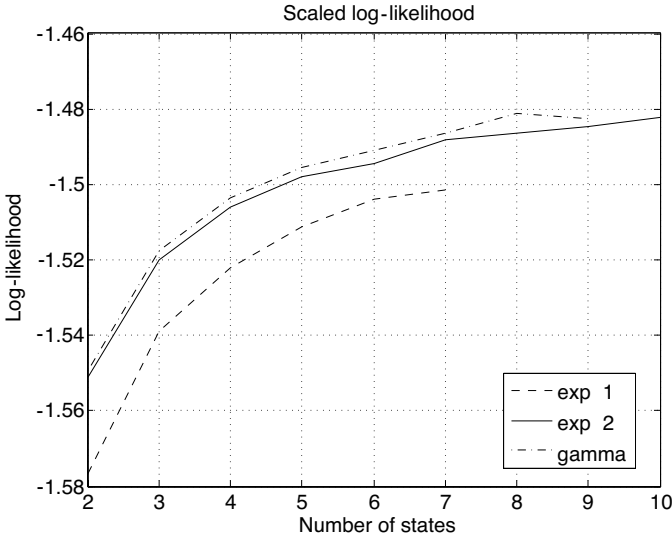


Figure 3. Out-of-sample log-likelihood for amounts models using: single exponential (dashed), mixture of two exponentials (solid), and gamma (dot-dashed) distributions for rainfall on wet days.

with  $p_{imc} \geq 0$  and  $\sum_{c=0}^C p_{imc} = 1$  for all  $m = 1, \dots, M$  and  $i = 1, \dots, K$ . Previous studies have demonstrated that a mixture of two exponentials represents well the distribution of daily rainfall amounts (e.g. Wilks and Wilby 1999).

The parameters of the model are estimated from the observed rainfall amount data in a standard manner using the EM algorithm. Note that rainfall amounts are thus incorporated directly into the formulation of the HMM, similar to the approach of Bellone *et al.* (2000). This contrasts with Charles *et al.* (1999), where amounts were modelled a posteriori, in an HMM of rainfall occurrence. Details of the EM estimation algorithm were presented by Robertson *et al.* (2003) for a model which is similar except that binary precipitation occurrence data is modelled instead of amount data. The mixture of exponentials fits well into the maximum-likelihood framework, and the additional EM equations required to handle estimation of the parameters for the state-dependent amount models are described in Kirshner (2005).

#### 4. STATES OF DAILY RAINFALL AMOUNTS

##### (a) Number of states

As in Robertson *et al.* (2004a), cross-validation is used to evaluate the quality of the fitted HMMs in terms of log-likelihood

$$\ln p(\mathbf{R}_{1:T}) = \ln \sum_{S_{1:T}} p(\mathbf{R}_{1:T}, S_{1:T}) \quad (6)$$

as a function of  $K$ , the number of states. Here, five-year blocks of data were withheld, the model trained on the remaining 35 years, and the simulations compared with observed rainfall for the eight 5-year validation periods. In each case the EM algorithm was run 10 times from different initial seeds, selecting the run with the highest log-likelihood. The resulting normalized out-of-sample values of the log-likelihood for each model are plotted for  $K = 2$  to 10 in Fig. 3. Three different amounts models are plotted. Each consists of a delta function at zero amount, together with (a) a single exponential

(dashed), (b) a mixture of two exponentials (solid) and (c) a gamma function (cf. Bellone *et al.* 2000) (dot-dash). The curves in Fig. 3 do not reach  $K = 10$  for the single exponential and gamma models, because the number of parameters to be estimated increases with  $K$ , and none of the EM runs from the 10 initial seeds converges for large  $K$  in these cases. This is a technical issue, due to the poorer fit to the data of these models (especially the single exponential). The double exponential is the most well behaved model, and we focus on it in the following. Its performance is similar to the gamma model (both are two-parameter models) and it is superior to the single exponential model.

The out-of-sample log-likelihood of the models increases monotonically with  $K$  in an asymptotic manner. Thus, the model does not ‘overfit’ for larger  $K$ , suggesting that the rainfall data is considerably more complex than our models. For parsimony, we choose  $K = 5$ . The results discussed below are not sensitive to choosing  $K = 4 - 6$ . Choosing larger  $K$  makes the states more difficult to interpret physically. Thus,  $K = 5$  represents a compromise between model performance and physical interpretability.

### (b) Estimation of the model parameters

Having chosen the 5-state model, its parameters were estimated from the entire 7880-day rainfall record. The EM algorithm was restarted 10 times, selecting the run with the highest log-likelihood. The resulting rainfall parameters are illustrated in Fig. 4, in terms of the probability of rain (panels a–e), and the mean rainfall amount on wet days (panels f–j). The latter was computed from the parameters of the mixed exponential distribution.

State 5, the wettest state, exhibits similar spatial distributions of occurrence and amount to those of the observed summertime climatology (Fig. 1), with largest values along the east coast. State 1 is also characterized by rainfall at all stations, but probabilities and mean amounts are lower than for state 5, especially in the north-east. State 2 has very low rainfall probabilities everywhere. Both states 3 and 4 are characterized by high rainfall probabilities along the coast, with dry conditions inland. However, coastal intensities are much lower than for state 5, and state 3 may be characterized by frequent drizzle along the coast. For states 1, 4 and 5, rainfall probability tends to be correlated with rainfall intensity, but this is less the case for states 2 and 3.

The state-transition matrix is given in Table 1. The Markov property of the HMM is clear in the relatively large self-transition probabilities, i.e. persistence. The dry and wet states, 2 and 5 respectively, are most persistent, with the ‘coastal’ states 3 and 4 less so. Other transition probabilities are generally small, though there are hints of preferred transitions from state 4 to 3, and from state 3 to 2.

### (c) Simulations

Once the parameters of the model have been estimated, multiple simulations of daily rainfall amounts can be generated. Figure 5 shows the marginal daily distributions of simulated rainfall amount for a dry and a wet station (stations 4 and 7 respectively), versus the observed. These two stations are typical of the model’s performance. Histograms are plotted with a bin-width of  $5 \text{ mm day}^{-1}$ , with the  $0 \text{ mm day}^{-1}$  delta function given by the leftmost pair of bars. The simulated values were generated from 100 simulations of 40 years.

The mixture of a delta function at zero rainfall together with a mixture of two exponentials is seen to fit the observed data closely, especially away from the tail of the distribution. The simulations (black bars) exhibit a smoother distribution at high

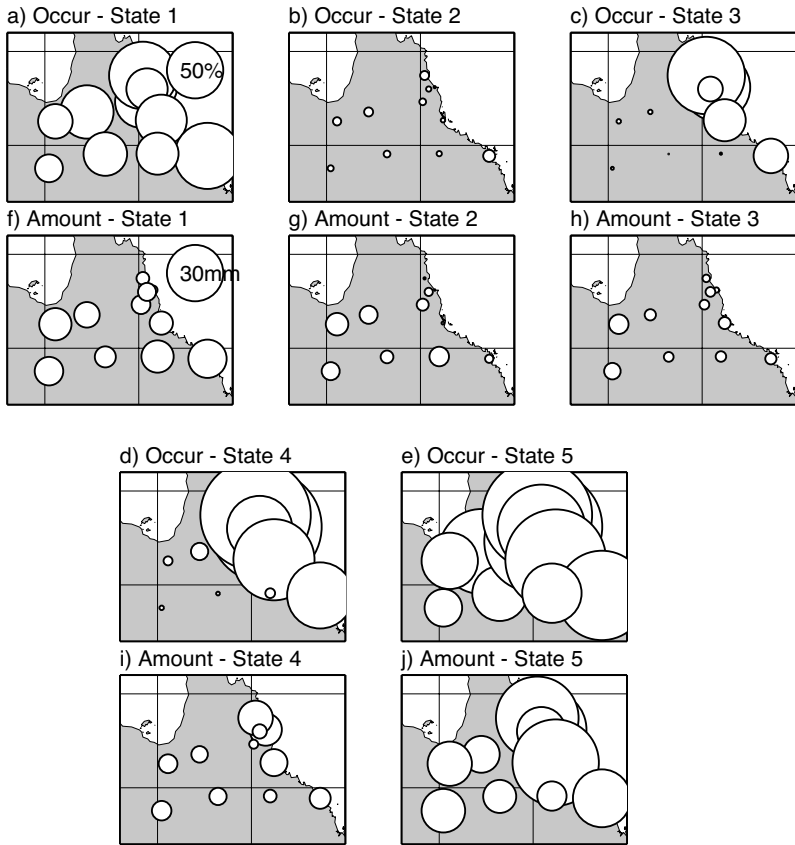


Figure 4. Five-state HMM rainfall parameters. (a)–(e): occurrence probabilities (circle radius). (f)–(j): mean intensities (wet-day amounts) of the mixed exponential.

TABLE 1. TRANSITION PROBABILITIES

		To state				
		1	2	3	4	5
From state	1	0.654	0.047	0.060	0.100	0.140
	2	0.080	0.784	0.114	0.022	0.000
	3	0.045	0.256	0.552	0.144	0.002
	4	0.051	0.043	0.271	0.559	0.076
	5	0.143	0.008	0.018	0.122	0.709

amounts because sampling variability is small for the 100-member simulation run. The contrast between the dry and wet station is clear, and the former has many more dry days (note the log scale on the ordinate). The conditional rainfall distributions for each individual state were found to be similarly accurate (not shown).

#### (d) *The estimated state sequence*

The most-probable daily sequence of the five states over the 40-summer period can be estimated using a dynamic programming algorithm, known as the Viterbi algorithm (Forney 1978). This assignment allows an interpretation of the observed rainfall record



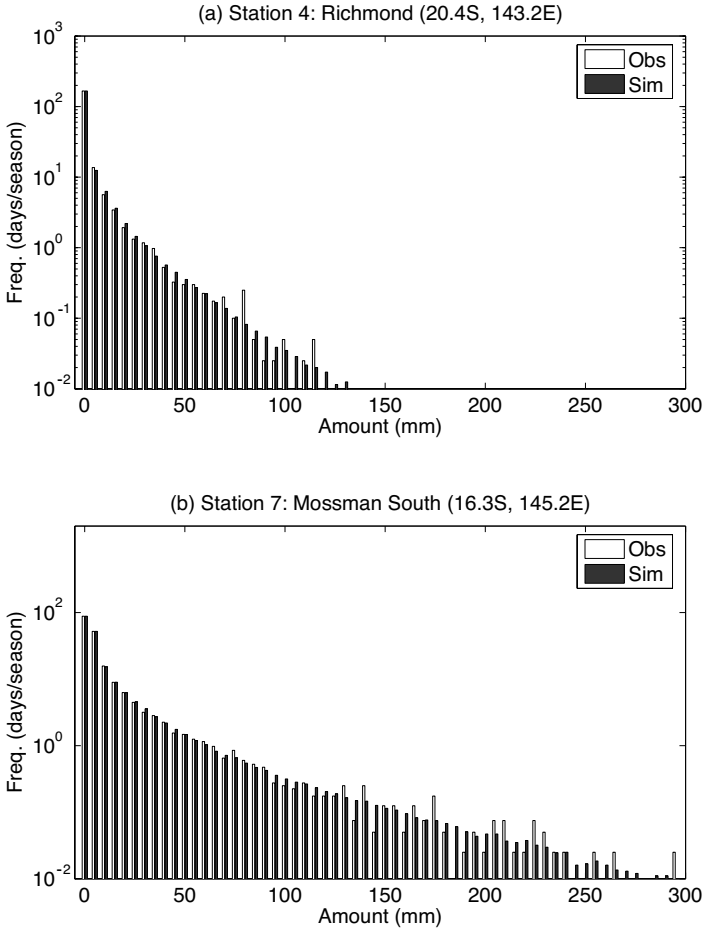


Figure 5. Histograms of the simulated (black) and observed (white) daily rainfall amount distributions ( $\text{mm day}^{-1}$ ). The leftmost pair gives the zero-rainfall frequencies. The bin-width is  $5 \text{ mm day}^{-1}$ .

in terms of these states, and thereby, the atmospheric circulation patterns that accompany each state. The sequence, plotted in Fig. 6, exhibits a clear seasonality as well as marked interannual and subseasonal variations. The average seasonal cycle is plotted in Fig. 7, in terms of pentad means of state occurrence averaged across all 40 years. The dry state (no. 2, 34% of days) dominates during October–December, with the wettest state (no. 5, 13% of days) becoming most prevalent in January–February, waning during March. States 3 and 4 (wet along the coast; 20% and 15% of days, respectively) become more prevalent in March–April, while state 1 (moderately wet, 18% of days) has a seasonality similar to state 5, but with higher prevalence in the austral spring.

Figure 7 suggests a description of the average monsoon evolution in terms of the rainfall states identified by the HMM, while the individual daily sequences of states in Fig. 6 point to the large degree of within-season and year-to-year variability. To pursue this further, we next examine the nature of the accompanying atmospheric circulation patterns.

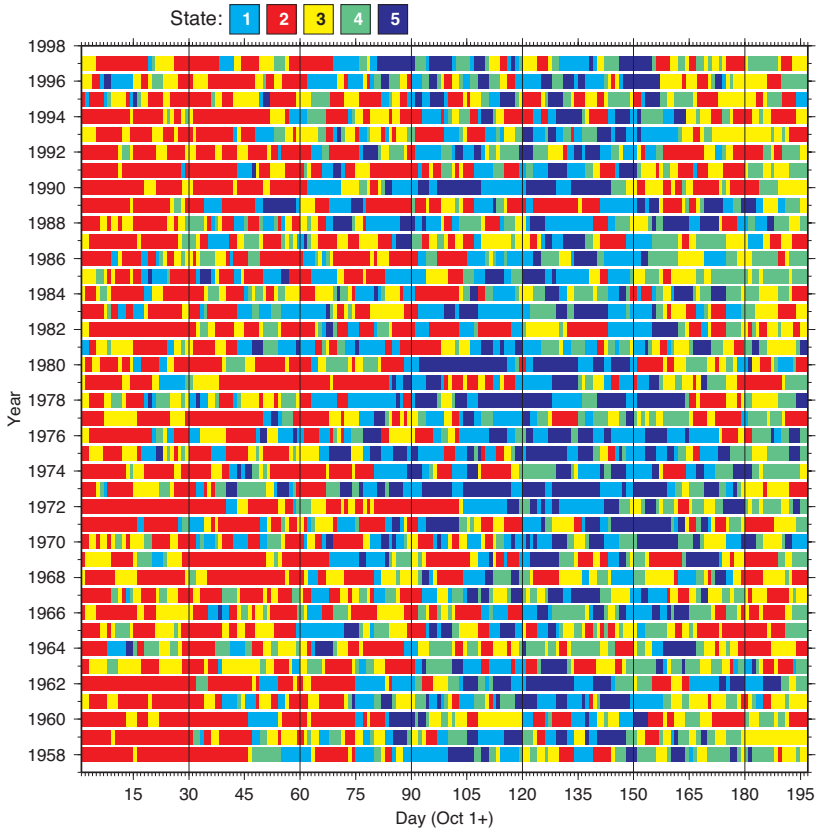


Figure 6. The estimated state sequence. The number of days falling into the five states are 1413, 2685, 1571, 1211, 1000 respectively.

### (e) *Synoptic conditions*

Relationships with atmospheric circulation are explored using the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data (Kalnay *et al.* 1996), together with the European Centre for Medium-Range Weather Forecasts reanalysis (ECMWF ERA-40) for the potential vorticity field. Composites of atmospheric circulation variables from reanalysis data are plotted for each state in Figs. 8 and 9, computed by averaging over the days assigned to each state by the Viterbi algorithm.

Figure 8 shows composites of 850 hPa winds and 500 hPa omega composites, constructed from unfiltered daily data. Large-scale ascent dominates in the equatorial region, extending into northern Australia in states 1, 4 and especially state 5; subtropical descent dominates over Australia in states 2 and 3. Monsoon westerlies and subtropical easterlies occur to varying degrees, consistent with the vertical motion: the westerlies are strongest in state 5 (wettest), weakening through states 1, 3, 4, to state 2 (dry) in which they are almost completely replaced by the south-east trade winds. States 3 and 4 (wet near the coast) both show strong south-easterlies, consistent with rain along the east coast. At upper levels the winds reverse (not shown), with subtropical westerlies and equatorial easterlies sandwiching the monsoonal anticyclone in state 5.

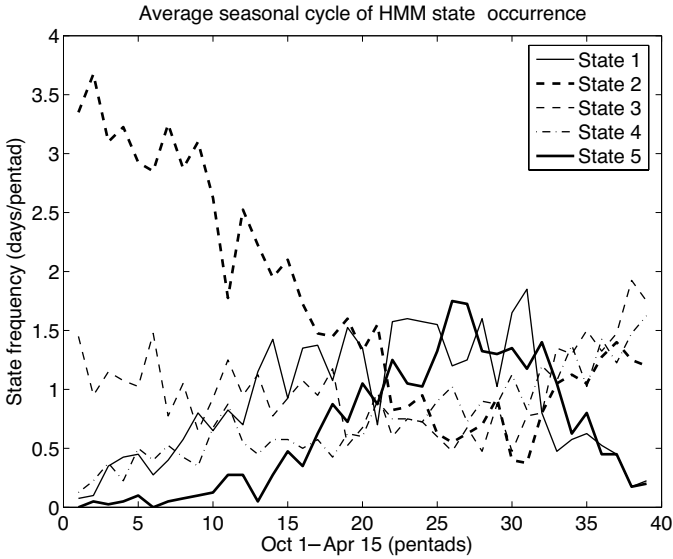


Figure 7. The mean seasonal cycle of HMM state occurrence, computed from pentad means (days/pentad).

In terms of the mean seasonal evolution (cf. Fig. 7), the composites for state 2 resemble the mean climatology for July (i.e. winter), and state 5 resembles February (i.e. the peak of the monsoon). States 3 and 4 resemble the transition season, as the south-east trades extend northward up the coast. The seasonal extension of the subtropical convergence zone into south-eastern Australia is also reflected in states 1 and 5.

Composite anomalies relative to the mean seasonal cycle are plotted in Fig. 9. Only statistically significant wind anomaly vectors are plotted, with shading denoting significant vertical motion anomalies. All the rainfall states are associated with statistically significant atmospheric circulation anomalies, with states 2, 4 and 5 being the most nominally significant. The strongly monsoonal state-5 composite of anomalous vertical motion over Queensland stands out as being comparable in magnitude to that of the total field (Fig. 8(e)), while the anomalies associated with other states are substantially weaker. State 5 is associated with large-scale anomalous ascent over Queensland, and a low-level anomalous cyclonic circulation displaced slightly to the west. This eddy is accompanied by low-level north-westerly wind anomalies over Queensland, and implies an influx of tropical moisture. The implied advection of anomalously high moist static energy during state 5 will tend to maintain the monsoon’s thermally-direct circulation, while the anomalous southward low-level wind component will act to offset vortex stretching at low levels, typical features of a monsoonal circulation. There is also a region of (weakly significant) anomalous descent over Western Australia during states 1 and 5, suggesting a Rossby wave-train response to the tropical heating over Queensland. Compared to state 5, the cyclonic eddy in state 1 is displaced further south-east, and is less intense.

The right-hand panels in Fig. 9 show circulation anomalies at 200 hPa, in terms of NCEP–NCAR reanalysis winds, and ERA-40 potential vorticity (PV). The two reanalysis datasets are in good correspondence with each other, with qualitative geostrophic balance between the wind and PV anomalies. The 200 hPa zonal wind anomalies of state 5 are easterly over Indonesia, consistent with a baroclinic vertical structure at low latitudes. A region of anomalous anticyclonic PV overlies the Queensland coast,

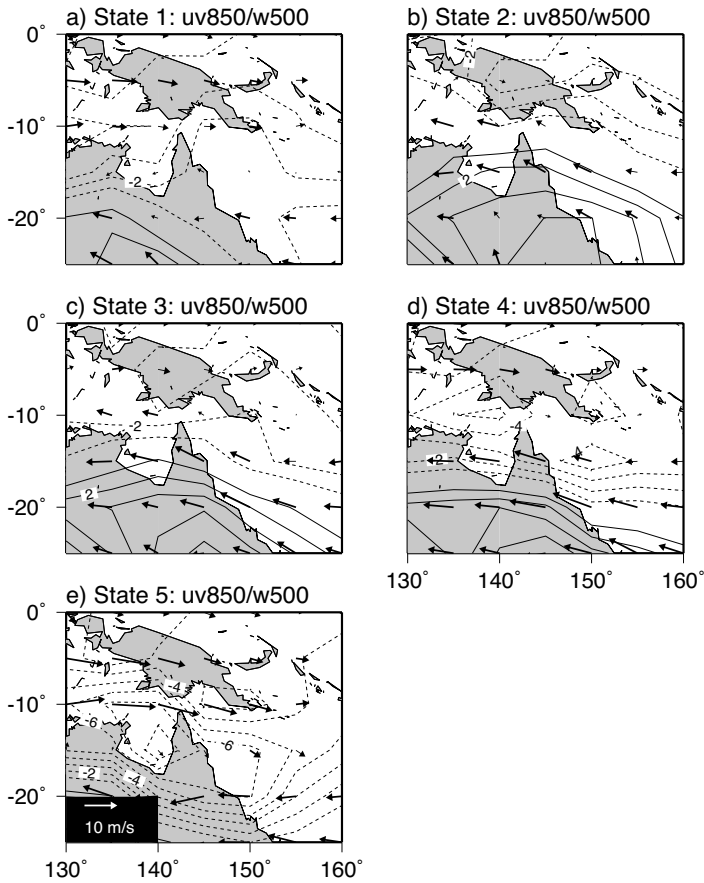


Figure 8. Composites of 850 hPa winds (vectors:  $\text{m s}^{-1}$ ) and 500 hPa omega vertical velocity (contour interval  $0.01 \text{ Pa s}^{-1}$ ) over the days assigned to each state, computed from NCEP–NCAR reanalysis unfiltered daily data. As in Fig. 6, the number of days falling into the five states are 1413, 2685, 1571, 1211, 1000 respectively.

with anomalous northward flow on its eastern flank. This implies anomalous northward advection of large cyclonic mean PV in state 5, that would tend to offset upper-level vortex compression over the monsoon region associated with ascent, while at the same time tending to destabilize the atmospheric column and thereby amplify the monsoonal circulation.

State-1 anomalies are generally weaker with less statistical significance, especially at upper levels, compared to state 5. The most marked feature in states 2 and 4 is the circulation anomaly to the east of Tasmania, whose influence extends into North Queensland. A cyclonic disturbance (state 2) leads to dry conditions, while an anticyclonic one (state 4) leads to rainfall especially along the coast.

#### (f) *Monsoon onset dates and subseasonal characteristics*

According to Fig. 7, the evolution of the monsoon over North Queensland consists, on average, of a seasonal progression from the dry state toward increasing prevalence of first the weakly-monsoonal state (no. 1), followed by the strongly-monsoonal state (no. 5). State 4 with south-easterlies and rainfall along the coast increases in prevalence toward the end of the monsoon season.

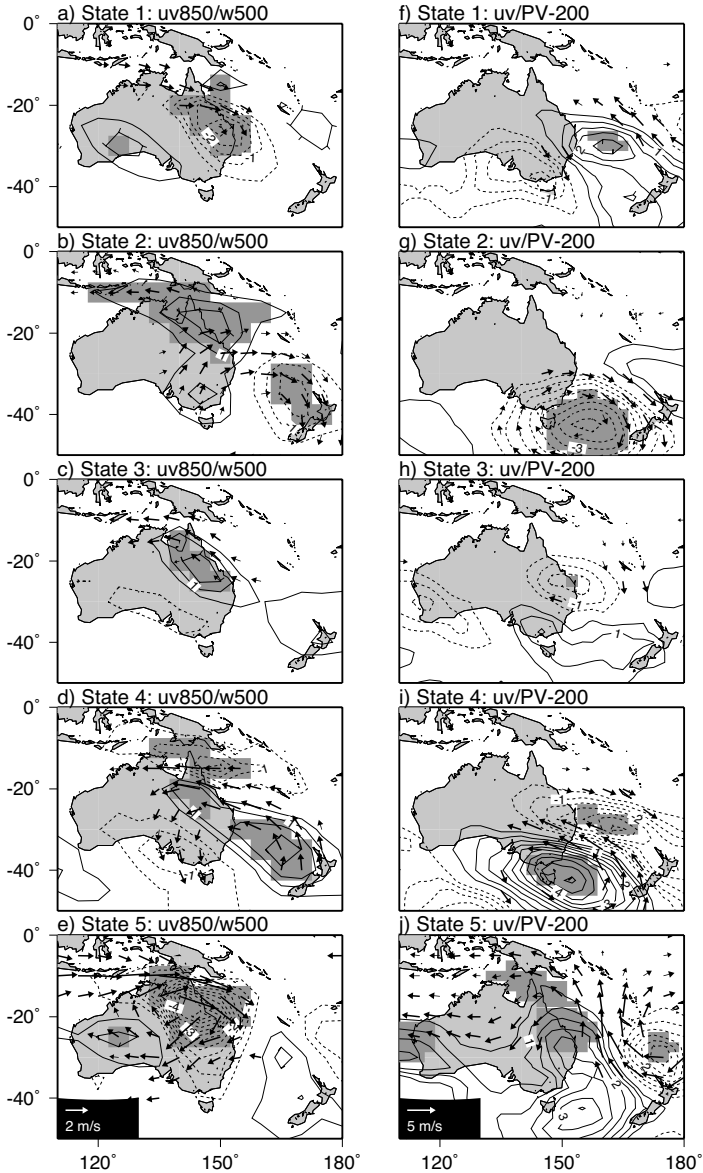


Figure 9. Anomaly composites with respect to the mean seasonal cycle, over the days assigned to each state. Left panels: 850 hPa winds (vectors:  $\text{m s}^{-1}$ ) and 500 hPa omega vertical velocity (contour interval  $0.005 \text{ Pa s}^{-1}$ ). Right panels: 200 hPa winds (vectors:  $\text{m s}^{-1}$ ) and 200 hPa potential vorticity from ERA-40 (contour interval  $0.5 \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$ ). Only wind vectors statistically significant at the 95% level are plotted, while shading indicates 95%-significant omega and PV anomalies. The number of days in each composite was divided by 10, as a conservative estimate of the number of effective degrees of freedom. The mean seasonal cycle was computed after firstly 10-day low-pass filtering the data.

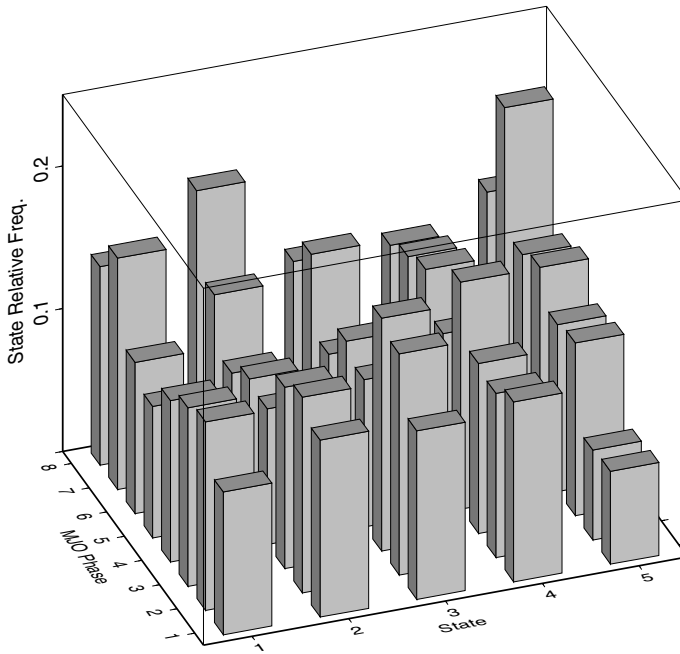


Figure 10. Histogram showing the occurrence frequency of each state (December–March, 1974–98), stratified by the phase of the MJO, as defined by Wheeler and Hendon (2004). Phase 7 is wet over North Queensland.

The evolution of the monsoon in any particular year can be interpreted in terms of the estimated state sequence. The intermittent nature of the monsoon is characterized by the HMM in terms of transitions between discrete states. Rainfall shows less subseasonal variability along the coast where states 1, 4 and 5 all bring substantial rainfall, while its intermittency is large inland. State 2 is clearly associated with a ‘break’ in the monsoon at all stations, while states 3 and 4 only signal a break over inland stations. The occurrence of persistent low-level westerlies near Darwin has been associated with monsoon onset (Troup 1961), so that the first persistent spell of states 1 and 5 could be identified with onset. However, the intermittent nature of Fig. 6 suggests that the definition of an onset date may not be meaningful in all years, at least away from the coast.

To determine whether the HMM’s estimated state sequence is influenced by the Madden–Julian oscillation, a two-dimensional histogram is plotted in Fig. 10. The histogram stratifies HMM state occurrence during the December–March seasons (1974–98) according to the eight phases of the MJO defined by Wheeler and Hendon (2004) from empirical orthogonal functions (EOFs). Their all-season real-time multivariate MJO Index was obtained from the [www.bom.gov.au](http://www.bom.gov.au) website. The very wet state (state 5) is most frequent during phase 7, which corresponds to the wet phase of the MJO over North Queensland (see Wheeler and Hendon 2004, their Fig. 16, and [www.bom.gov.au](http://www.bom.gov.au)). The dry state (state 2) is most frequent during phase 8, which is a neutral phase of the MJO as regards North Queensland rainfall. Thus, the occurrences of both the very wet and dry states are consistent with the MJO, with the stronger relationship during the active phase of the monsoon.

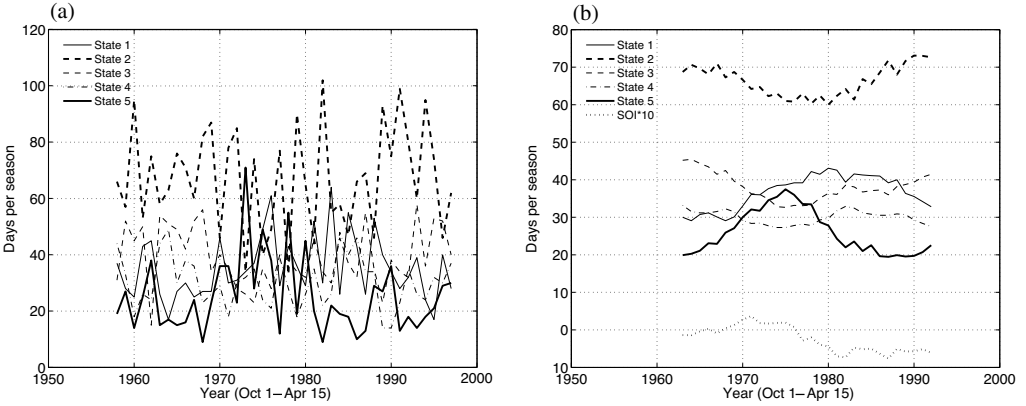


Figure 11. (a) Interannual variability and (b) 11-year running mean of state frequency. The 11-year running mean of the SOI index (multiplied by 10) is also shown.

(g) *Interannual and interdecadal variability*

The interannual variability in state occurrence frequency is plotted in Fig. 11(a), computed by summing the number of days per season in the estimated state sequence. The interannual prevalence of the dry state 2 tends to vary inversely to the monsoonal states 1 and 5. Correlations between these series and the October–April SOI index are (0.33,  $-0.62$ ,  $-0.01$ , 0.01, 0.58), indicating that the monsoonal states 1 and 5 both tend to be associated with La Niña, with the strongest correlation for state 5, while occurrence of the dry state 2 is strongly correlated with El Niño. These ENSO relationships are further brought out in Fig. 12, which shows the relationship between state frequency and SSTs for states 1, 2 and 5. Shown are SST anomaly composites for the October–December (OND) and January–March (JFM) seasons in which state frequency exceeds one standard deviation from the mean, using the NOAA Extended Reconstructed SST dataset (Smith *et al.* 1996). The composites are defined to be the difference between seasons in which frequency is anomalously high, minus seasons with anomalously low frequency. Consistent with the SOI correlations, all three states are related to ENSO SST anomalies in at least one season. The most significant La Niña relationships with states 1 and 5 occur during OND, while the dry state 2 is strongly associated with El Niño during both seasons. During OND, there are significant relationships with SST over the South Pacific east of Australia. During JFM, there are significant associations with Indian Ocean SSTs, consistent with ENSO-related SST anomalies there.

In addition to the large interannual variability, the frequency of state occurrence also varies on interdecadal time-scales as shown in Fig. 11(b). The strong-monsoonal regime was most prevalent during the 1970s, and became less frequent after about 1981. Concomitant with this decrease, the weak monsoonal state became more prevalent, indicative of a weakening of the monsoon between the mid-1970s and 1998. This drying trend is even more pronounced since about 1985, with an increase in the prevalence of the dry state. The 1960s were also marked by less-frequent monsoonal states, and higher prevalence of the dry states 2 and 3. The tendency toward drier conditions since the late 1970s is consistent with more-negative values of the SOI, plotted at the bottom of Fig. 11(b).

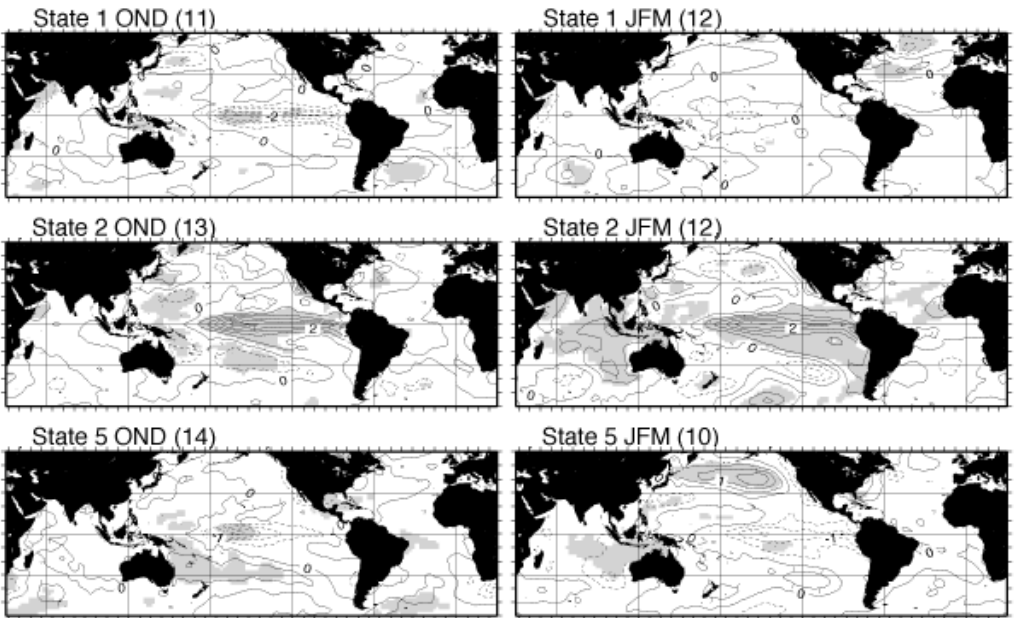


Figure 12. Composites of seasonal mean SST anomalies, according to occurrence frequency of states 1, 2 and 5, for OND (left) and JFM (right) seasons. Shown are SST differences between summers when the HMM states are most prevalent, minus those in which they are least prevalent as defined by  $\pm$  one standard-deviation excursions of the curves in Fig. 11(a). The number of seasons in each composite is given in brackets. Shading represents 95% statistical significance according to a two-sided Student  $t$ -test with 10 degrees of freedom. Negative contours are dashed, and the contour interval is 0.2 degC.

## 5. SEASONAL SIMULATION AND PREDICTION WITH A NON-HOMOGENEOUS HMM

In this section we introduce input variables into the HMM, by creating an explicitly non-homogeneous HMM (NHMM). In a non-homogeneous HMM the state-transition matrix  $\Gamma$  is no longer stationary, and the transition probabilities are defined to be a function of a (possibly) multivariate ‘predictor’ input time series  $\mathbf{X}_{1:T}$ , corresponding (for example) to other variables that can influence the evolution of the weather state sequence  $\mathbf{S}_{1:T}$ . Here we describe experiments with such a model, where the transition probabilities are defined as a logistic function  $f$  of the predictor variables, i.e.

$$\gamma_{ij}(t) = \gamma_{ij} \times f(\mathbf{X}_t).$$

More complete details on this model are provided in Hughes *et al.* (1999) and Robertson *et al.* (2003).

The ‘predictors’ are defined from large-scale precipitation fields derived from the NCEP–NCAR reanalysis dataset, and from simulations of the ECHAM4.5 atmospheric general circulation model (GCM) (Roeckner *et al.* 1996). We also show the result of using statistical predictors based on antecedent June–August Pacific SSTs.

The reanalysis precipitation fields can be interpreted as a single GCM simulation, which has been ‘nudged’ toward available (non-precipitation) observations. Precipitation is known to be one of the most problematic variables to model in a GCM, and its usage here can be regarded as a test of the ability of the reanalysis model to synthesize the large-scale precipitation process, and provide an appropriate large-scale input to the NHMM.



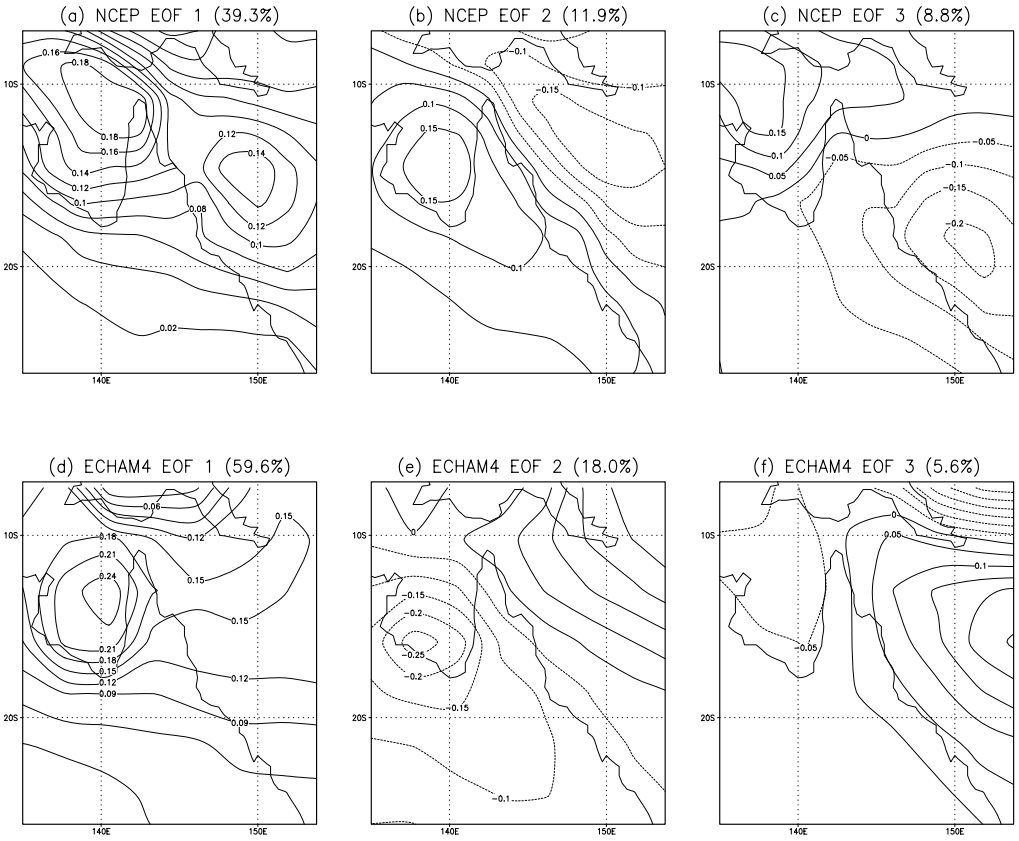


Figure 13. Leading three regional EOFs of precipitation for (a–c) NCEP/NCAR reanalysis (135°E–154°E, 28°S–9°S), and (d–f) ECHAM4 precipitation (135°E–157°E, 29°S–7°S). Units are relative, and negative contours are dashed. The variance explained is given in brackets. The EOFs were constructed from 10-day low-pass filtered daily fields 1 October–15 April.

Twenty-four long-term simulations with the ECHAM4.5 GCM were available, with historical SSTs prescribed (each differing only in its initial condition) from which we derive the ensemble mean. These inputs to the NHMM represent the portion of large-scale precipitation that can be characterized by a GCM response to SST anomalies. As such, we cannot expect any correspondence on a particular day between the GCM simulations and observations, but the simulations of ensemble-mean three-month averages generally contain some skill in the tropics, depending on the geographical region (Goddard *et al.* 2003; Robertson *et al.* 2004b).

For both the reanalysis and GCM, inputs are defined from a principal components analysis (PCA) of gridded daily precipitation over the regional domains given in Fig. 13, using 10-day low-pass (Blackmon and Lau 1980) filtered daily precipitation. The loading patterns (empirical orthogonal functions, EOFs) are plotted in Fig. 13. The PCA is used here as a data compression tool, so as to define the predictors in a semi-objective fashion from the large-scale precipitation field. The leading three principal components (PCs) were selected in each case; their variances are given in Fig. 13, accounting together for 60.0% (reanalysis) and 83.2% (ECHAM4.5) of the low-pass filtered variance respectively. We have used 10-day low-pass filtered precipitation values

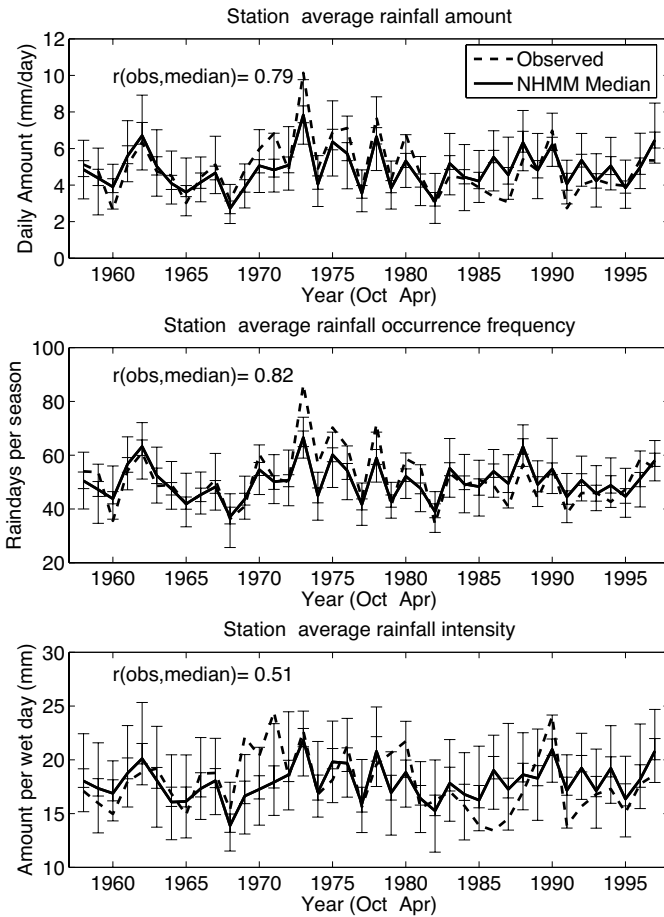


Figure 14. Interannual variability of NHMM-simulated seasonally-averaged rainfall amount (top), occurrence frequency (middle), and intensity (average amount on wet days, bottom). Plotted is the median of 100 NHMM simulations averaged over all 11 stations (solid), versus the observed (dashed). Inputs are from NCEP–NCAR reanalysis precipitation PCs. The error bars indicate the entire range of the 100 simulations, with the interquartile range given by the inner ticks.

so as to capture the seasonal cycle of precipitation, and to include a limited amount of subseasonal variability while averaging over day-to-day weather.

The NHMM was then trained in cross-validated mode, as described in section 4(a), using the observed station rainfall together with the three daily PC inputs, to generate 100 stochastic daily simulations of rainfall. The NHMM simulations made with reanalysis inputs are shown in Fig. 14, in terms of seasonal means, averaged over the 11 stations. The median of the simulations is plotted versus the observed, together with the interquartile and full range of the 100-member simulation distribution of seasonal and station averages. Interannual variability of rainfall occurrence frequency and seasonal amount are well simulated ( $r = 0.82$  and  $0.79$  respectively), with seasonal-mean intensity less so ( $r = 0.51$ ). The correlations between the median of seasonal-averaged simulated rainfall amount, with the observed, are tabulated in Table 2 for each individual station in turn. All stations, except one (Cloncurry, inland,  $r = 0.34$ ), achieve correlations of about 0.6. Also shown in Table 2 are correlations of observed station seasonal

TABLE 2. INTERANNUAL CORRELATIONS

Station	NHMM amount	Reanalysis-grid amount	Dry-spell frequency
1	0.68	0.14	0.33
2	0.63	-0.14	0.29
3	0.34	-0.04	0.13
4	0.69	0.22	0.40
5	0.65	0.30	0.50
6	0.61	0.31	0.14
7	0.68	0.23	0.53
8	0.66	0.41	0.29
9	0.59	0.27	0.52
10	0.56	0.21	0.47
11	0.65	0.24	0.50
Average	0.79	0.29	0.63

Correlations between seasonal averages of simulated rainfall (NHMM or reanalysis grid box) and the observed values at each station. Rainfall amount and 10-day dry-spell frequency.

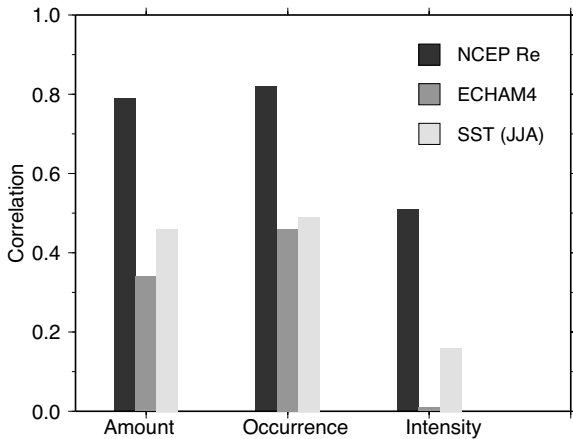


Figure 15. Interannual correlations of simulated versus observed station-averaged October–April rainfall, using predictors derived from (a) NCEP/NCAR reanalysis precipitation, (b) ECHAM4 precipitation, and (c) June–August SST. The correlation coefficient between simulated and observed seasonal averages is plotted on the ordinate. See text for details.

amount with reanalysis precipitation taken from the grid box overlying each respective station. These correlations are much lower than those derived from the NHMM. Ten-day dry-spell frequency is a quantity of particular relevance to agriculture. The NHMM simulations perform less well and more unevenly for this statistic (column 4 in Table 2), although the correlation of the station average reaches  $r = 0.63$ . An attempt to estimate dry-spell frequency directly from the reanalysis grid-point data gave very poor results.

Figure 15 compares the performance of NHMM simulations made with three different choices of predictors: reanalysis precipitation (black), GCM precipitation (dark grey), and antecedent Pacific SSTs (light grey). The NCEP/NCAR reanalysis and ECHAM4 GCM ‘predictors’ are the October–April contemporaneous precipitation PCs described above, whose EOFs are plotted in Fig. 13. The Pacific SST predictors were defined by averaging antecedent June–August SST over two boxes—one over the central equatorial Pacific (167°W–133°W, 5°S–5°N), and the other over the south-west Pacific, adjacent to Australia (147°E–167°E, 31°S–17°S). Thus we are using June–August SST

anomalies to predict the subsequent October–April rainfall over North Queensland. The two SST regions are characterized by correlations reaching 0.5 between June–August SST and October–April rainfall occurrence summed over the 11 North Queensland stations.

Again, the veracity of the simulations is measured in terms of the correlation coefficient between simulated and observed station-seasonal averages, using the median of the 100-member simulated distribution. Not surprisingly, both the ECHAM4-NHMM and SST-NHMM models exhibit lower station-averaged correlations than the reanalysis-NHMM. However, the purely statistical SST-NHMM model, based on SSTs from the preceding winter season, is just as skilful as the ECHAM4-NHMM, even though the GCM simulations are made with prescribed observed SSTs, and are thus not true forecasts. Direct use of the GCM's grid-box precipitation values (analogous to column 3 of Table 2) leads to low correlations of seasonal amounts at the station level, with a range of  $-0.09$  to  $+0.42$  over the 11 stations.

## 6. DISCUSSION

The HMM yields a state-based description of daily rainfall measured on a network of stations, which we have used to interpret the variability of the monsoon over North Queensland on time-scales of subseasonal to interdecadal. The picture that emerges is one of substantial intrinsic rainfall variability associated with transitions between rainfall states, rather than of a smooth seasonal evolution. The concept of a monsoon onset date is scarcely supported, but rather one in which certain regimes have a probability of occurring that is both seasonally dependent, as well as strongly modulated on interannual and longer time-scales. Even the dry state occurs quite frequently during February in some years, and could be interpreted as a monsoon 'break'. This interpretation is consistent with previous studies on the intermittent nature of the Australian monsoon, starting with Troup (1961), and discussed recently by Wheeler and McBride (2005). It is analogous to recent descriptions of the Indian summer monsoon, in terms of modes with predominantly intraseasonal time-scales (Goswami and Mohan 2001; Molteni *et al.* 2003; Gadgil 2003).

The state sequence characterizes temporal variability at the level of the whole rainfall network. Three of the rainfall states describe variability that is relatively homogeneous across the network (states 1, 2 and 5), while states 3 and 4 describe coastal rainfall. Thus, there is only one dry state at the coast, and the monsoon tends to be much more persistent there. Onset would be less variable from year to year, and breaks in the monsoon less frequent. Inland, on the other hand, the monsoon is much more erratic, with 3 dry states and 2 wet ones. Thus, the HMM states describe the nature of subregional variations in the monsoon.

Active and break phases of the monsoon have been associated with northward propagating synoptic waves, as well as with storm surges over the South China Sea (Suppiah and Wu 1998). Our results suggest that episodes of monsoonal convection over North Queensland (state 5) are accompanied by a synoptic-scale ridge centred east of Tasmania, while monsoon breaks (state 2) are associated with a cyclonic disturbance there. There is no evidence in our composites of synoptic waves impinging from the northern hemisphere, even when plotting composites that extend further north. Previous studies have found troughs over the west coast of Australia to precede or modulate the onset of the monsoon (Davidson *et al.* 1983; Keenan and Brody 1988; Hung and Yanai 2004). Our focus on North Queensland, rather than the broader-scale monsoon, may account for this difference. A midlatitude wave to the east is also seen in state 4.

The HMM classification is based on a factorization of the daily rainfall distribution on a network of stations. It is possible that different atmospheric circulation patterns could yield the same pattern of rainfall. This is a potential issue for the dry state where, for example, either the MJO or a midlatitude trough east of Tasmania could each yield dry conditions at all stations, and these effects may be seasonally stratified. Further work is required to investigate this issue.

ENSO is an important modulator of state frequency especially for the dry state 2 (El Niño), and to a lesser extent for the monsoonal states 1 and 5 (La Niña). However, using GCM simulations of large-scale precipitation, or Pacific SST anomalies, as input 'predictors' to the NHMM led to only moderately successful rainfall simulations; the correlation values between observed and simulated seasonal means of rainfall occurrence are substantially lower than those obtained in a similar study over north-eastern Brazil (Robertson *et al.* 2004a). This is consistent with Syktus *et al.* (2003) who found that GCM precipitation simulation skill tends to be low over Queensland during austral summer.

It is interesting to note that the weaker monsoonal state 1 (typically associated with early summer, but less strongly associated with ENSO conditions) has become more prevalent than the stronger monsoonal state since around 1980. Thus, the monsoon has, on average, failed to reach full maturity in recent decades, at least until 1998. This is also consistent with a predominance of more-negative SOI values since the late 1970s (Fig. 11(b)).

We have used NHMMs with various predictors based either on large-scale reanalysis or GCM precipitation, or antecedent Pacific SSTs. These are subjective choices, and predictor selection remains an unresolved issue. In the case of the GCM or reanalysis, other variables may be more appropriate than precipitation. On the other hand, preliminary work with a multi-model ensemble of six GCMs led to improved skill, suggesting that it may be adequate to choose GCM precipitation as the predictor field, provided the simulations of several GCMs are included together. In the case of antecedent Pacific SST predictors, our criteria for model selection used the Queensland rainfall itself, so that the skill reported in Fig. 15 may be overstated. The simulations of seasonally-averaged rainfall amount from the reanalysis-NHMM model are almost as good as those of rainfall occurrence frequency. However, the interannual variability of average rainfall intensity is less well recovered. This may have less to do with the treatment of amounts in the HMM, and more to do with the lower seasonal predictability of rainfall intensity (V. Moron, personal communication).

## 7. CONCLUSIONS

We have examined daily rainfall amount at 11 stations over North Queensland during summer 1958–1998 (Figs. 1, 2), by means of a Hidden Markov Model (HMM). Five discrete rainfall states are used to factor the observed distribution of daily rainfall: three states characterized by overall very wet, moderately wet, and dry conditions respectively, together with two states with enhanced rainfall along the east coast and dry conditions inland (Fig. 4). The HMM rainfall states are found to be associated with statistically significant atmospheric circulation anomalies, supporting the existence of distinct weather states. The two wet states are found to be accompanied by monsoonal circulation patterns, with large-scale ascent, low-level inflow from the north-west, and a phase reversal with height. The dry state is characterized by circulation anomalies of the opposite sense. The coastal rainfall states are characterized by low-level south-easterlies from the ocean, and north-west/south-east midlatitude troughs.

Simulations with the HMM accurately reproduce the marginal distributions of daily rainfall amount at each station (Fig. 5). Variability of the monsoon on time-scales from daily to interdecadal is interpreted through the estimated state sequence (Fig. 6). The seasonal cycle shows a progression from the dry to the monsoonal states (Fig. 7), though with a highly variable monsoon ‘onset’ date. The state sequence exhibits large subseasonal variability, allowing the states to be interpreted as active and break phases of the monsoon. An MJO signal is found in the occurrence of the very-wet monsoonal state (Fig. 10). There are also large year-to-year variations in state frequency that coincide with El Niño–Southern Oscillation conditions (Fig. 12). On decadal time-scales, the state sequence indicates stronger monsoons during the 1970s, and an increased prevalence of the weaker monsoon and dry states since then, consistent with an overall decrease in the SOI index (Fig. 11(b)).

Stochastic simulations of daily rainfall occurrence and amount at the 11 stations are then made by introducing inputs to form a non-homogeneous HMM. These daily input time series are based on large-scale precipitation from (a) NCEP–NCAR reanalysis, (b) the ECHAM4.5 general circulation model run with observed SST forcing and (c) antecedent Pacific SST anomalies. The reanalysis large-scale precipitation yields fairly accurate simulations of the interannual variability of seasonally-averaged rainfall occurrence frequency and amount (Fig. 14, Table 2). This skill is markedly degraded when the GCM predictors are used, and antecedent June–August SST predictors are found to yield a skill comparable to that of the GCM. We expect the NHMM will provide a useful tool in the future, for assessing the potential for seasonal climate forecasts to be downscaled in time.

#### ACKNOWLEDGEMENTS

We wish to thank Josef Syktus and Ken Day for fruitful e-mail exchanges. This work was supported by Department of Energy Grant DEFG02-02ER63413, and by NOAA through a block grant to the International Research Institute for Climate Prediction (IRI). SK and PS were supported by the National Science Foundation under grants No. SCI-0225642 and IIS-0431085. SPC and BCB were funded by the Australian Government’s National Greenhouse Research Program and the Government of Queensland Greenhouse Taskforce for portions of this work. NCEP–NCAR reanalysis data were provided by the NOAA–CIRES Climate Diagnostics Center, Boulder, Colorado, from their Web site available online at <http://www.cdc.noaa.gov>. ECMWF ERA-40 data were provided by ECMWF, through the ECMWF data server. The Queensland Centre for Climate Applications supplied the daily rainfall data.

#### REFERENCES

- |  |      |   |
|--|------|---|
| Bellone, E., Hughes, J. P. and Guttorp, P.     | 2000 | A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. <i>Climate Res.</i> , <b>15</b> , 1–12  |
| Blackmon, M. L. and Lau, N.-C.                 | 1980 | Regional characteristics of the Northern Hemisphere wintertime circulation: A comparison of the simulation of a GFDL general circulation model with observations. <i>J. Atmos. Sci.</i> , <b>37</b> , 497–514 |
| Charles, S. P., Bates, B. C. and Hughes, J. P. | 1999 | A spatiotemporal model for downscaling precipitation occurrence and amounts. <i>J. Geophys. Res.</i> , <b>104</b> , 31657–31669   |
| Chen, T.-C.                                    | 2003 | Maintenance of summer monsoon circulations: A planetary-scale perspective. <i>J. Climate</i> , <b>16</b> , 2022–2037  |
| Cheng, X. and Wallace, J. M.                   | 1993 | Cluster analysis of the Northern Hemisphere wintertime 500 hPa height field: Spatial patterns. <i>J. Atmos. Sci.</i> , <b>50</b> , 2674–2696  |

- Chou, C. and Neelin, J. D. 2003 Mechanisms limiting the northward extent of the northern summer monsoons over North America, Asia, and Africa. *J. Climate*, **16**, 406–425
- Davidson, N. E., McBride, J. L. and McAvaney, B. J. 1983 The onset of the Australian monsoon during winter MONEX: Synoptic aspects. *Mon. Weather Rev.*, **111**, 496–516
- Dempster, A. P., Laird, N. M. and Rubin, D. B. 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc.*, **39B**, 1–38
- Drosowsky, W. 1996 Variability of the Australian summer monsoon at Darwin: 1957–1992. *J. Climate*, **9**, 85–96
- Forney, G. D., Jr. 1978 The Viterbi algorithm. *Proc. IEEE*, **61**, 268–278
- Gadgil, S. 2003 The Indian Monsoon and its variability. *Ann. Rev. Earth. Planet. Sci.*, **31**, 429–467
- Ghil, M. and Childress, S. 1987 *Topics in geophysical fluid dynamics: Atmospheric dynamics, dynamo theory and climate dynamics*. Springer-Verlag, New York
- Ghil, M. and Robertson, A. W. 2002 ‘Waves’ vs. ‘particles’ in the atmosphere’s phase space: A pathway to long-range forecasting? *Proc. Natl. Acad. Sci. of the USA*, **99**(Suppl. 1), 2493–2500
- Goddard, L., Barnston, A. G. and Mason, S. J. 2003 Evaluation of the IRI’s ‘Net Assessment’ seasonal climate forecasts: 1997–2001. *Bull. Am. Meteorol. Soc.*, **84**, 1761–1781
- Goswami, B. N. and Ajaya Mohan, R. S. 2001 Intraseasonal oscillations and interannual variability of the Indian summer monsoon. *J. Climate*, **14**, 1180–1198
- Hendon, H. H. and Liebmann, B. 1990a A composite study of onset of the Australian summer monsoon. *J. Atmos. Sci.*, **47**, 2227–2240
- 1990b The intraseasonal (30–50 day) oscillation of the Australian summer monsoon. *J. Atmos. Sci.*, **47**, 2909–2924
- Holland, G. J. 1986 Interannual variability of the Australian summer monsoon at Darwin: 1952–82. *Mon. Weather Rev.*, **114**, 594–604
- Hughes, J. P. and Guttorp, P. 1994 Incorporating spatial dependence and atmospheric data in a model of precipitation. *J. Appl. Meteorol.*, **33**, 1503–1515
- Hughes, J. P., Guttorp, P. and Charles, S. P. 1999 A non-homogeneous hidden Markov model for precipitation occurrence. *J. Royal Stat. Soc.: Series C (Applied Statistics)*, **48**, 15–30
- Hung, C.-W. and Yanai, M. 2004 Factors contributing to the onset of the Australian summer monsoon. *Q. J. R. Meteorol. Soc.*, **130**, 739–758
- Jeffrey, S. J., Carter, J. O., Moodie, K. B. and Beswick, A. R. 2001 Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software*, **16**, 309–330
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D. 1996 The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.*, **77**, 437–471
- Keenan, T. D. and Brody, L. R. 1988 Synoptic-scale modulation of convection during the Australian summer monsoon. *Mon. Weather Rev.*, **116**, 71–85
- Kimoto, M. and Ghil, M. 1993a Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.*, **50**, 2625–2643
- 1993b Multiple flow regimes in the Northern Hemisphere winter. Part II: Sectorial regimes and preferred transitions. *J. Atmos. Sci.*, **50**, 2645–2673
- Kirshner, S. 2005 ‘Modeling of multivariate time series using hidden Markov models’. Ph.D. thesis, University of California, Irvine
- Legras, B. and Ghil, M. 1985 Persistent anomalies, blocking and variations in atmospheric predictability. *J. Atmos. Sci.*, **42**, 433–471
- McBride, J. L. 1987 The Australian monsoon. Pp. 203–231 in *Monsoon Meteorology*. Eds. C. P. Chang and T. N. Krishnamurti. Oxford University Press, UK
- McBride, J. L. and Nicholls, N. 1983 Seasonal relationships between Australian rainfall and the Southern Oscillation. *Mon. Weather Rev.*, **111**, 1998–2004

- Madden, R. A. and Julian, P. R. 1971 Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708
- Manton, M. J. and McBride, J. L. 1992 Recent research on the Australian monsoon. *J. Meteorol. Soc. Jpn*, **70**, 275–284
- Mo, K. C. and Ghil, M. 1987 Statistics and dynamics of persistent anomalies. *J. Atmos. Sci.*, **44**, 877–901
- Molteni, F., Tibaldi, S. and Palmer, T. N. 1990 Regimes in the wintertime extratropical circulation over northern extratropics. I: Observational evidence. *Q. J. R. Meteorol. Soc.*, **116**, 31–67
- Molteni, F., Corti, S., Ferranti, L. and Slingo, J. M. 2003 Predictability experiments for the Asian summer monsoon: Impact of SST anomalies on interannual and intraseasonal variability. *J. Climate*, **16**, 4001–4021
- Nicholls, N. 1989 Sea surface temperatures and Australian winter rainfall. *J. Climate*, **2**, 965–973.
- Pittock, A. B. 1975 Climate change and the patterns of variations in Australian rainfall. *Search*, **6**, 498–504
- Reinhold, B. B. and Pierrehumbert, R. T. 1982 Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Weather Rev.*, **110**, 1105–1145
- Robertson, A. W. and Ghil, M. 1999 Large-scale weather regimes and local climate over the western United States. *J. Climate*, **12**, 1796–1813
- Robertson, A. W., Kirshner, S. and Smyth, P. 2003 ‘Hidden Markov models for modeling daily rainfall occurrence over Brazil’. Tech. Rep. ICS-TR 03-27, Information and Computer Science, University of California, Irvine
- 2004a Downscaling of daily rainfall occurrence over Northeast Brazil using a Hidden Markov Model. *J. Climate*, **17**, 4407–4424
- Robertson, A. W., Lall, U., Zebiak, S. E. and Goddard, L. 2004b Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Weather Rev.*, **132**, 2732–2744
- Rodwell, M. J. and Hoskins, B. J. 1996 Monsoons and the dynamics of deserts. *Q. J. R. Meteorol. Soc.*, **122**, 1385–1404
- 2001 Subtropical anticyclones and summer monsoons. *J. Climate*, **14**, 3192–3211
- Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U. and Schulzweida, U. 1996 The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate. Max-Planck-Institut für Meteorologie Report No. 218
- Smith, T. M., Reynolds, R. W., Livezey, R. E. and Stokes, D. C. 1996 Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420
- Smyth, P., Ide, K. and Ghil, M. 1999 Multiple regimes in northern hemisphere height fields via mixture model clustering. *J. Atmos. Sci.*, **56**, 3704–3723
- Suppiah, R. 1992 The Australian monsoon: a review. *Prog. Phys. Geogr.*, **16**, 283–318
- 2004 Trends in the Southern Oscillation phenomenon and Australian rainfall and changes in their relationship. *Int. J. Climatol.*, **24**, 269–290
- Suppiah, R. and Wu, X. 1998 Surges, cross-equatorial flows and their links with the Australian summer monsoon circulation and rainfall. *Aust. Meteorol. Mag.*, **47**, 113–130
- Syktus, J., McKeon, G., Flood, N., Smith, I. and Goddard, L. 2003 ‘Evaluation of a dynamical seasonal climate forecast system for Queensland’. Pp. 160–173 in Proc. National Drought Forum 2003: Science for Drought
- Troup, A. J. 1961 Variations in upper tropospheric flow associated with the onset of the Australian summer monsoon. *Indian J. Meteorol. Geophys.*, **12**, 217–230
- Vautard, R. 1990 Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Mon. Weather Rev.*, **118**, 2056–2081
- Wheeler, M. C. and Hendon, H. H. 2004 An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Weather Rev.*, **132**, 1917–1932
- Wheeler, M. C. and McBride, J. L. 2005 Intraseasonal variability of the Australian-Indonesian monsoon region. Chapter 5 in *Intraseasonal variability of the atmosphere-ocean climate system*. Eds. W. K. M. Lau and D. E. Waliser, Praxis Publishing
- Wilks, D. S. and Wilby, R. L. 1999 The weather generation game: a review of stochastic weather models. *Prog. Phys. Geogr.*, **23**, 329–357