Probabilistic Forecasts of the Onset of the North Australian Wet Season

FIONA LO AND MATTHEW C. WHEELER

Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

HOLGER MEINKE

Department of Primary Industries and Fisheries, Toowoomba, Queensland, Australia, and Department of Plant Sciences, Wageningen University, Wageningen, Netherlands

ALEXIS DONALD

Department of Primary Industries and Fisheries, Toowoomba, Queensland, Australia

(Manuscript received 11 December 2006, in final form 16 January 2007)

ABSTRACT

The amount and timing of early wet-season rainfall are important for the management of many agricultural industries in north Australia. With this in mind, a wet-season onset date is defined based on the accumulation of rainfall to a predefined threshold, starting from 1 September, for each square of a 1° gridded analysis of daily rainfall across the region. Consistent with earlier studies, the interannual variability of the onset dates is shown to be well related to the immediately preceding July–August Southern Oscillation index (SOI). Based on this relationship, a forecast method using logistic regression is developed to predict the probability that onset will occur later than the climatological mean date. This method is expanded to also predict the probabilities that onset will be later than any of a range of threshold dates around the climatological mean. When assessed using cross-validated hindcasts, the skill of the predictions exceeds that of climatological forecasts in the majority of locations in north Australia, especially in the Top End region, Cape York, and central Queensland. At times of strong anomalies in the July–August SOI, the forecasts are reliably emphatic. Furthermore, predictions using tropical Pacific sea surface temperatures (SSTs) as the predictor are also tested. While short-lead (July–August predictor) forecasts are more skillful using the SOI, long-lead (May–June predictor) forecasts are more skillful using Pacific SSTs, indicative of the longer-term memory present in the ocean.

1. Introduction

Rainfall over north Australia is highly seasonal, with the great majority of annual precipitation falling in the summer half of the year (November–April). At the peak of the summer season, this rain tends to preferentially occur within a number of separate multiday spells of widespread heavy falls (e.g., Troup 1961). Leading up to the first such spell, there is usually a transitional period, lasting a few weeks or more, in which relatively isolated rain-bearing storms become more frequent (Keenan and Carbone 1992). The timing of this transition to wet conditions, or "onset," however, can vary substantially from one year to the next, and between regions, making it an important consideration for decision making in agriculture. For some applications, this timing is more important than the overall wet-season rainfall total, as once the wet-season rains are established, the water availability becomes less of a limiting factor for agricultural production (McCown 1981; Mollah and Cook 1996). The goal of the work presented in this paper is to quantify the variability, and make predictions, of wet-season onset over north Australia, using a simple index of onset based on rainfall accumulation. While it is appreciated that there are other aspects of rainfall variability that are equally or perhaps more important for some agricultural decisions, such as the distribution within the wet season, in this paper we concentrate specifically on that near the start of the season. Our aim has been to produce a

Corresponding author address: Dr. Matthew C. Wheeler, BMRC, GPO Box 1289, Melbourne, VIC 3001, Australia.
E-mail: m.wheeler@bom.gov.au

DOI: 10.1175/MWR3473.1

forecast technique that can make predictions months in advance of the onset, can be applied operationally, and should be readily applicable to many agricultural practices and industries.

The meteorological explanation for at least part of the interannual variability in north Australian wet-season onset is the influence of the El Niño–Southern Oscillation (ENSO). Many previous studies have shown the existence of a relationship between Australian rainfall and indicators of ENSO, especially in the spring months of September–November (e.g., Priestley 1962; Nicholls et al. 1982; McBride and Nicholls 1983). During El Niño events, when surface pressures in north Australia are typically anomalously high, early wet-season rainfall tends to be reduced, causing onset (as defined in terms of rainfall accumulation) to be later. Very little relationship exists between ENSO and rainfall in the midsummer months of December–February, however, causing the ENSO association with the total wet-season rainfall to be weaker (Nicholls et al. 1982).

The relationship between indicators of ENSO and Australian spring rainfall also exists when a lag of several months is incorporated. Nicholls et al. (1982) found a significant correlation between austral winter (June–August) Darwin surface pressure (as contributes to half of the Southern Oscillation index, SOI) and rainfall-based indices of onset at Darwin (which occurs around November–December), thus demonstrating an ability to predict the onset of the wet season some months in advance. Nicholls (1984) furthered this work by developing an operationally feasible scheme for predicting an index of the wet-season onset date, as defined by rainfall data from 10 stations across the Australian Tropics. Using a logistic regression model and Darwin pressures as the predictor, Nicholls made forecasts of the probability that the wet season would commence later than the area-averaged climatological mean date.

This current study follows and expands upon the work of Nicholls (1984). Using an updated rainfall dataset with broader coverage, we apply a similar logistic regression model to predict the probability of wet-season onset being later than normal. Rather than measuring wet-season onset with a single index representative of the entire region like Nicholls (1984), we predict onset separately for each location in a gridded dataset. The principal predictor for which we present results is the July–August SOI. Other predictors, such as large-scale patterns of sea surface temperature (SST), and the SOI in earlier months, are applied for comparisons of forecast skill. It is notable that the timing of the onset shows a near-zero correlation with mid- to late wet-season (December–May) total rainfall (Nicholls et al. 1982); hence, the results presented should not be interpreted as indicative of the expected rainfall during the later part of the season.

Following a description of the data (section 2), this paper explores the characteristics of north Australian wet-season onset as defined using the gridded daily analysis of rainfall in section 3. For the calculations presented in this paper, we concentrate on a definition of onset that is the date at which an accumulation of 50 mm of rainfall is received from 1 September. We argue that an onset definition based on a constant threshold (e.g., 50 mm) is easier to understand, and in some instances more applicable, than is a date based on the accumulation of a certain percentage of the annual or wet-season mean rainfall, such as used by Nicholls (1984). Nevertheless, our methodology, which we present in sections 4–6, is adaptable and applicable to alternative onset definitions, and we provide a set of figures based on one such alternative, with the threshold set at 15% of the mean wet-season (1 September–30 April) rainfall (information online at http://www.bom.gov.au/bmrc/clfor/cfstaff/matw/NAWS/Onset/papersupp.html). Consistent with the work of Nicholls, the results show little sensitivity to such a change in the onset definition; if the forecast probability of a late onset by the 50-mm definition is high, then the probability of a late onset by the 15% definition will also be high.

Importantly, we discover a relatively large amount of spatial variability in the raw forecast probabilities, and we argue that it results from noise that is physically unrelated to the SOI predictor. A technique for removing this spatial noise, with a resultant improvement in skill, is developed in section 5d. Extension of the forecast methodology to multiple "threshold" dates, besides just the climatological mean date, is presented in section 6. This extension allows for the provision of the probability of exceedance information, that is, the probability that onset will be later than any of a range of threshold dates. A comparison of results obtained with the gridded dataset and a station dataset is provided in section 7, while the forecast skill obtained when using other predictors is presented in section 8.

2. Data

Two daily rainfall datasets are used in this study. The majority of results are obtained using a gridded analysis; however, a dataset comprising a subset of high quality stations is also used for evaluation and comparison purposes. Both datasets were obtained from the National Climate Centre of the Australian Bureau of Meteorology.

The gridded analysis is derived from daily observa-

tions of station rainfall data, utilizing the entire climatological rainfall network of approximately 6000 stations across Australia. A Barnes successive-correction analysis is applied to the station data with a correlation length scale of 80 km (Mills et al. 1997). The grid onto which the analysis has been performed is a regular 0.25° latitude–longitude grid, although the data used for this study are area averaged onto a 1° grid, and we focus on the portion north of 28°S. Importantly, the analysis aims to provide accurate estimates of daily rainfall averaged over an area rather than accurate estimates of point values, hence the need for a comparison of results with those obtained from the station data (section 7). By construction, the gridded analysis has no missing data, but areas where there are insufficient observations are masked out of the plots. These areas occur in the relatively unpopulated deserts of inland Western Australia, the southern Northern Territory, and northern South Australia (see Fig. 9 for the location of state boundaries and the Web site http://www.bom.gov.au/climate/averages/ for climatological rainfall distributions). We are confident of the validity of the gridded analyses over the rest of north Australia beginning in 1948; thus, there are 57 wet seasons available for use from 1 August 1948 to 31 July 2005.

The high quality station dataset is composed of rainfall data from a subset of the entire network of stations, chosen with a long-term climate analysis in mind (Haylock and Nicholls 2000). Station rainfall records are not adjusted for discontinuities but are rejected if they are suspected of having data-quality problems. There are 26 high quality stations available for the region north of 28°S. Unless otherwise specified, the results presented here use the gridded rainfall dataset.

The SOI we use is the Troup version, as published by the Australian Bureau of Meteorology (online at http://www.bom.gov.au/climate/current/soihtm1.shtml). Sustained negative SOI values are indicative of El Niño–like conditions, and positive SOI values of La Niña–like conditions. The majority of the analysis and development uses the mean of the July and August monthly SOI values as a predictor. In section 8, however, skill scores of forecasts using other months of SOI data and a large-scale pattern of SST anomalies as predictors are presented.

The large-scale pattern of SST anomalies we evaluate as a predictor is that derived by Drosdowsky and Chambers (2001). They generated SST patterns through a rotated principal component analysis of monthly Indo-Pacific SSTs. The spatial pattern of the first principal component shows strong loadings in the Pacific, and its associated time series (SST1) is well correlated with other typical sea surface signatures of ENSO (e.g., Niño-3), while the second component (SST2) is correlated with Indian Ocean SST patterns. Both the first (SST1) and second (SST2) rotated principal components are used operationally as predictors of Australian seasonal rainfall by the Bureau of Meteorology, but only SST1 clearly demonstrates an ability to provide useful skill in north Australia in spring (Drosdowsky and Chambers 1998). Thus, SST1 is tested here as an alternative predictor to the SOI, for which values from 1949 to the present are available (online at http://www.bom.gov.au/climate/ahead/sst_data_table.html).

3. Definitions and characteristics of wet-season onset dates

a. Definitions of onset

Rainfall variability is a major source of risk in agricultural systems (Meinke and Stone 2005). In this paper, we focus on one of the characteristic features of rainfall variability in the Tropics—the timing of the seasonal transition to wet conditions, which is known as wet-season onset. For this we require an objective definition of onset that focuses specifically on rainfall received at the beginning of the wet season, particularly that during the so-called transitional period.

It should be noted that we distinguish between the onset of the wet season and the onset of the meteorological monsoon. Monsoon onset in north Australia is associated with a large-scale rearrangement of the tropospheric circulation, characterized by a relatively rapid shift in the low-level predominant winds from easterly to westerly, and typically occurs later than the first agriculturally or ecologically significant rains (Troup 1961; Nicholls et al. 1982; Drosdowsky 1996; Cook and Heerdegen 2001). Indeed, a considerable portion [~30%; Nicholls et al. (1982)] of total wet-season rain can fall before the large-scale change in circulation associated with the monsoon. In this study we concentrate on definitions of wet-season onset that are meaningful from either a biological (e.g., sufficient rainfall to stimulate pasture growth) or a managerial perspective (e.g., an absolute threshold of rainfall that might prevent any further movement of grazing stock). These definitions were developed via a consultative process with the grazing industry in north Australia (data not shown) and are defined exclusively by rainfall. While they each provide a date for onset, this does not necessarily imply that the transition to wet conditions is rapid.

One definition under consideration was the date at which an accumulation of 50 mm of rainfall is reached

in 10 days or less, starting from any day after 1 September (but before 31 March). Such a duration-limited definition requires a relatively strong weather event, and while it can be quite applicable in the far north of the region, it is not particularly suited to making management decisions in the more arid areas.

Another definition tested was the date at which an accumulation of 15% of a station's climatological mean wet-season (September–April) rainfall is reached, from 1 September. This is similar to that used by Nicholls (1984), except he used 15% of the annual mean. This definition results in a relatively uniform onset date over north Australia, with a climatological mean date in most locations falling in December. The amount of rainfall defined by 15% of the wet-season mean, however, is quite spatially variable; near Darwin it exceeds 220 mm, but it is less than 40 mm in the southern third of the Northern Territory. Thus, as argued by Cook and Heerdegen (2001), the same plant species growing in these different locations would likely be at very different stages of growth at the time of onset defined this way. Nevertheless, the plant species mix varies considerably with location (Ash et al. 2002), and such a definition is still quite biologically meaningful (as ascertained through our consultative process).

Perhaps the simplest of definitions is based on the date at which an accumulation of an absolute amount of rainfall is reached after 1 September, with no duration limit. We consider a 50-mm-accumulation threshold. This definition has several advantages: it requires no prior knowledge of the climatological annual or wet-season mean, it is easily measured and simple to understand, and it is readily applicable to certain, mainly management related, decisions (e.g., the mustering and mechanized transport of cattle becomes impossible once a certain absolute rainfall threshold is reached). Nevertheless, arguments can be made for and against the applicability of each definition, and each management decision is likely to require a different threshold. In this paper we have chosen to present results entirely for this final definition: the accumulation of 50 mm from 1 September. However, noting differences in needs, we provide a complementary set of results and figures using the 15% threshold online (http://www.bom.gov.au/bmrc/clfor/cfstaff/matw/NAWS/Onset/papersupp.html).

The onset dates are a function of both location and year. For the selected 50-mm definition, onset typically occurs earlier in the north and on the coast of Queensland, and occurs later farther south. In some dry areas, it may take multiple rain events spread over multiple months for the rainfall to accumulate to 50 mm, and in some years this threshold may never be reached during the traditionally defined wet-season period. If the 50 mm has not been accumulated by 31 March, onset is defined as not occurring, and a special value is assigned. In most cases the locations and years of a very late or nononset are physically meaningful. However, some of the late or "not reached" onset dates computed from the gridded analyses are artificial, and can be attributed to the lack of data or the scarcity of stations. When there is a lack of nearby observations, the analysis is regressed toward zero rainfall, thus biasing the onset toward later dates than actual. Regions with few stations are masked out of our analyses and plots, but this does not preclude the possibility of biased onset dates in areas bordering those masked-out regions or in some other locations during periods when station data are unavailable.

b. Onset mean and variability

It is useful to know the typical or time-mean onset date for a particular location. Knowledge of a mean date allows the calculation of onset date anomalies. However, a problem with the calculation of the arithmetic mean arises at locations where there are years in which the onset threshold is not reached; if the not-reached years are excluded from the calculation, the mean calculated from the rest of the dates is artificially early. We thus use the approach of a "trimmed" mean (Wilks 1995) and compute a climatologically typical date in the following manner: 1) count the number of not-reached years in the record and omit them, (2) sort the onset dates of the remaining years from earliest to latest, 3) omit an equivalent number of years with the earliest dates, and 4) compute the mean of the remaining dates, rounded to the nearest integer. These remaining dates represent the least extreme years of onset on record. If the wet-season onset threshold is not reached in half of the years on record or more, then the above-described calculation cannot be made for that location, and it is flagged as being later than 31 March. Henceforth, we refer to the date computed using the above procedure simply as the "mean" onset date at a location. In most locations it is no different from the arithmetic mean of all years, but at some locations it helps considerably to reduce artificial biases.

In north Australia, mean onset (50-mm threshold definition) occurs earliest in the Top End[1] region and two areas on the Queensland east coast, beginning as early as mid-October (Fig. 1a). It gradually becomes

[1] The "Top End" is the name given to the northernmost quarter of the Northern Territory (see Fig. 9).

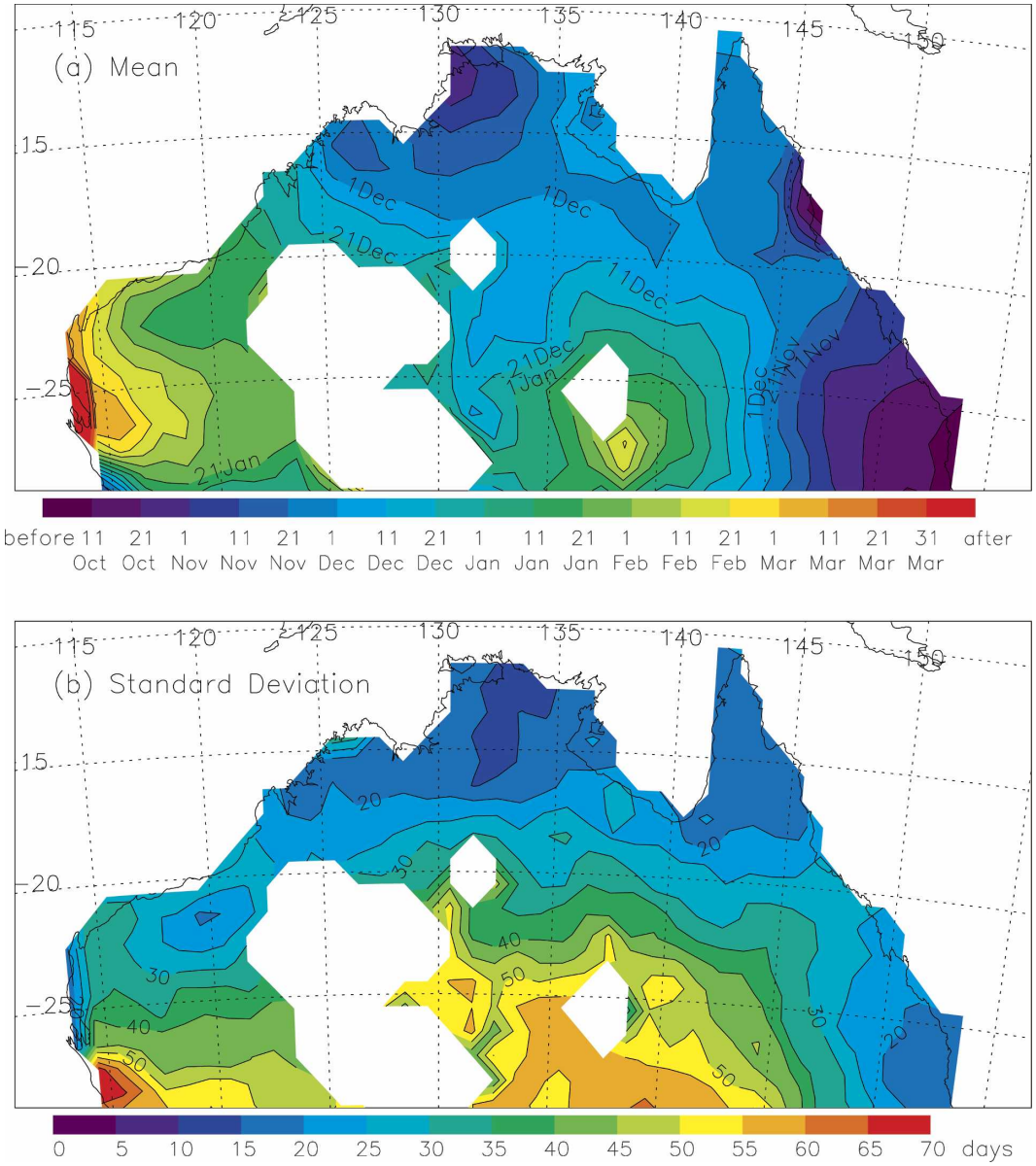

FIG. 1. (a) Climatological mean (1948/49–2004/05) wet-season onset date (50-mm-threshold definition). White areas denote regions of insufficient observational rainfall data. (b) Standard deviation of wet-season onset dates (50-mm-threshold definition).

later with increasing latitude in the west, and distance from the coast in the east. The latest mean onset dates, by the 50-mm definition, are in the far west, occurring in late March. Whether or not such a late onset is climatologically meaningful in these extreme cases is debatable. Here, we are more concerned with the usefulness of the information for decision making. Importantly for decision makers in the cattle industry, this map of mean onset dates bears a close resemblance to the map of median dates of commencement of the "green season" as defined by McCown (1981), which corresponds to the main period of cattle weight gain when fed on native pastures, computed using a water balance model.

The magnitude of the interannual variability of the onset dates can be appreciated by examining their standard deviation (Fig. 1b). If the onset threshold is not reached in a particular year, the default date of 31 March is used as input to the calculation. There is a pattern of increasing variability to the south and inland from the east coast. The least variability occurs in the Top End region, where the standard deviation is about

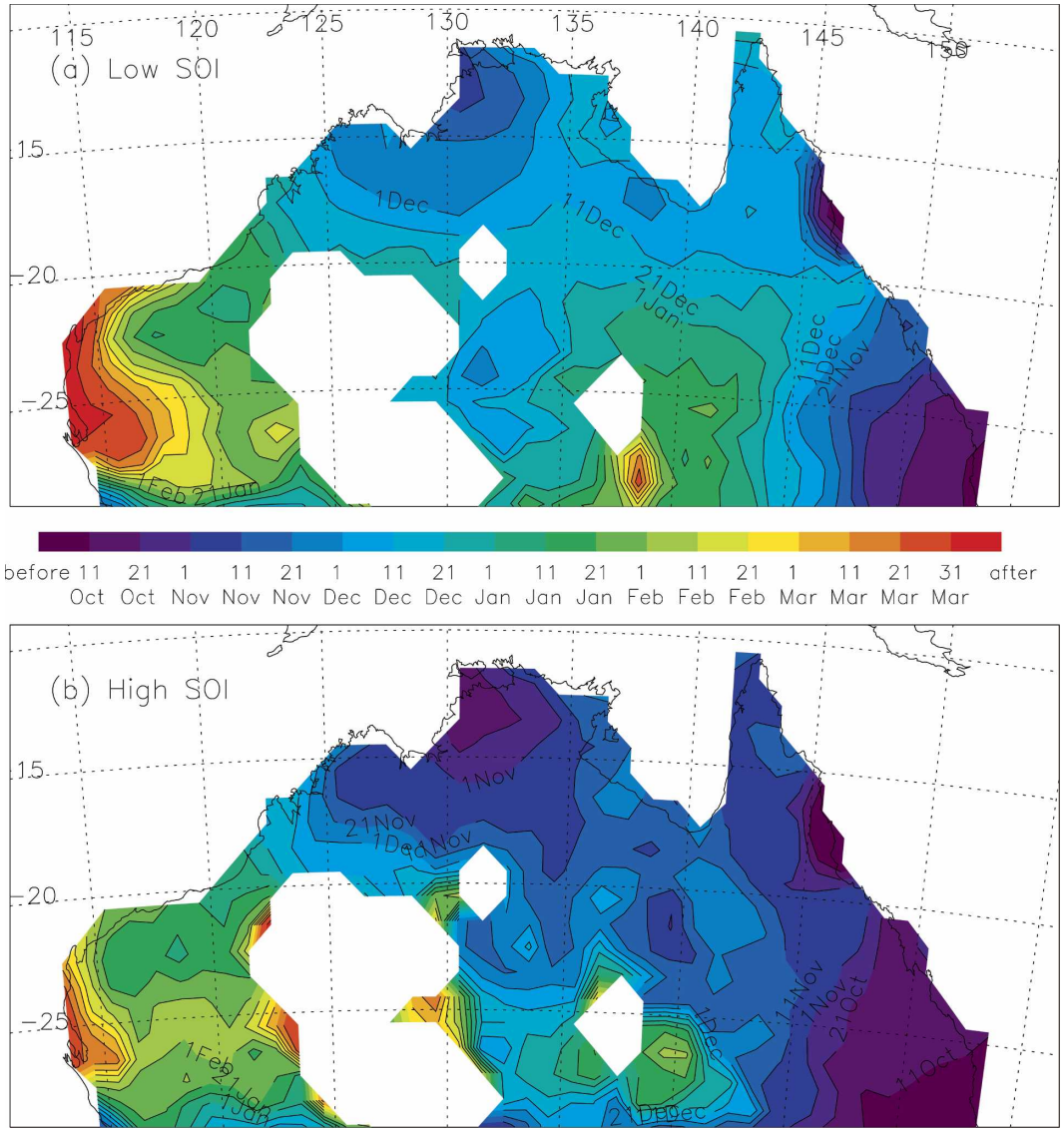

FIG. 2. Stratified climatological mean wet season onset dates for (a) low-SOI (<−10) years (1965, 1972, 1976, 1977, 1982, 1987, 1993, 1994, 1997, and 2002) and (b) high-SOI (>+10) years (1950, 1955, 1956, 1964, 1975, 1988, and 1998).

10 days, and greatest in the desert regions of northern South Australia, reaching values of up to 55 days and over.

4. ENSO influence on onset date

As discussed in the introduction, the existence of a relationship between ENSO and the onset of the north Australian wet season is already well established. To explore the specific relationship using our defined dates of wet-season onset, we examine the stratified climatological mean onset date for contrasting high and low SOI years. Figure 2 shows stratified means generated using thresholds of ±10 of the July–August SOI. This results in 7 "high" and 10 "low" years from a record of 57 yr.

The stratified climatological mean dates, computed using the same procedure as described in the previous section, confirm the tendency for wet-season onset to occur later in low-SOI (El Niño–like) years, and earlier in high-SOI (La Niña–like) years (Fig. 2). The difference in mean dates between the low- and high-SOI years ranges from about 3 weeks in the region north of 20°S, to 10 weeks in the southeast of the Northern Territory and southwestern Queensland. Notably, the magnitude of such differences is comparable to the inter-

annual standard deviation of the onset dates (Fig. 1b); thus, we surmise that ENSO can potentially account for a large portion of the interannual variability. This result confirms the feasibility of using wintertime SOI values as a predictor of our defined wet-season onset dates.

5. Forecast methodology

a. Logistic regression model

As in Nicholls (1984), we employ logistic regression as our statistical forecast model. Logistic regression is particularly suited for making probabilistic forecasts, as it has the property that it will never forecast a probability of less than zero or greater than one (Wilks 1995). The logistic regression equation fit to the data is

$$\hat{P} = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)},$$

where $\hat{P}$ is the predicted probability of a late occurrence of onset and x is the predictor. The fitted coefficients, b_0 and b_1, are computed through an iteratively reweighted least squares approach.

Unless indicated otherwise, the predictor used is the July–August SOI preceding the wet-season onset. The logistic regression model is applied separately to each square of the gridded rainfall analyses. At each grid location, each year on record is categorized as being late or early, according to the sign of the onset date anomalies. For onsets occurring later than the mean date, the binary predictand, P, is given a value of 1; for onsets on or earlier than the mean date, P is given a value of 0.

For example, Fig. 3 shows the logistic regression model fitted for the grid square closest to Darwin, centered at 12.5°S, 130.5°E. Crosses indicate the observed onset probabilities, P, and the solid curve indicates the forecast probabilities of the fitted model, $\hat{P}$, each as a function of the SOI predictor. There are 57 crosses in total, one for each observation year. The tendency for low-SOI years to be more likely to have a late onset ($P = 1$) is apparent.

The sensitivity of the logistic regression model to small changes in input data is assessed by successively removing 1 yr of data, and recalculating, producing an ensemble of 57 logistic regression curves. Taking the upper and lower extremes of the 57 curves gives the dashed curves in Fig. 3. Thus, the sensitivity to removing a year of input data can be seen to be relatively small, causing at most a variation to the forecast probability of about 0.07 in this example.

Also displayed in Fig. 3 are the observed probabilities when averaged into six bins of 9 or 10 yr each (the diamonds). The probabilities are grouped according to like values of the SOI. That is, the left-most diamond is for the observed probability of late onset when averaged for the nine lowest SOI years, plotted against the average SOI value for those same years. The next diamond from the left is for the next group of 9 or 10 yr, and so on. Note that the averaged probabilities (diamonds) are computed independently of the logistic regression model (curve); thus, the match between the averaged observations and logistic regression model supports the use of this model.

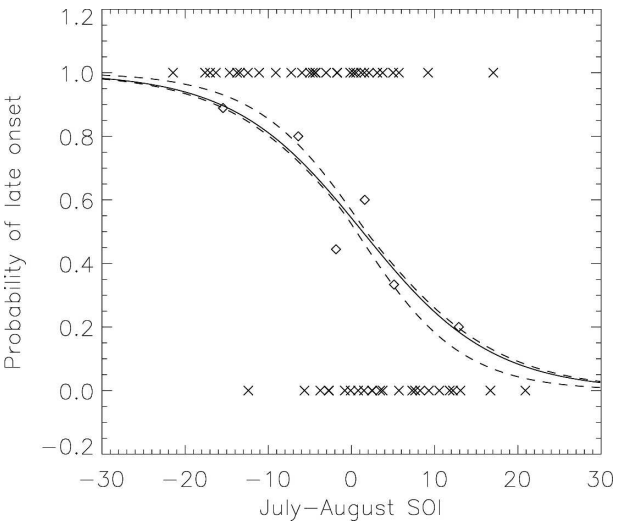

FIG. 3. Example logistic regression model ($\hat{P}$; solid curve) fitted for the grid square closest to Darwin (centered at 12.5°S, 130.5°E), using all 57 yr of input data. Probabilities (as a fraction) are the chance of onset occurring later than the climatological mean date, as a function of the July–August SOI. Crosses are the observed onset probabilities (P) for all 57 yr. An observed probability of zero indicates that onset occurred on or before the mean date, while a value of 1 indicates onset occurred later. The diamonds are calculated averages of observed probabilities and SOI values, binned into six groups of 9 or 10 yr each. The dashed curves provide the upper and lower bounds of logistic regression models computed by all combinations of 56 yr only (i.e., by successively leaving 1 yr of input data out), and provide an indication of the sensitivity of the model.

b. Cross-validated hindcasts

An assessment of the skill of the forecast model requires testing it on a set of data that is independent to the data used to develop the model. To achieve this, a cross-validation technique is used to generate hindcasts for each year at each grid location. The year to be forecast is left out of the training dataset, leaving 56 yr for the computation of the logistic regression model coefficients. This recalculation of the logistic regression

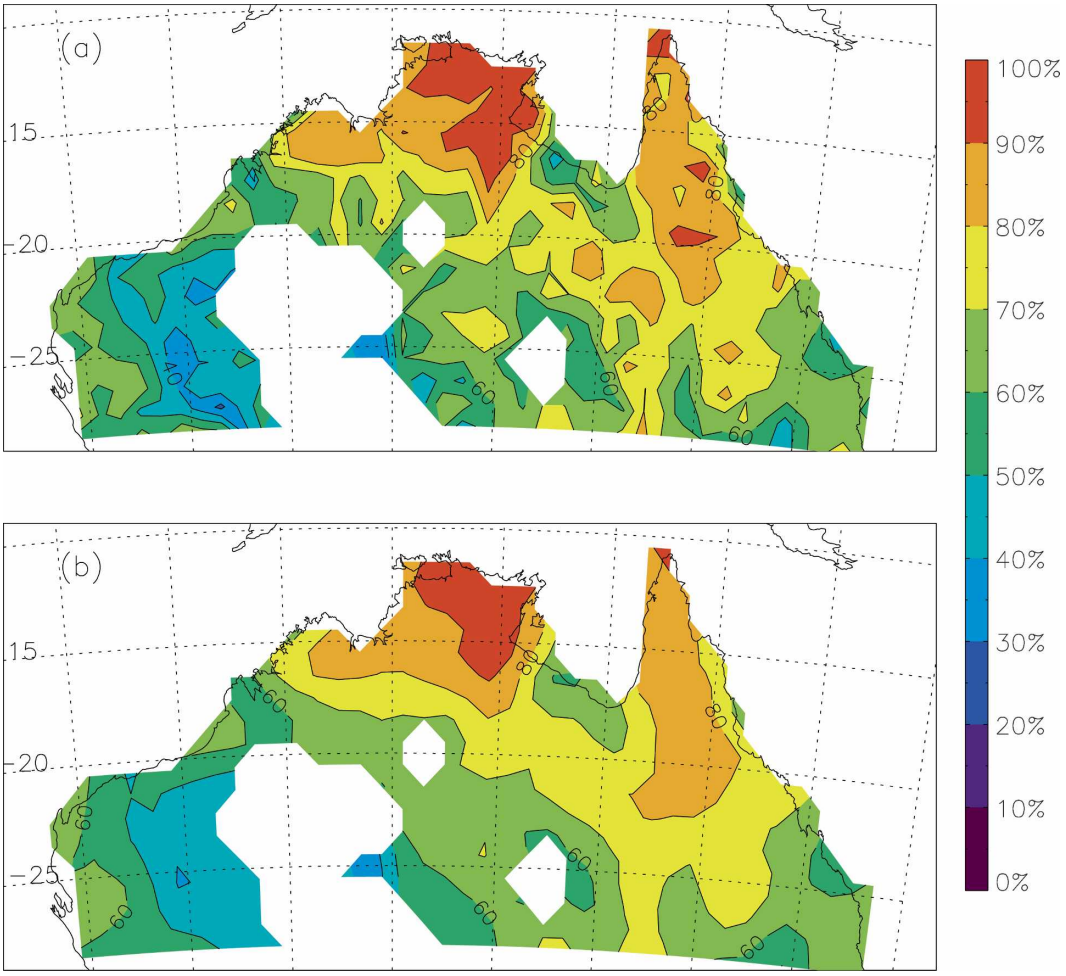

FIG. 4. (a) Forecast probabilities (as a percentage) of later than mean onset from the example cross-validated hindcast of the wet season of 1997–98. (b) As in (a) except that forecast probabilities are smoothed in space using one application of a 1–2–1 filter in each spatial direction.

parameters is done 57 times for each grid square, producing a complete set of "cross-validated" hindcasts (and thus permitting the calculation of the dashed curves in Fig. 3). The autocorrelation of the onset dates at 1 yr lag is very low, being no greater than 0.3 for any grid location, indicating an almost-complete independence between years, and justifying the removal of only 1 yr at a time from the training dataset.

c. *Removing spatial noise*

The forecast probabilities, generated independently at each grid location, may be mapped together, as we show for the cross-validated hindcast for 1997–98 in Fig. 4a. The SOI in July–August 1997 was strongly negative; hence, this example hindcast shows generally high probabilities of a later than normal onset, varying from around 50% in the south and west, to greater than 90% in the far north. However, such maps contain a great deal of spatial variability. Given the general lack of large topography in north Australia, except near the Queensland coast, we believe such spatial variability is unrelated to the ENSO signal; the average climatic impact of ENSO events, resulting from shifts in large-scale circulation patterns, should be highly spatially coherent, except in locations where topography can provide contrasting effects on the lee- and windward sides. Thus, much of the spatial variability can be ascribed to be a consequence of noise in the input data and the limited record length. Therefore, we remove some of this spatial noise by smoothing the forecast probabilities in space using a 1–2–1 smoother, applied once in each of the latitudinal and longitudinal dimensions (identical to a 3×3 binomial smoother; Fig. 4b). We are confident that the spatially smoothed forecasts

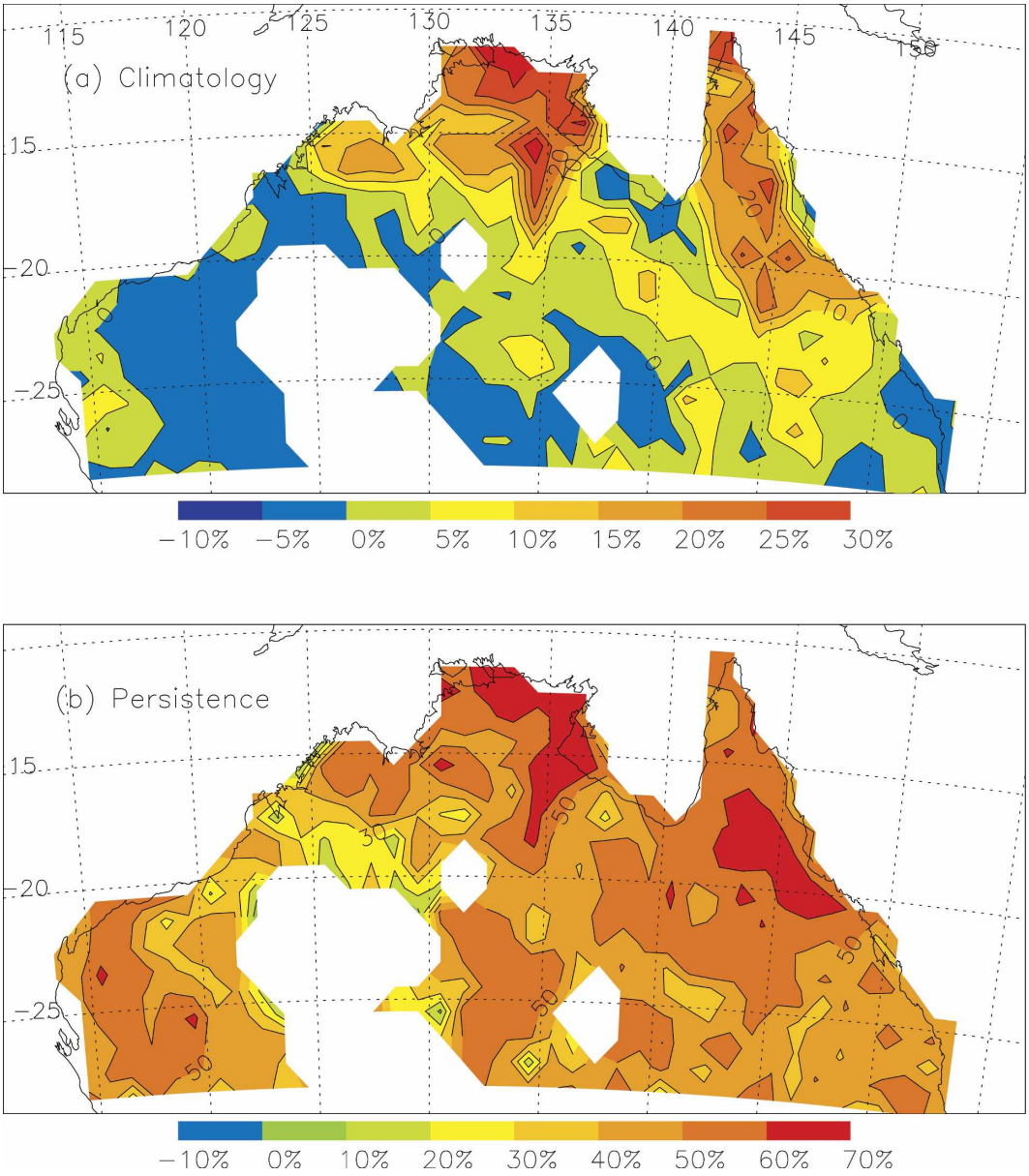

FIG. 5. Brier skill scores (as a percentage improvement) of the spatially smoothed cross-validated hindcasts in comparison to reference forecasts made with (a) climatology and (b) persistence.

more accurately reflect the true probability, confirmed by increased skill, to be discussed in the next section.

d. Measures of skill and reliability

A forecast is only useful if it is both skillful and reliable. A common method of evaluating the skill of probabilistic forecasts is the Brier skill score [Wilks 1995, Eq. (7.23)]. The Brier skill score provides an indication of the level of improvement of a set of forecasts compared to that of a reference forecast strategy. Brier skill scores are presented here as a percentage improvement over reference forecasts obtained using either climatology or persistence (Fig. 5; Table 1). The climatological forecast is computed from all years except the year to be forecast, and persistence is taken from the previous year.

The Brier skill scores for the spatially smoothed cross-validated hindcasts demonstrate that improvement over climatology is almost always less than the improvement over persistence. This is because persistence is generally a poor forecast, a consequence of the relatively low 1-yr lag autocorrelation of the onset

TABLE 1. Brier skill scores of cross-validated hindcasts made relative to the mean onset as a threshold date. All valid grid squares north of 28°S are used.

	Raw (%)	Spatial smoothing (%)	Spatial and temporal smoothing (%)
Climatology	4.36	5.65	5.62
Persistence	46.62	47.29	47.32

dates. As our consultative process with the grazing industry revealed, however, a comparison to persistence is still very relevant. This is because there is a high rate of turnover of managers of some climate-sensitive industries in north Australia, resulting in decisions sometimes being based on a short range of experience only, such as knowledge of the previous year.

Greatest skill occurs in the Top End region of the Northern Territory, in Cape York, and in central Queensland. In these regions the improvement over climatology is greater than 5%, and in some locations is higher than 25% (Fig. 5a). The improvement over persistence in those locations ranges from 50% to greater than 60% (Fig. 5b). In the desert regions of central Western Australia, the southern Northern Territory, and northern South Australia, however, there is no improvement over climatology. The coastal fringe of the southern end of the Gulf of Carpentaria also has a low level of skill. As expected, this spatial pattern of skill bears similarities to the spatial pattern of correlations shown between wintertime indices of ENSO and September–November rainfall by McBride and Nicholls (1983), and bears close similarity to the skill map for forecasts of September–November rainfall presented by Drosdowsky and Chambers (2001).

Brier skill scores computed when all hindcasts at all locations north of 28°S were combined are presented in Table 1. The raw, unsmoothed, forecasts show an improvement over climatology of 4.36%, and an improvement over persistence of 46.6%. Confirming the arguments made in section 5c, the spatial smoothing of the forecasts increases the skill, resulting in skill improvements of 5.65% and 47.3% over climatology and persistence, respectively.

In addition to the Brier skill score, a reliability diagram shows an informative graphical display of the reliability of a set of probabilistic forecasts. Figure 6 presents the reliability diagram constructed using all cross-validated hindcasts from every valid grid square north of 28°S. The forecasts and corresponding observations are grouped by increasing forecast probability into bins of 95 members each. The observed frequency of occurrence is computed and plotted against the average of the 95 forecast probabilities in the bin. There are 273 bins in total. The diagonal black line indicates the line of perfect forecast reliability. The forecasts appear to be quite reliable, and the spread of the observed probabilities from values of less than 0.2 to greater than 0.9 indicates a relatively high degree of resolution. However, a slight tendency to overforecast low probabilities and underforecast high probabilities is also apparent.

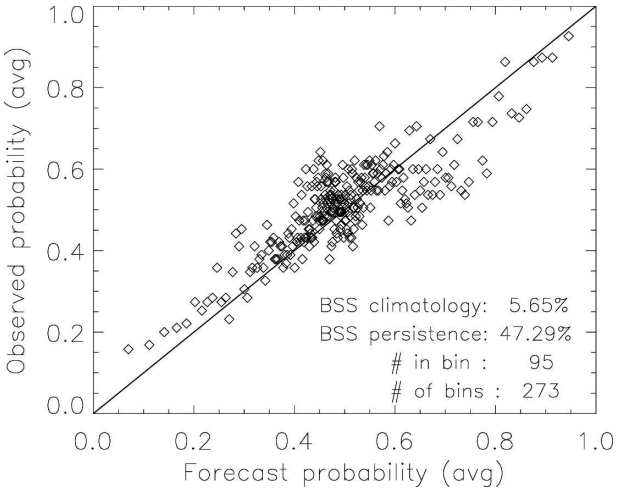

FIG. 6. Reliability diagram representing all cross-validated hindcasts at all valid grid locations north of 28°S. Average probabilities were calculated by sorting forecasts into 273 bins of 95 members each, in order of increasing forecast probability. "BSS" refers to the Brier skill score of these forecasts, with respect to either climatology or persistence.

6. Extending the forecast methodology to multiple threshold dates

In addition to knowing the probability of onset being later than the mean date, it is useful to be given an indication of how late, or early, onset is likely to be. Such an indication is computed by forecasting probabilities of onset being later than a range of arbitrary dates, not just later than the mean date alone. The date at which we evaluate the lateness of onset will be referred to as the "threshold date." In previous sections the threshold date was the climatological mean onset date. The forecast methodology is extended to a range of threshold dates, from a month before the mean, to a month afterward. That is, the steps in sections 5a–c are repeated for each different threshold date.

For example, at the nearest grid location to Darwin, the mean onset date is 25 October. Previously, the logistic regression model was constructed to evaluate and forecast the lateness of onset with respect to the mean date. The process is now repeated using threshold dates from 25 September to 24 November, constructing a

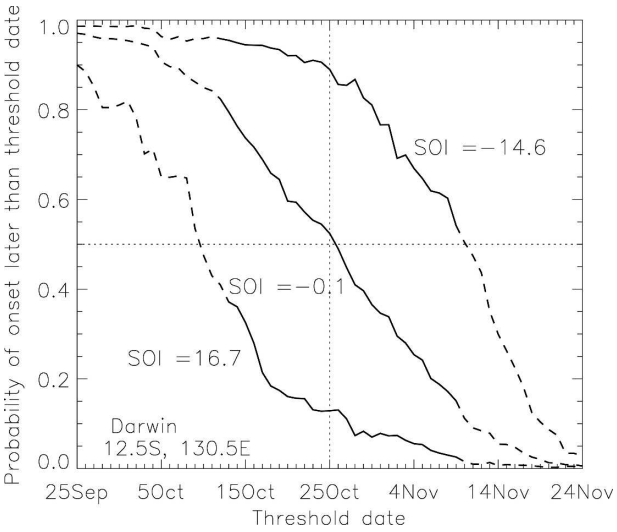

FIG. 7. Probability of exceedance curves, from the cross-validated hindcasts, for the grid location nearest to Darwin (12.5°S, 130.5°E) for an example high-SOI year (1950; SOI = 16.7; lower curve), low-SOI year (1997; SOI = −14.6; upper curve), and neutral-SOI year (1952; SOI = −0.1; middle curve). Dashed lines are plotted when the model is developed with fewer than 10 observations of either a late of early onset and, thus, represent where there is a higher level of uncertainty in the forecast. The mean onset date at this location is 25 Oct.

separate logistic regression model for each calendar date. For onsets occurring later than the threshold date, the predictand, P, is given a value of 1; for onsets on or earlier than the threshold date, P is given a value of 0. For each threshold date and for each grid square, multiple (57) logistic regression models are constructed leaving 1 yr of input data out at a time (as in section 5b), generating a complete set of cross-validated hindcasts, which are then smoothed in space (as in section 5c).

It is now possible to plot the probability of onset for a particular forecast year (or SOI value), as a function of threshold date (e.g., Fig. 7). As expected, the probability of onset later than a certain threshold date decreases as the date becomes later. At Darwin, in a near-neutral ENSO year such as 1952 (SOI = −0.1), we forecast a 90% probability that onset will occur after 5 October, but only a 10% probability that it will be later than 10 November (Fig. 7). These thresholds change, however, with the value of the SOI, as we display in Fig. 7 for three different values. Such a presentation of forecast probabilities is akin to a "probability of exceedance" curve.

Probability of exceedance curves are useful for selecting the date at which a management action should be taken for a certain level of risk (Hammer et al. 2001; Maia et al. 2007). For example, an action that is risk averse to dry conditions could be taken at the date at which there is only a 20% probability of onset having not yet occurred. In a high-SOI year, like 1950, such an action could be taken as early as 17 October at Darwin, but in a low-SOI year, like 1997, the action could be delayed until 16 November (Fig. 7). The skill of such forecasts, using all threshold dates from 2 weeks before the mean date to 2 weeks afterward, is presented in Table 2. The Brier skill scores indicate that the forecasts with respect to these varying threshold dates (5.12% improvement over climatology) are almost as skillful as those with respect to the mean date (5.65% improvement over climatology; Table 1).

TABLE 2. Brier skill scores of cross-validated hindcasts made relative to threshold dates ±2 weeks of the mean date. All valid grid squares north of 28°S are used.

	Spatial smoothing (%)	Spatial and temporal smoothing (%)
Climatology	5.12	5.20
Persistence	47.14	47.19

Ideally, the probability of exceedance curves should be monotonically decreasing. However, because a separate logistic regression model is fitted to the data for each threshold date, this is not guaranteed. When the threshold date is shifted to later in the calendar, this causes one or more of the observed onsets to change its binary classification from being late ($P = 1$) to early ($P = 0$). If the onset year that changes its classification is an outlier to the logistic regression model, a nonnegligible change in the fitted logistic regression curve will result and cause a bump in the probability of exceedance curve. Such irregularities are of the same size as the deviations in probability produced when random years were left out of the model training set, as assessed with the dashed curves presented in Fig. 3. That is, the irregularity is a consequence of our limited record length, rather than an intrinsic feature of the climate system.

To eliminate the irregularities of the probability of exceedance curves, and to ensure a potentially more meaningful forecast, the probability of exceedance curves are smoothed in time with 10 applications of a 1–2–1 smoother. An example of the result of this smoothing is shown for the grid square nearest to the town of Katherine in the Northern Territory (centered at 14.5°S, 132.5°E) in Fig. 8. In this example, showing the cross-validated forecast probabilities for 1988 (SOI = 13.1), the smoothing transforms a bumpy probability of exceedance curve into one that is monotonically decreasing. The skill of the cross-validated hindcasts, as assessed with the Brier skill score, is either

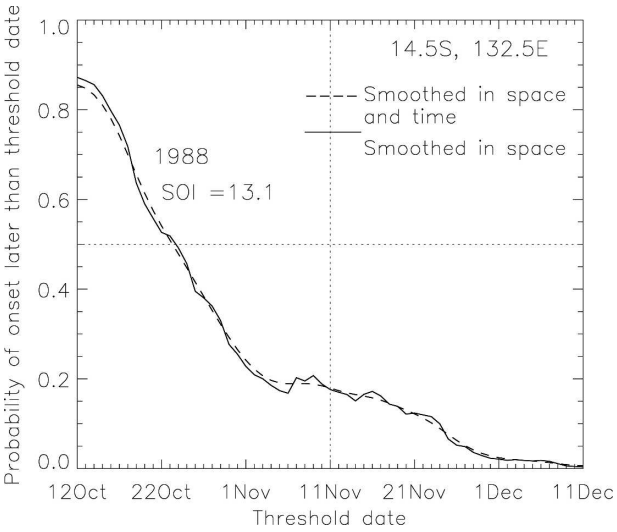

FIG. 8. As in Fig. 7 but for the grid location nearest to Katherine (14.5°S, 132.5°E) for the year 1988. The solid line is from the hindcasts smoothed in space only, while the dashed line also has smoothing in the temporal (threshold date) dimension.

essentially unchanged (Table 1), or slightly improved (Table 2), by this additional smoothing.

7. Comparison with station data

The gridded rainfall dataset used is intended to be representative of rainfall area averaged over a 1° grid. While it can be argued that a 1° average is very applicable to some industries in north Australia (e.g., the cattle industry, whose property holdings are as large as 12 000 km^2), it is important to demonstrate how well the wet-season onset dates and forecasts calculated from the gridded dataset compare to onset dates and forecasts calculated from station data. For this purpose, we have used a subset of high-quality stations described in section 2, for which there are 26 available north of 28°S (Fig. 9).

The observed onset dates calculated from the station data agree quite well with those calculated from their closest corresponding grid square (not shown). Excluding years in which there were insufficient station data, the correlation coefficients of observed onset dates for the different datasets are not less than 0.7.

But how skillful are the gridded data forecasts at predicting onset as measured at a station? Are forecasts developed with the station data more skillful at predicting the station observations than the forecasts made with the gridded data? To determine this, cross-validated hindcasts were computed from the station data using the same techniques as described in sections 5 and 6, except no spatial smoothing was conducted because the high quality stations have insufficient uniformity and density. Comparisons were made with the cross-validated hindcasts generated with the gridded data that were smoothed in space and time. Both sets of forecasts were verified against the station observations. Note that this is a considerably more stringent test than verifying the gridded data forecasts with gridded observations.

The Brier skill score was computed for forecasts with respect to threshold dates extending from 2 weeks before the station mean date, to 2 weeks afterward. At 19

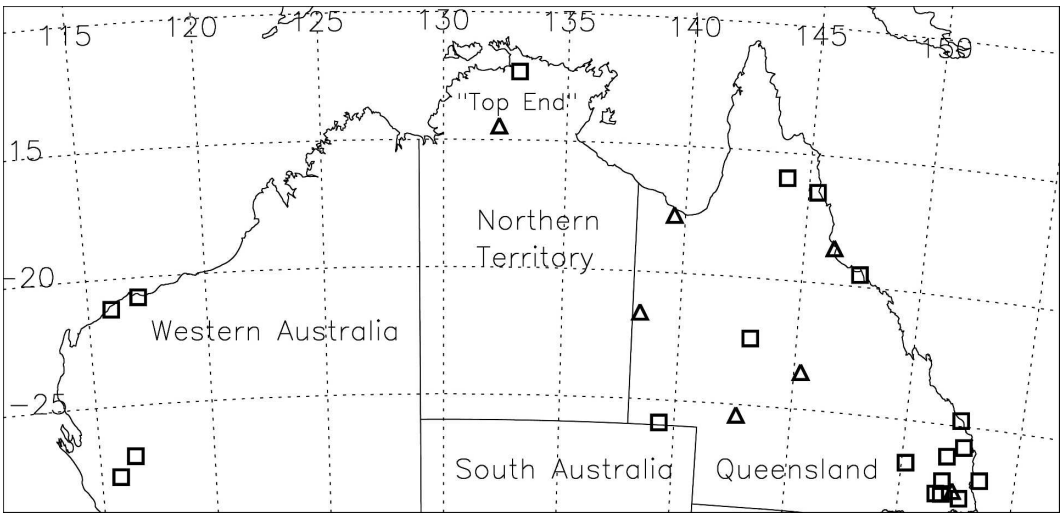

FIG. 9. Locations of high quality stations and state boundaries. A square symbol is used at locations where the hindcasts derived from the gridded data were more skillful, and a triangle is used at locations where the station-derived hindcasts were more skillful (see section 7).

TABLE 3. Brier skill scores of cross-validated hindcasts (smoothed spatially and temporally) with either the SOI or SST1 during particular pairs of months as predictors, relative to both climatology and persistence as reference forecasts. All valid grid squares north of 28°S are used, and the threshold date is the mean date.

	SOI (climatology, %)	SOI (persistence, %)	SST1 (climatology, %)	SST1 (persistence, %)
May–Jun	2.92	45.89	3.26	45.99
Jun–Jul	4.29	46.61	4.61	46.72
Jul–Aug	5.62	47.32	4.80	46.81

of the 26 station locations (73%), the gridded hindcasts were more skillful than the corresponding station-derived hindcasts (squares versus triangles in Fig. 9). For example, at Oenpelli (12.3°S, 133.1°E), the northernmost station, the gridded hindcasts provided a 2.72% improvement over the station-derived hindcasts. Even in areas where the SOI and onset date relationship is relatively weak, such as the Queensland coast (Fig. 2), the gridded hindcasts still generally performed better than the station-derived hindcasts (Fig. 9). For example, at South Mossman (16.5°S, 145.4°E), the gridded hindcasts provided a 3.56% improvement.

In summary, while not designed to provide accurate estimates of rainfall at point locations, the gridded data can produce more skillful forecasts of wet-season onset when verified at nearby stations, than forecasts developed from the stations themselves. The reason for this is the increased information coming from surrounding locations in the gridded analysis. Information from nearby areas effectively increases the signal to noise ratio in the limited 57-yr dataset. The gridded data also have another great advantage; with almost-complete spatial coverage, they allow for the production of forecast maps and, thus, provide almost all geographic locations with potentially useful information.

8. Other predictors

The forecast skill obtained when using other predictors, besides the July–August SOI, was investigated (Table 3). As expected, when using SOI values from earlier pairs of months, the longer the forecast lead time, the poorer the skill. For example, the Brier skill score computed using the May–June SOI as the predictor shows an improvement over climatology of only 2.92%, whereas for the July–August SOI it was 5.62%. A similar trend of decreasing skill with longer lead time exists in the Brier skill scores computed with respect to persistence.

The Australian Bureau of Meteorology currently issues seasonal rainfall forecasts based on the statistical technique of linear discriminant analysis, using two large-scale patterns of SST as predictors (Drosdowsky and Chambers 2001). The first is representative of SST conditions across the Pacific (SST1), and the second representative of the Indian Ocean (SST2). The rainfall forecast skill during spring in north Australia is derived mostly from the Pacific pattern (Drosdowsky and Chambers 1998); therefore, SST1 should be a good alternate predictor of wet-season onset. The forecast skill of our logistic regression model using SST1 as the predictor is tested for various lags (Table 3). The May–June and June–July SST1 forecasts perform slightly better than the SOI at the same lead time. However, the July–August SOI is more skillful than the July–August SST1 forecasts. This result is consistent with our physical understanding of the sources of predictability in the climate system, and the dynamics of ENSO, with the greatest predictive memory being in the ocean. Hence, at long lead times the best predictions are derived from using ocean information, but since there is some inherent uncertainty in the influence of ocean conditions on the atmospheric circulation, at short lead times the best predictions are derived from the atmospheric component of the oscillation (i.e., the SOI).

9. Summary

Noting the importance of early wet-season rainfall for north Australian agricultural industries, we define a wet-season onset date by simple accumulation criteria based solely on rainfall. In this paper we present results for a date of onset defined by an accumulation of 50 mm of rainfall from 1 September. The methodology has also been extended to alternative onset definitions. At most locations across north Australia, the onset dates show a marked relationship to the immediately preceding winter's (July–August) SOI, emphasizing the previously found feasibility of predicting onset with an ENSO-based predictor (e.g., Nicholls 1984). We employ a logistic regression model to forecast the probability of a later than normal onset at each location from the July–August SOI. As an extension of this forecast model, probabilities of a late onset in reference to various threshold dates can also be forecast and combined together to create "probability of exceedance" curves.

The forecasts have varying degrees of skill for different locations. The cross-validated hindcasts show most skill in the northern and easternmost parts of the study area, especially for the Top End, Cape York, and central Queensland, where they outperform climatological forecasts by more than 10%. Furthermore, despite being developed from gridded data that are representative of an area average, the forecasts generally outperform those developed from high quality station data as well, even when verified against the station observations.

Due to the limited record length (57 yr), the forecasts produced independently at each grid location contain a certain amount of spatial noise. Hence, the forecasts are smoothed in the spatial dimension, resulting in considerably greater forecast skill. Similarly, the probability of exceedance curves produced from combining forecasts for different threshold dates contain a certain degree of noise, which is removed by smoothing in the temporal (threshold date) dimension.

Predictors other than the July–August SOI were also tested. At longer leads, a pattern of Pacific SSTs (specifically SST1) performs better than the SOI, which is consistent with the knowledge of the dynamics of ENSO, and the longer-term memory provided by the ocean.

Most importantly, the forecasts should be readily useable by agricultural managers, as ascertained through a continuous consultative process with representatives of the agricultural industry, provided there are no large and sudden shifts in the relevant climate processes (as may result from climate change). Further enhancements may come from the incorporation of information from multiple predictors, for example, from Pacific thermocline temperatures and western Pacific wind variability at relatively long leads (McPhaden et al. 2006), or from the Madden–Julian oscillation at shorter leads (Wheeler and Hendon 2004; Donald et al. 2006). As well as serving to extend the earlier work of Nicholls, this study will serve as a useful benchmark to such further enhancements. Finally, work is also under way to extend these techniques to both the wet-season duration and decline.

Acknowledgments. This research was supported in part by the Managing Climate Variability Program of Land and Water Australia, through the multi-institution collaborative project QPI62. Thanks to Robert Fawcett for his efforts in producing and providing the gridded rainfall dataset, Wasyl Drosdowsky for his guidance early in the project, Sarah Lennox for her statistical advice, Sam Cleland and Tim Schatz for hosting our visits to the Northern Territory, and Scott Power for his enthusiasm and suggestion to also test SST1. The logistic regression software was obtained from Alan Miller's suite of statistical routines in Fortran (information online at http://users.bigpond.net.au/amiller/). We thank Pandora Hope and Robert Fawcett for their internal reviews, as well as two anonymous external reviewers.

REFERENCES

Ash, A., J. Corfield, and T. Ksiksi, 2002: The Ecograze Project: Developing guidelines to better manage grazing country. CSIRO Sustainable Ecosystems Rep. 4366, Townsville, Queensland, Australia, 44 pp.

Cook, G. D., and R. G. Heerdegen, 2001: Spatial variation in the duration of the rainy season in monsoonal Australia. *Int. J. Climatol.,* **21,** 1723–1732.

Donald, A., H. Meinke, B. Power, A. H. N. Maia, M. C. Wheeler, N. White, R. C. Stone, and J. Ribbe, 2006: Near-global impact of the Madden–Julian oscillation on rainfall. *Geophys. Res. Lett.,* **33,** L09704, doi:10.1029/2005GL025155.

Drosdowsky, W., 1996: Variability of the Australian summer monsoon at Darwin: 1957–1992. *J. Climate,* **9,** 85–96.

——, and L. E. Chambers, 1998: Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. BMRC Research Rep. 65, Bureau of Meteorology, Melbourne, Victoria, Australia, 39 pp. [Available online at http://www.bom.gov.au/bmrc/pubs/researchreports/rr_65/RR65.htm.]

——, and ——, 2001: Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *J. Climate,* **14,** 1677–1687.

Hammer, G. L., J. W. Hansen, J. G. Phillips, J. W. Mjelde, H. Hill, A. Love, and A. Potgieter, 2001: Advances in application of climate prediction in agriculture. *Agric. Syst.,* **70,** 515–553.

Haylock, M., and N. Nicholls, 2000: Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998. *Int. J. Climatol.,* **20,** 1533–1541.

Keenan, T. D., and R. E. Carbone, 1992: A preliminary morphology of precipitation systems in tropical northern Australia. *Quart. J. Roy. Meteor. Soc.,* **118,** 283–326.

Maia, A. H. N., H. Meinke, S. Lennox, and R. C. Stone, 2007: Inferential, nonparametric statistics to assess the quality of probabilistic forecast systems. *Mon. Wea. Rev.,* **135,** 351–362.

McBride, J. L., and N. Nicholls, 1983: Seasonal relationships between Australia rainfall and the Southern Oscillation. *Mon. Wea. Rev.,* **111,** 1998–2004.

McCown, R. L., 1981: The climatic potential for beef cattle production in tropical Australia: Part III—Variation in the commencement, cessation and duration of the green season. *Agric. Syst.,* **7,** 163–178.

McPhaden, M. J., X. Zhang, H. H. Hendon, and M. C. Wheeler, 2006: Large scale dynamics and MJO forcing of ENSO variability. *Geophys. Res. Lett.,* **33,** LI6702, doi:10.1029/2006GL026786.

Meinke, H., and R. C. Stone, 2005: Seasonal and inter-annual climate forecasting: The new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic Change,* **70,** 221–253.

Mills, G. A., G. Weymouth, D. Jones, E. E. Ebert, M. Manton, J. Lorkin, and J. Kelly, 1997: A national objective daily rainfall analysis system. BMRC Techniques Development Rep. 1,

Bureau of Meteorology, Melbourne, Victoria, Australia, 30 pp.

Mollah, W. S., and I. M. Cook, 1996: Rainfall variability and agriculture in the semi-arid tropics—The Northern Territory, Australia. *Agric. For. Meteor.,* **79,** 39–60.

Nicholls, N., 1984: A system for predicting the onset of the north Australian wet-season. *J. Climatol.,* **4,** 425–435.

——, J. L. McBride, and R. J. Ormerod, 1982: On predicting the onset of the Australian wet season at Darwin. *Mon. Wea. Rev.,* **110,** 14–17.

Priestley, C. H. B., 1962: Some lag associations in Darwin pressure and rainfall. *Aust. Meteor. Mag.,* **38,** 32–42.

Troup, A. J., 1961: Variations in upper tropospheric flow associated with the onset of the Australian summer monsoon. *Ind. J. Meteor. Geophys.,* **12,** 217–230.

Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.,* **132,** 1917–1932.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.