Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

Improving predictions of the North Australian wet season: onset and duration

CAWCR Technical Report No. 001

Fiona Lo, Matthew C. Wheeler, Sarah Lennox

June 2008

www.cawcr.gov.au

Improving predictions of the North Australian wet season: onset and duration

CAWCR Technical Report No. 001

Fiona Lo[1], Matthew C. Wheeler[1], Sarah Lennox[2]

[1]Centre for Australian Weather and Climate Research, Melbourne, Victoria, Australia

[2] Department of Environmental Protection Agency, Toowoomba, Queensland, Australia
Current address: CSIRO Division of Mathematical and Information Sciences, Queensland, Australia

ISSN: 1836-019X

National Library of Australia Cataloguing-in-Publication

Lo, Fiona, 1972-

Improving predictions of the North Australian wet season : onset and duration / Fiona Lo ; Matthew C. Wheeler ; Sarah Lennox.

ISBN: 978-1-921424-44-1 (pdf)

Series: Technical report (Centre for Australian Weather and Climate Research. Online) ; no. 1.
Notes: Bibliography. Also issued in print.

1. Monsoons--Australia, Northern--Forecasting. 2. Rain and rainfall--Australia, Northern. 3. Australia, Northern--Climate.

Other Authors/Contributors: Wheeler, Matthew C, 1972- ; Lennox, Sarah.

551.51840994

Enquiries should be addressed to:

Matthew Wheeler
Centre for Australian Weather and Climate Research:
A Partnership between the Bureau of Meteorology and CSIRO
GPO Box 1289
Melbourne VIC 3001
AUSTRALIA
m.wheeler@bom.gov.au
Phone: 61 3 9669 4068
Fax: 61 3 9669 4660

Copyright and Disclaimer

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A statistical probabilistic forecast scheme for predicting wet season onset in north Australia was developed by Lo et al. (2007) using the Southern Oscillation Index (SOI) in austral winter as the sole predictor. This current study investigates the use of different predictors with the goal of improving forecast skill, and extends the scheme to also forecast wet season duration. An empirical orthogonal function (EOF) analysis of austral winter sea surface temperature (SST) anomalies in the Indo-Pacific is used to characterize the major modes of interannual climate variability affecting rainfall in the austral spring and summer. The first three EOFs have statistically significant correlations with both onset and duration, indicating the EOFs are potentially good predictors. Brier skill scores for cross-validated hindcasts show that these predictors are more skilful than the SOI or other defined SST predictors. The predictors are applied here to provide forecasts for both onset and duration.

1. INTRODUCTION

Recent work by Lo et al. (2007; hereafter L07) resulted in the development of a statistical prediction scheme for wet season onset across north Australia. Noting the strong influence of the El Niño-Southern Oscillation (ENSO) on north Australian spring rainfall, L07 built on the previous work of Nicholls (1984) and others (e.g., Priestley 1962; Nicholls et al. 1982; McBride and Nicholls 1983), to provide, and demonstrate the applicability, of a probabilistic forecast of wet season onset for a grid of locations. In L07, the main predictor used was the Southern Oscillation index (SOI) preceding the wet season (e.g. July-August). In this paper, we extend and improve on the work covered in L07. Namely, we demonstrate increased skill resulting from the use of multiple large-scale patterns of sea surface temperatures (SSTs) as predictors, and additionally investigate forecasts of the duration of the wet season in north Australia.

As in L07, we define wet season onset to occur once a certain threshold of rain is reached beginning from 1 September. While the choice of this threshold is somewhat arbitrary and may depend on user requirements, here, as in L07, a constant threshold of 50 mm accumulation is used at all locations. Similarly, the wet season retreat may be defined to occur by working backwards in time from 30 April. The duration of the wet season is then computed as the number of days from onset to retreat. A consultative process with the grazing industry in north Australia revealed that duration was potentially of greater importance than the total wet season rainfall, hence the motivation for considering this additional parameter for prediction. In north Australia the duration of the wet season is highly related to the onset, thus it is likely that duration can be skilfully predicted using the same predictors and statistical system as onset.

In L07, an initial assessment of SSTs as predictors of onset was made. Here this is developed further by deriving SST spatial patterns in the Indian and Pacific Oceans that are specifically aimed at making forecasts from the austral winter season. We derive patterns, using Empirical Orthogonal Function (EOF) analysis, that are aimed at isolating the principal patterns of SST variability that tend to exist at the time of year at which we desire to issue forecasts (i.e., in the austral winter). In contrast, in L07 the SST pattern used was derived without any consideration of season.

After introducing our datasets (Section 2), this paper gives a comprehensive consideration of the new SST predictor patterns (Section 3). The methodology of L07 is adapted to allow for multiple predictors. We use a logistic regression model to forecast the probability of onset being later than any threshold date, or duration greater than a certain number of days (Section 4). We demonstrate that using the first three EOFs as predictors can provide skilful forecasts of both the onset and duration (Section 5), and that these predictors provide an improvement over what was previously achieved (Section 6).

2. DATA AND DEFINITIONS OF ONSET AND DURATION

The SST anomalies are derived from monthly values of the NOAA Extended Reconstructed Sea Surface Temperature, version 2 (http://www.cdc.noaa.gov/cdc/data.noaa.ersst.html) for 1948-2006. The Indo-Pacific Ocean is selected in the latitude range of 30°S- 30°N. However, only data from every second grid point of the original 2° grid is used to reduce the computation time for the EOFs. Monthly anomalies are calculated with respect to the climatological monthly means computed for 1948-2006.

Daily rainfall data is used to determine wet season onset and duration. The rainfall dataset is provided by the National Climate Centre of the Australian Bureau of Meteorology. It is a gridded analysis derived from daily station observations, area-averaged to a 1° grid (see L07 for details). The area of this study is the continent of Australia north of 28°S. Data for this study has been updated from L07 and now extends from 1 September 1948 to 31 July 2007, thus providing data for 59 wet seasons: 1948/49 – 2006/07.

As in L07, onset is defined as the date at which there is an accumulated rainfall of 50 mm beginning from 1 September. If onset is not reached by 31 March, then onset is very late and given a special value. The calculations and figures in this report focus solely on the onset definition threshold of 50 mm. However, this forecast technique may be easily adapted to alternative onset definitions and thresholds according to different agricultural disciplines and management needs.

Duration of the wet season is defined as the number of days from onset to retreat inclusively. The retreat date is defined as the date at which 50mm of rain falls between it and 30 April. Note that the retreat date cannot be calculated until after 30 April.

In the case that there is little rain during the wet season and the onset does not occur before 31 March, then the onset is considered to be too late for the wet season to be useful, and both onset and duration are given special values. Further, in the case in which onset occurs before 31 March but there is insufficient rainfall after onset to reach 50mm before 30 April (i.e. the defined retreat occurs before the onset), then the duration is similarly not able to be defined and given a special value. This occurs approximately 6% of the time when averaged across all grid boxes and all 59 wet seasons, occurring more frequently in the climatologically dry areas. Note that these years of undefined onsets and durations are still included in the development of the forecast scheme, and our forecast probabilities in each grid square accurately reflect these years with failed wet seasons.

Mean dates of onset were provided in L07. The median value for duration is shown in Fig. 1. In the very dry areas of the region, such as the western part of Western Australia and northern South Australia, the duration is too short to be defined in more than half of the years on record, resulting in a median duration that is effectively negative and undefinable. Elsewhere, the median duration ranges from greater than 140 days in the far north of the Top End and along the eastern coast, to less than 40 days in the central and western deserts.

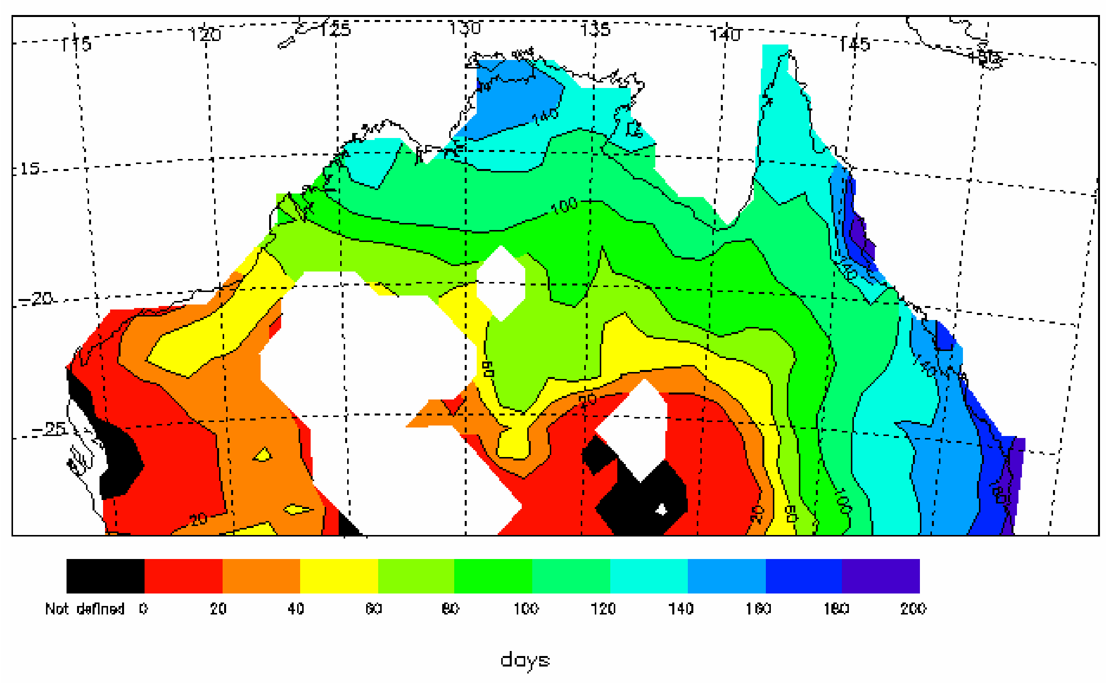

Figure 1: Median length of duration of wet season in days (1948/49-2006/07). White areas denote regions of insufficient observed rainfall data, and black areas denote where the duration was too short to be defined in more than half the years on record.

3. EOFS OF SSTS AS PREDICTORS

Utilizing EOFs as predictors is simple and efficient because a small number of EOF spatial patterns can represent a large fraction of the total variance of the predictor data, in this case, SST anomalies. Additionally, EOFs are by definition orthogonal, thus their use as predictors can minimise problems associated with highly correlated predictors. Because we are interested in making forecasts prior to the beginning of the wet season, we choose to specifically use SST anomalies from the austral winter months to generate the predictor spatial patterns. Therefore, EOFs were computed for May–August averaged SST anomalies in the Indo-Pacific (30°N-30°S) for 1948-2006, with the eigenvector calculation made from the correlation matrix. The first four EOFs (EOF1-4) are shown in Fig. 2. Their corresponding principal component (PC) time series, are presented in Fig. 3.

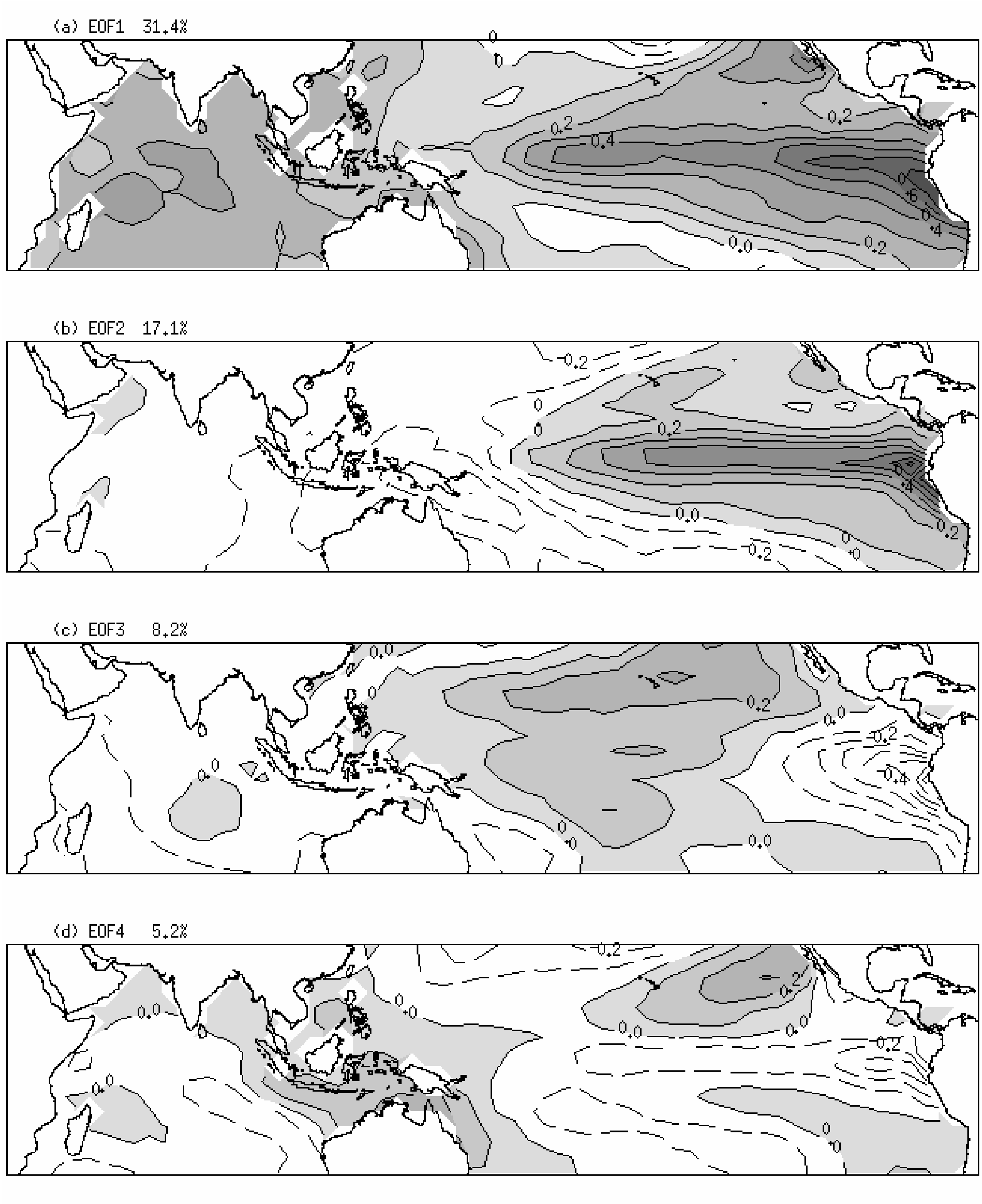

Figure 2: (a-d) First four EOFs of Indo-Pacific (30S- 30N) May-August SST anomalies, 1948-2006. Contour interval of 0.1C. Positive values are shaded. Percentage variance explained by each EOF is indicated in the top left corner.

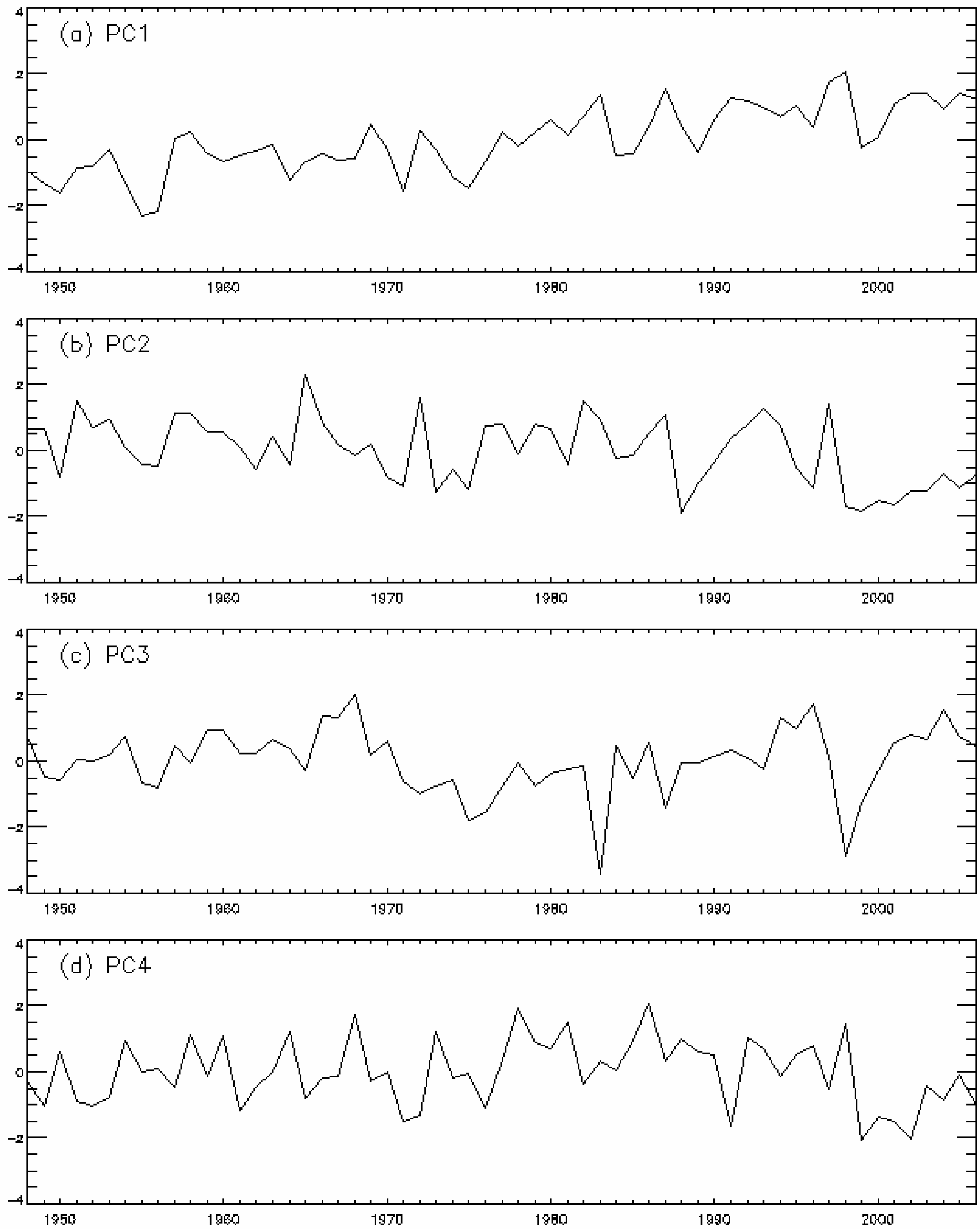

Figure 3: (a-d) First four principal components associated with EOF of May-August SST anomalies, 1948-2006. PCs are normalized (units are dimensionless).

The leading few EOF spatial patterns resemble what is known of ENSO variability. EOF2, in particular, looks like the classic ENSO pattern. EOF2 has the strongest loadings in the eastern equatorial Pacific, oppositely-signed anomalies located in a boomerang-shaped arc to their west, north, and south, and has only small loadings in the Indian Ocean. EOF1, by contrast, has positive loadings throughout the whole Indo-Pacific, but with strongest loadings in the eastern equatorial Pacific as is characteristic of ENSO. EOF3 represents some of the event-to-event variability of ENSO; when superimposed with EOF 2, it provides an east-west translation of the peak of the ENSO-associated SST anomalies in the Pacific (e.g., Wang and Hendon 2007). Additionally, EOF3 highlights the non-linear variability of ENSO between the cold and warm phase; while El Niño SST anomalies tend to be maximized near the coast of Peru, La Niña SST anomalies tend to be maximized further west in the central Pacific (Hoerling et al. 1997). Finally, EOF4 is a multi-poled pattern that we are unable to relate to any particular mode of ocean-atmosphere variability, and we include it here for reference only.

Returning our attention to EOF1 (Fig. 2a), while having some ENSO component, its predominantly like-signed loadings everywhere are indicative of the global warming of SSTs that has occurred in recent decades (Trenberth et al. 2007). This is supported by its associated PC time series, PC1 (Fig. 3a), which shows a positive trend as well as interannual variability. The interannual component is related to PC2, with PC1 leading PC2 with a one year lag correlation of -0.41. That is, peaks (dips) in PC1 tend to be followed by dips (peaks) in PC2. Thus we conclude that EOF1 contains information both about the global warming trend and ENSO.

EOF2, which most represents ENSO, is very similar to the SST1 pattern developed by Drosdowsky and Chambers (2001). SST1 is the large-scale dominant pattern of SST anomalies over the Indo-Pacific region, found by the first rotated EOF of SST anomalies using all seasons of data. The main difference between SST1 and EOF2 is that SST1 is computed using all months of the year, whereas EOF2 is computed using averaged May-August SST anomalies. SST1 was tested as a predictor in L07, and, we found that while

SST1 was a reasonable predictor of wet season onset, it was not as skilful as the SOI at leads of less than a few months. We will show in section 6 that using the projection time series of EOF2 as the sole predictor gives a more skilful forecast than either SST1 or the SOI, thus justifying our construction of spatial predictor patterns that are specific to austral winter.

How is the onset and duration of the wet season related to the EOF patterns? To answer this question, each PC is correlated with the time series of onset dates (Fig. 4) and duration lengths (Fig. 5) over the wet seasons 1948/49-2006/07 for each grid square. The statistical significance of the correlations is indicated in the left-hand column of each figure. As seen from the large amount of area that shows statistical significance at 5% or better, EOF2 and EOF3 patterns linearly correlate best with the onset dates and duration lengths. Thus, PC2 and PC3 are potentially the best two predictors.

PC1 does not have a spatially-extensive statistically-significant correlation with onset date or duration length (Fig 4a and 5a). This is consistent with the raw data as there is no clear signal of trend in the onset date or duration length over northern Australia (not shown). However, there are a couple reasons to retain PC1 as a predictor. EOF1 overwhelmingly represents the global warming trend and also contains some information on the interannual variability of ENSO. To ensure that as much ENSO information is retained as possible, we include EOF1 as a predictor. Although the trend does not factor largely with onset, it does appear to have increasing importance for the Bureau's current operational seasonal forecast scheme. Furthermore, given the large amount of variance accounted for by the EOF1, 31.4%, (Fig. 1), it would be unwise to discard it; if the logistic regression model finds its influence to be negligible at a particular location, then it will be included with only a negligible weighting.

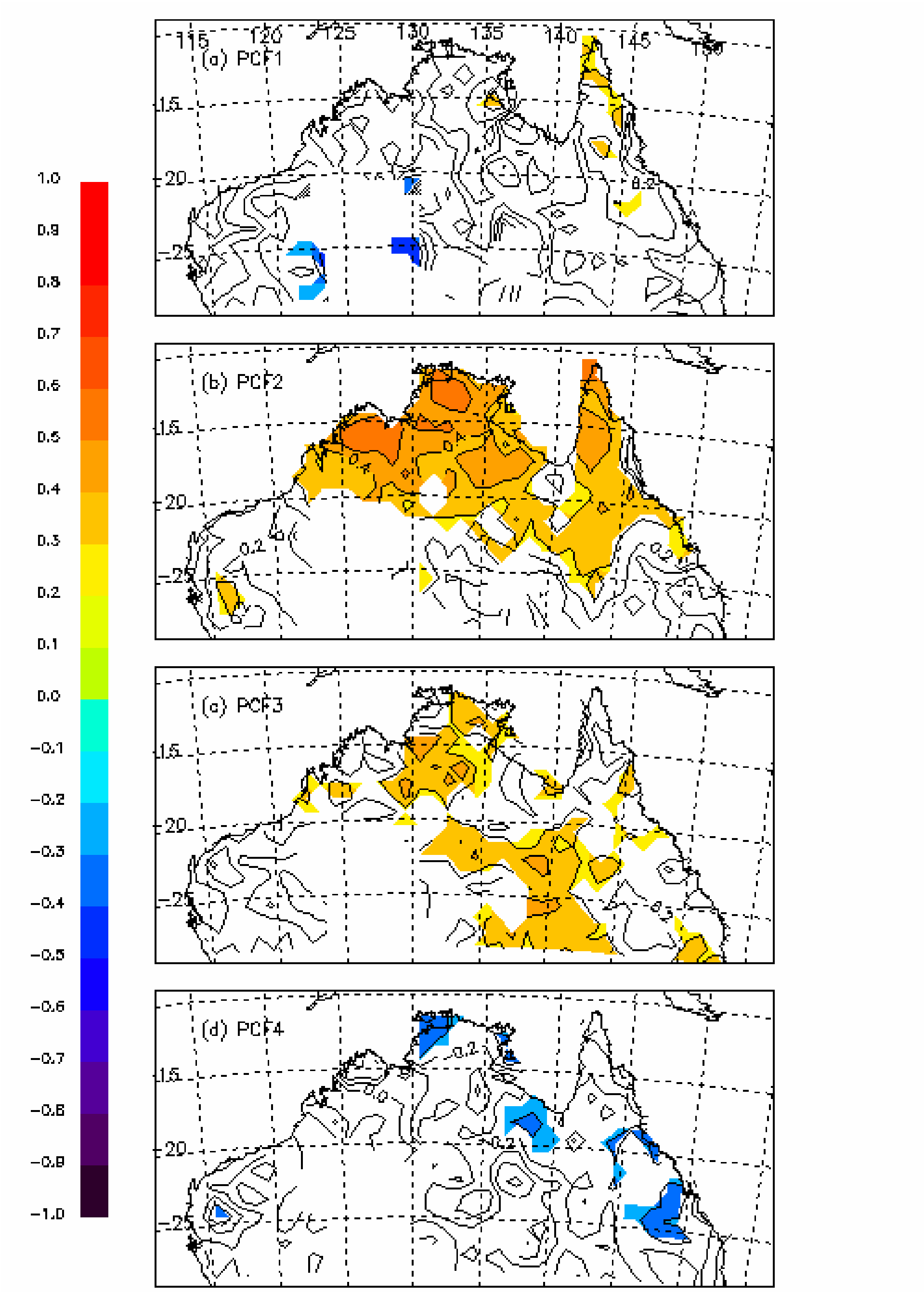

Figure 4: (a-d) Linear correlation of the leading four May-August averaged projection coefficients (PCFs; equivalent to the PCs) with the subsequent onset date in each year for the wet seasons 1948/49-2006/07. Shaded areas indicate the grid squares where the correlation coefficient is statistically significant at the 5% level.

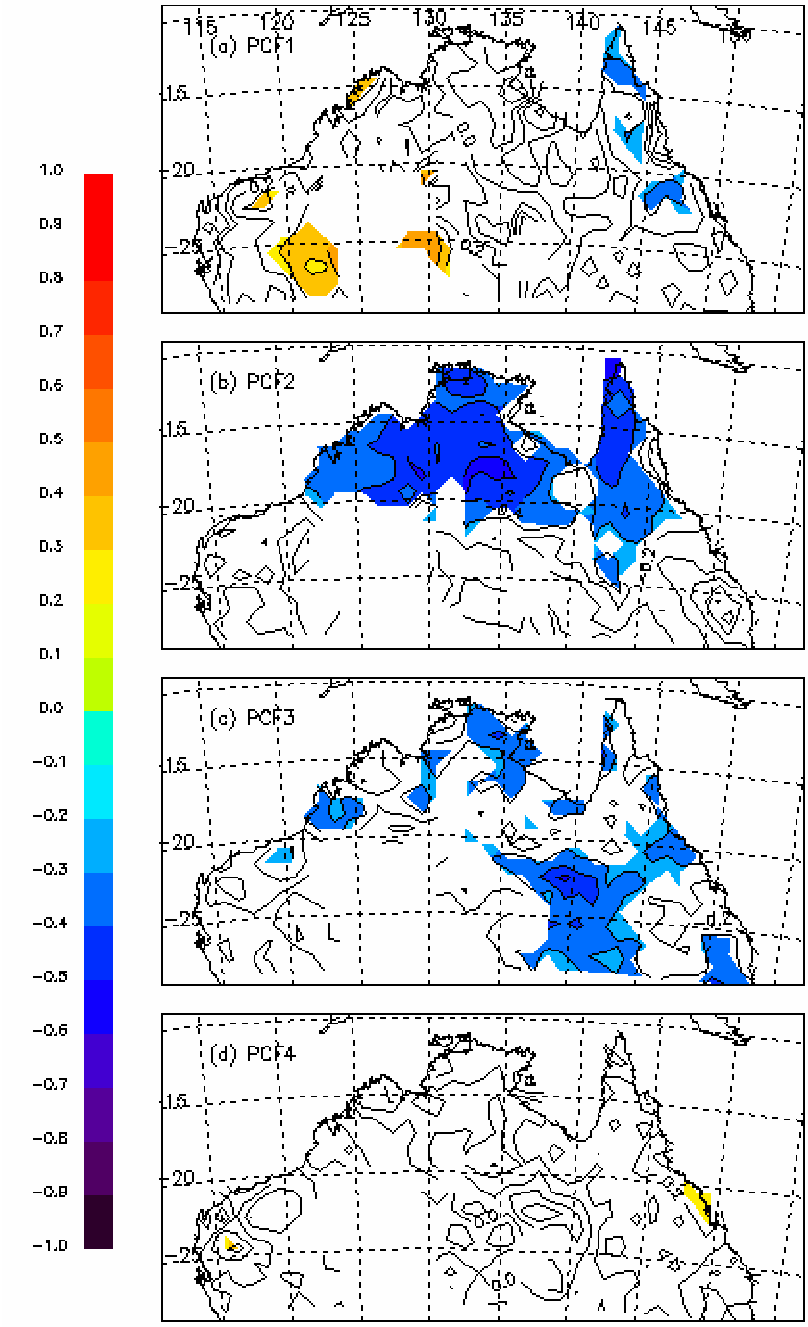

Figure 5: As in Fig.4, except for the correlation with wet season duration.

The correlation of PC4 with duration shows almost no significance at the 5% level (Fig. 5d). The correlation of PC4 with onset also shows a lack of a spatially-extensive statistical significance (Fig. 4d), but has a similar amount of statistically-significant area to the correlation of onset with PC1 (Fig. 4a). However, EOF4 accounts for only 5.2% of the total SST variance, far less than EOF1. In addition, North et al. (1982) suggest that the number of EOFs to retain can be picked from an analysis of the eigenvalues. The point at which the eigenvalues transition from a rapid decline to the more gradual decline, indicates an appropriate cutoff for the number of EOFs to retain (Fig. 6). By North et al., an estimation of standard error of the eigenvalues shows that EOF4 is not separable from EOF5, which in turn is not separable from EOF6. If PC4 were to be included as a predictor, by this rule PC5 and PC6 should be included as well, since their SST variability cannot be adequately separated from that of PC4. Since our preference is to create a relatively simple yet skilful technique, we do not include PC4, PC5 or PC6 as predictors. Limiting the number of predictors also minimizes the potential of overfitting the model.

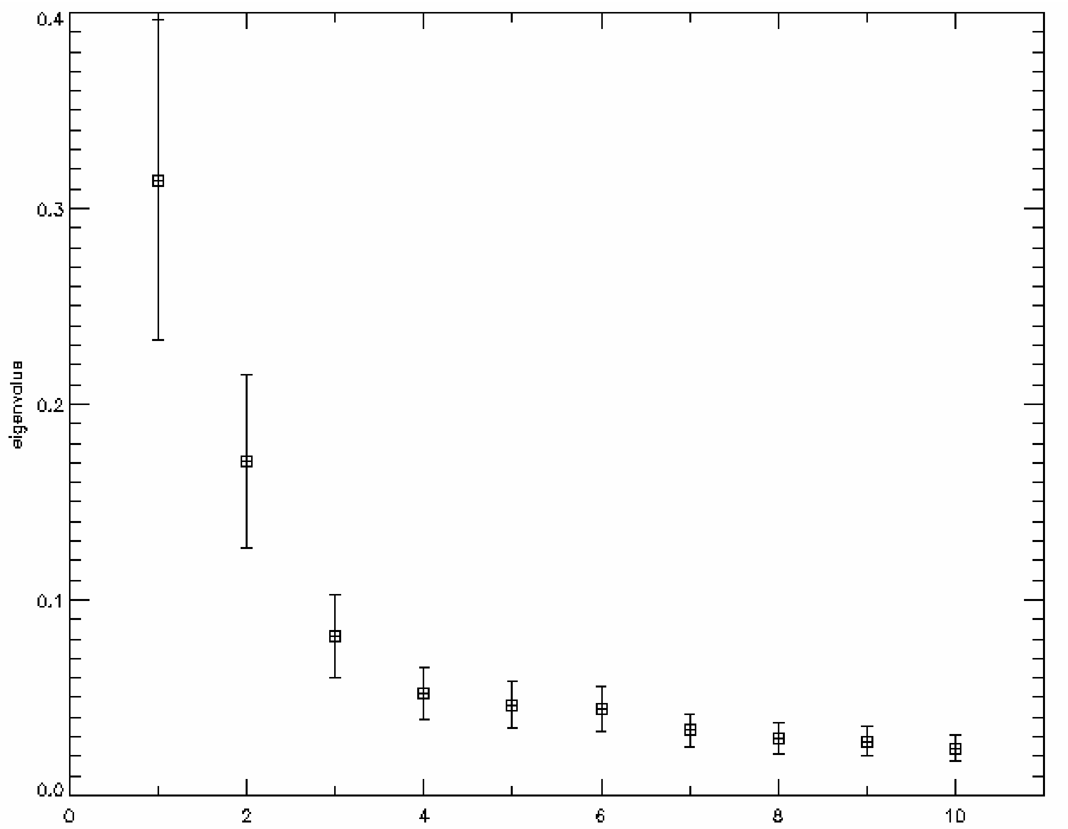

Figure 6: First ten eigenvalues of the EOF of May-August SST anomalies 1948-2006 with associated standard error bars.

Unfortunately, the use of pure PCs as predictors is not convenient because each year brings new SST data and a re-calculation of the EOFs would be required to generate the latest year's PC value. This is computationally intensive and cumbersome. Instead, an equivalent time series called the projection coefficients (PCFs) is used. Further, because we are interested in providing forecasts using SSTs from different subsets of months during the winter (e.g., May-June or July-August), the PCFs are computed by projecting monthly SST anomaly data onto the predefined EOF patterns. The EOF spatial patterns are kept constant throughout our subsequent analysis.

4. METHOD: LOGISTIC REGRESSION

Similar to L07, a logistic regression model is used for making a statistical probabilistic forecast:

$$\hat{P} = \frac{\exp(b_0 + \sum_i^m b_i x_i)}{1 + \exp(b_0 + \sum_i^m b_i x_i)}$$

where $\hat{P}$ is the predicted probability, x_i are the predictors, and m is the number of predictors. The fitted coefficients, b_i, are computed through an iteratively re-weighted least squares approach.

The logistic regression model is applied separately to each grid location. For a forecast of onset, at each grid location, each year is categorised as being late or early with respect to a threshold date. Onsets occurring later than a given threshold date are given a predictand value of 1 and onsets occurring earlier than or on the threshold date are given a value of 0. The logistic regression model is then fitted to those data. For a forecast of duration of the wet season, the duration is categorised as being longer or shorter than a threshold number of days. A wet season longer than the threshold duration is given a predictand of 1, and a wet season shorter than or of equal duration is given a value of 0.

The forecasts are generated using independent data through a cross-validation technique. Hindcasts are generated for each year at each grid square. The year to be forecast is left out of the training dataset that is used to develop the logistic regression model, thus guaranteeing the data to be independent and minimizing the generation of artificial skill.

5. RESULTS

5.1 Onset

Fig. 7 shows an example of the forecasts of onset for the 1997/98 wet season. The forecasts are given as probabilities of the onset occurring later than the mean onset date. This figure is provided as a direct comparison with Fig 4.of L07, which showed the forecast for the same year using the SOI as the only predictor. Because the forecasts are generated for each grid square independently, with no regard to adjacent grid squares, when the forecast probabilities are mapped together, there are signs of spatial noise. This is likely to be a consequence of a limited record length and noise in the input data, and unlikely to be a consequence of large scale SST changes in the Indo-Pacific. Following the procedure outlined in L07, we apply a 1-2-1 smoother once in each of the latitudinal and longitudinal dimensions.

Hindcasts were also run with respect to a range of threshold dates, from a month before to a month after the mean onset date. This provides information on the probability of onset to occur by any particular date, and is akin to the type of information contained within a "probability of exceedance" curve. Due to limited record length, these hindcasts sometimes give irregular, non-physical results; therefore, as in L07, the hindcasts were also smoothed in time with ten applications of a 1-2-1 smoother. We recognize this is not the optimum method of dealing with data inconsistencies and further improvement is required in this area. Unless stated otherwise, results shown are thus of hindcasts smoothed in space and time.

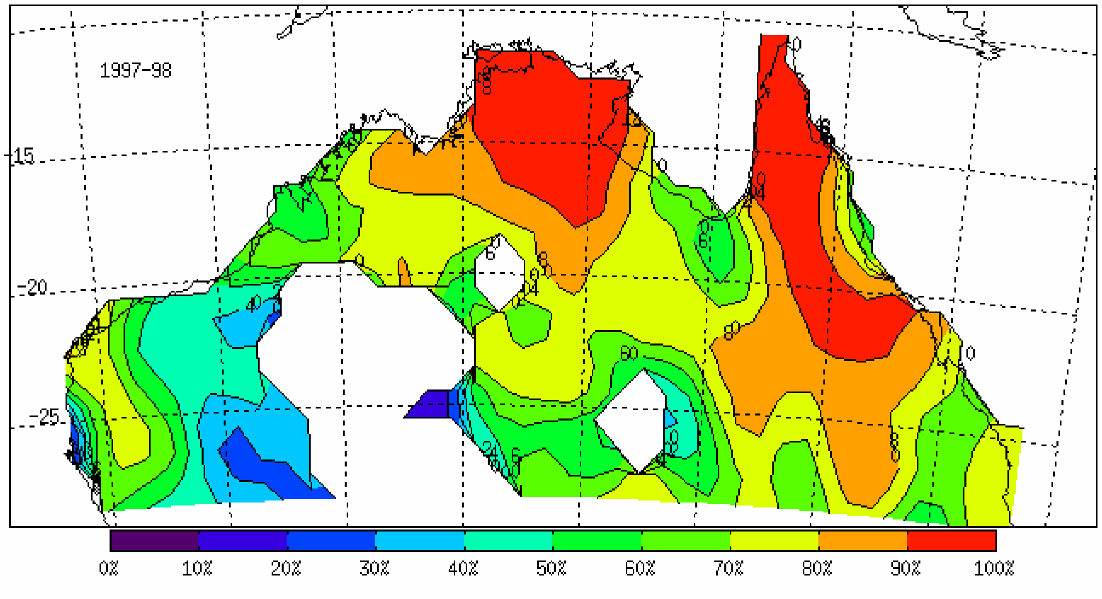

Figure 7: Example forecast probabilities (as a percentage) for onset in 1997/98, using July-August PCFs1-3 as predictors. The values express the probability that onset will occur later than the long-term mean.

The skill of hindcasts is evaluated using Brier skill scores, as an indication of improvement of forecasts compared to that of a reference forecast. Because climatology and persistence are commonly-used simple forecasts, they are used here as the reference forecasts. Fig. 8 shows the Brier skill scores computed for hindcasts, using PCF1-3 as predictors. The greatest skill occurs in the Top End and Cape York regions, where the skill is up to 40% better than climatology, and 60% better than persistence. However, the forecast performs worse than a forecast of climatology for some areas in western Western Australia and southern Queensland. For an overall indication of the forecast skill over the entire north Australia region, a Brier skill score was averaged for all valid grid boxes. A 10.04% total improvement of skill over climatology and 50.07% improvement over persistence were found.

An evaluation of the area-averaged Brier skill score with respect to climatology for forecasts using successively more PCFs as predictors, indicates that PCF2 and PCF3 add the most skill to the forecasts (Table 1). While PCF1-4 generates a slightly better skill score than PCF1-3, we feel the improvement is not great enough to justify including PCF4 in the final forecast scheme when we have no strong physical understanding of its significance for north Australian rainfall. Further, even with cross-validation, artificial skill may be generated when a larger number of predictors are allowed to be included. Table 1 also shows that a forecast with a shorter lag, using July-August predictor values, is more skilful than a longer lag using May-June predictor values.

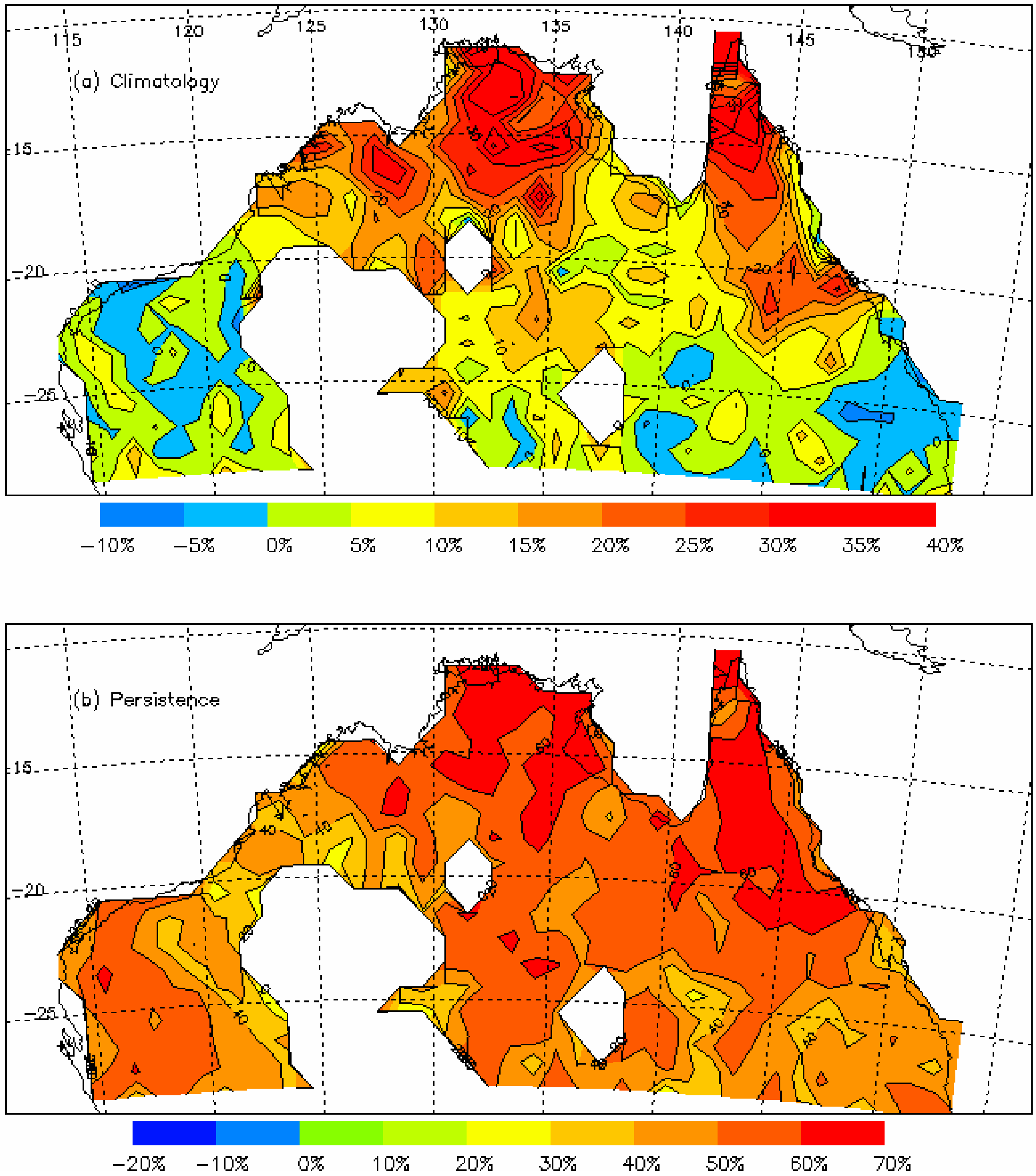

Figure 8: Brier Skill Score of onset forecast using July-August PCFs 1-3 as predictors, as a percentage improvement over (a) climatology and (b) persistence.

Table 1: Area-averaged Brier skill scores of cross-validated hindcasts of onset, compared with climatology, with one to five PCFs as predictors in (1) May-June and (2) July-August over the north Australia region.

Predictor	May-June	July-August
PCF1	0.44	0.99
PCF1-2	4.79	7.47
PCF1-3	7.58	10.46
PCF1-4	8.25	11.42
PCF1-5	7.01	9.71

5.2 Duration

Using the same reasoning and methods as for onset, hindcasts of the duration are smoothed with a 1-2-1 smoother in each the latitudinal and longitudinal direction, and smoothed with ten applications of the 1-2-1 smoother in time. Unless stated otherwise, results shown are of hindcasts smoothed in space and time; and forecast probabilities are shown with the threshold length set as the median duration length.

Forecast probabilities of the duration of the wet season were calculated using cross-validated logistic regression. Fig. 9 is an example of a forecast for the 1997/98 wet season and the corresponding observed duration, expressed as anomalies. Regions that were forecasted to have a low chance (<20%) of a longer than median duration were indeed observed to have a short duration, but elsewhere these particular forecasts are poor. The forecast map looks similar to the map of the probabilities for the onset forecast because the duration of the wet season is largely dependent on the onset. Of particular note is the region in the west, where, even though the 1997/98 season was a strong El Niño, the forecast indicates high probability that the wet season will be longer than the median duration length. This is contrary to expectation where an El Niño is usually indicative of a dry, short season. The longer wet season forecast is partly a result of the trend, governed by PCF1 and increasing SSTs in the Indo-Pacific, over the time range of this study.

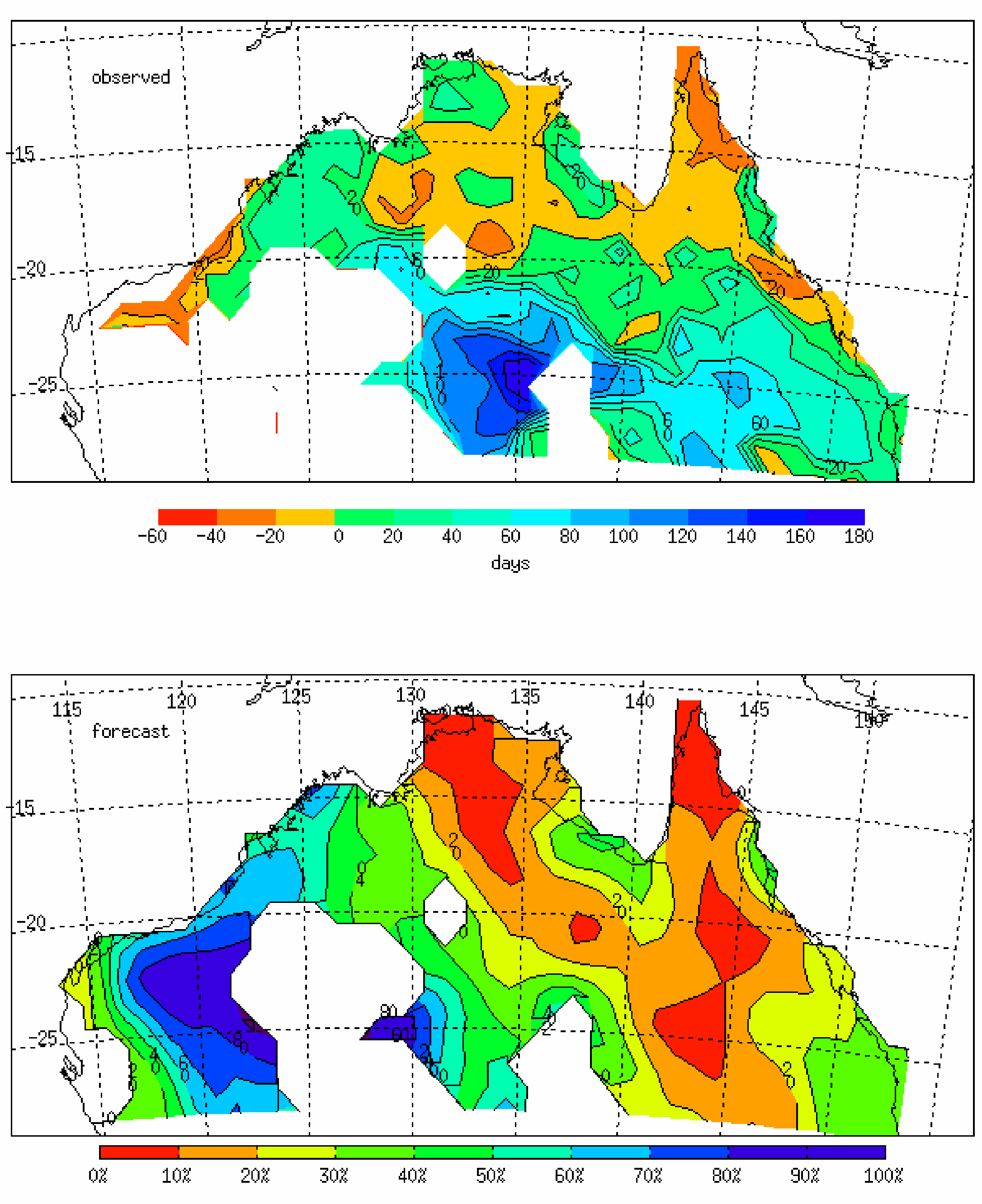

Figure 9: (a) Observed duration anomalies of wet season for 1997/98. Positive (negative) anomalies indicate the number of days the duration is longer (shorter) than the median duration length. (b) Forecast probabilities (as a percentage) for duration in 1997/98, using July-August PCFs 1-3 as predictors. The values express the probability that the duration will be longer than the long-term median.

Evidence of the trend is demonstrated by comparing the forecast probabilities of the wet seasons 1964/65 and 2005/06 (Fig. 10). These seasons share similar PCF2 and PCF3 values yet, due to the trend in PCF1 over the record, have different PCF1 values. The trend in the duration forecast splits the north Australia region into an east and west region: Western Australia and northern South Australia appear to have longer wet seasons with time, and the Northern Territory and Queensland have shorter wet seasons with time. In 1964/65 the western region shows a 10-40% probability of having a longer wet season; 41 years later, the same region has an increased probability of 60-90% of a longer than usual wet season. In contrast, the eastern region transitions from a higher probability of a long wet season (40-80%), to a drier, lower probability of a longer wet season (10-60%). Indeed, this trend in increasing rainfall is documented by the Bureau of Meteorology, and made available on the website (2007). A 30-50mm/decade increase in rainfall has been observed over the summer months in the northwest of Australia and a 30-50mm/decade decrease in rainfall in southern Queensland over the period 1950-2007. The trend is also present for onset forecasts, but less marked (not shown).

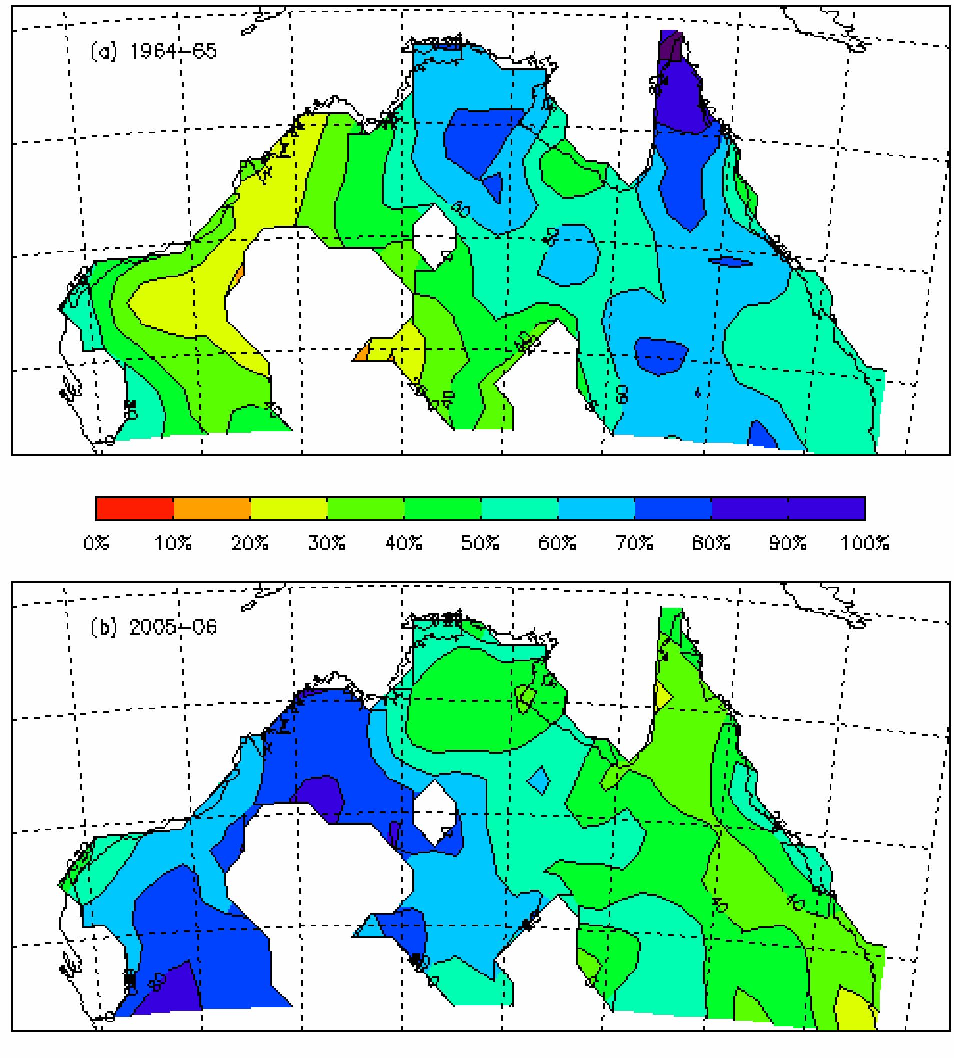

Figure 10: Forecast probabilities (as a percentage) for duration in (a) 1964/65 and (b) 2005/06, using July-August PCFs 1-3 as predictors. The values express the probability that the duration will be longer than the long-term median.

How does the probability of a long wet season change with threshold duration length? Fig. 11 shows the probability of exceedence plot for forecasts for Darwin. By definition, as the length of the duration of the wet season increases, the probability of a having a wet season longer than that threshold length decreases. The figure shows that a season with El Niño-like conditions, or positive PCF2 value, has increased probability of a shorter wet season. And, in contrast, a season with negative PCF2 values has greater probability of a longer wet season than a season with neutral PCF2.

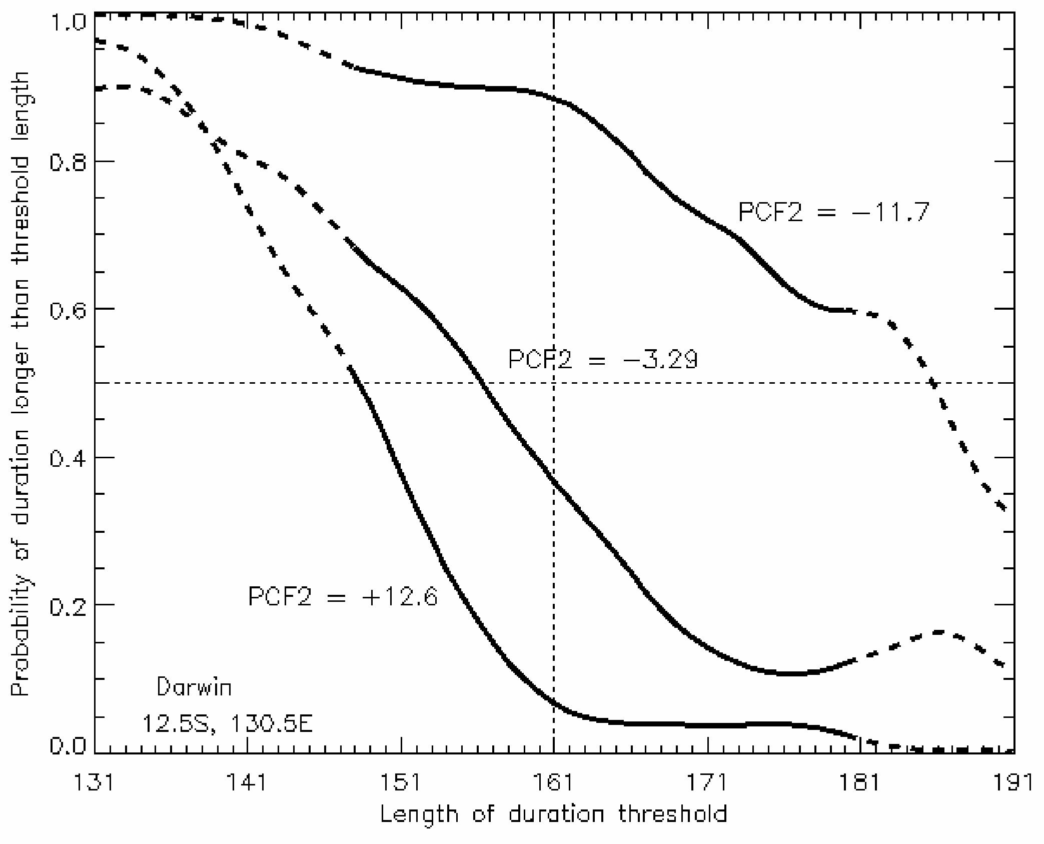

Figure 11: Probability of exceedance curves for the length of duration for the grid location nearest Darwin (12.5S, 130.5E) for positive PCF2 (El Niño) year, 1997/98; negative PCF2 (La Niña) year, 1988/89; and neutral PCF2 year, 1954/55. Dashed lines are plotted when the model is developed with fewer than 10 observations of either longer or shorter than median duration, and thus represent where there is a higher level of uncertainty in the forecast. Median duration length is 161days.

The hindcasts of wet season duration length are evaluated for skill using the Brier skill scores compared with forecasts of climatology and persistence (Fig. 12). The Top End and Cape York regions have considerable skill, with more than 30% better skill than climatology and 50% better than persistence. The area-averaged Brier skill scores for the north Australia region indicate that the duration forecasts are 10.40% better than a forecast of climatology and 48.13% better than a forecast of persistence. The forecasts of wet season duration are skilful and therefore potentially useful to the agriculture industry.

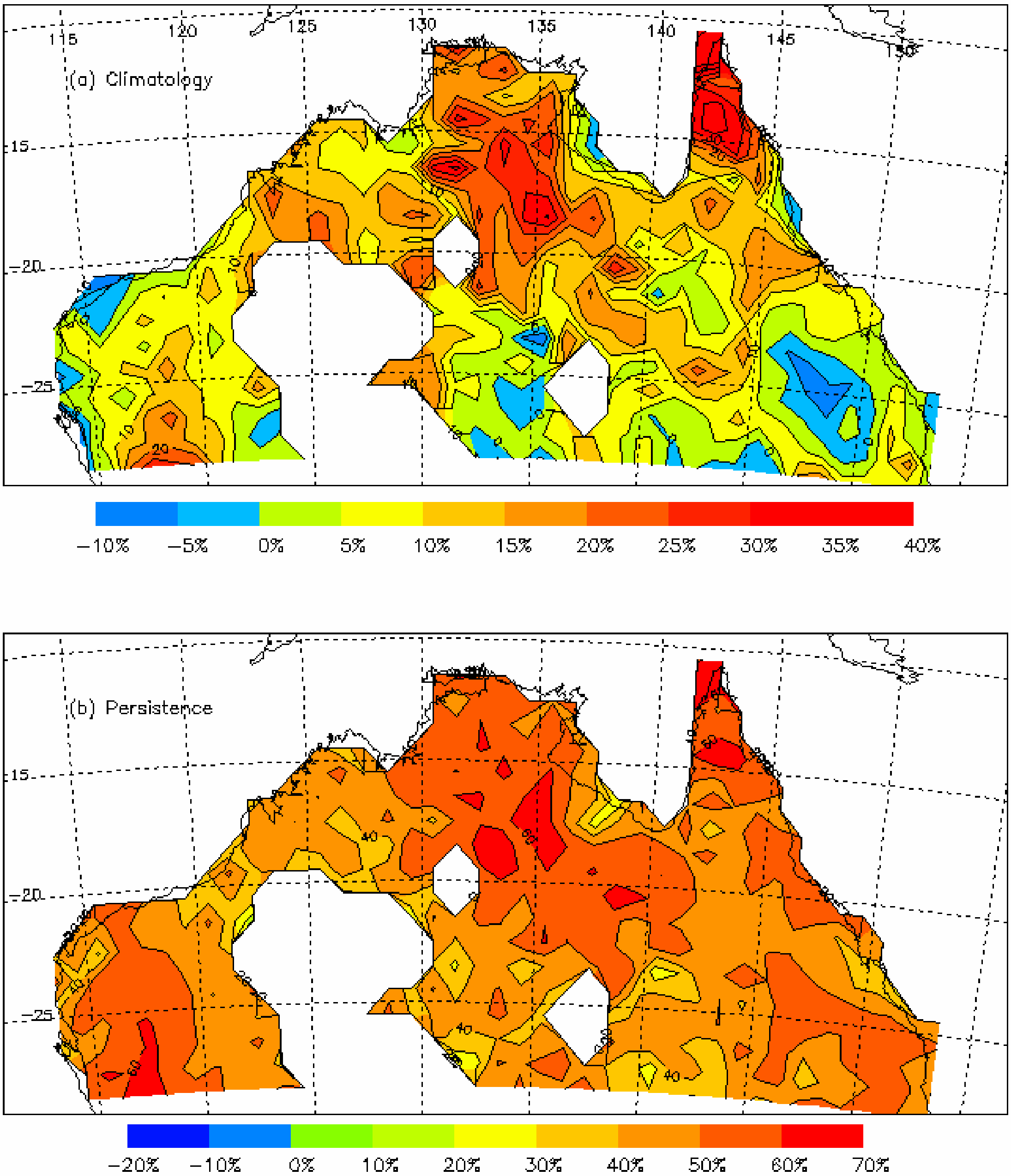

Figure 12: Brier Skill Score of duration forecast using July-August PCFs 1-3 as predictors, as a percentage improvement over (a) climatology and (b) persistence.

6. SKILL COMPARISON OF PREDICTORS WITH PREVIOUS WORK

The comparative skill of different predictors is evaluated by Brier skill scores. In L07, SOI and SST1 were evaluated as predictors and SOI was found to be a more skilful predictor than SST1 for onset forecasts issued using July-August data. As mentioned in section 3, SST1 is very similar to PCF2, but they differ in that PCF2 is calculated using EOFs with May-August SST anomalies, and SST1 is calculated using the entire year of SST anomalies. Computing the EOF with more time-appropriate data for forecasting the wet season contributes to increasing forecast skill. And, using the three dominate PCFs as predictors yield an even more skilful forecast.

Fig. 13 compares the skill of onset forecasts using maps of Brier skill scores of improvement over climatology for the predictors SOI, SST1, PCF2 and PCF1-3. Of the four predictors, PCF1-3 shows the most skill over the north Australia region and has a larger area in which the skill is better than climatology. Using PCF1-3 also results in an increase in skill in the Gulf of Carpentaria and in the southern part of the study region, where other predictors have little skill. For all predictors, the Top End and Cape York regions have the most skilful forecasts, while the western region and southern Queensland consistently show difficulty in generating a good forecast. As a summary of the findings over the north Australia region, the area-averaged Brier skill scores are shown in Table 2 for different predictors when compared with climatology and persistence.

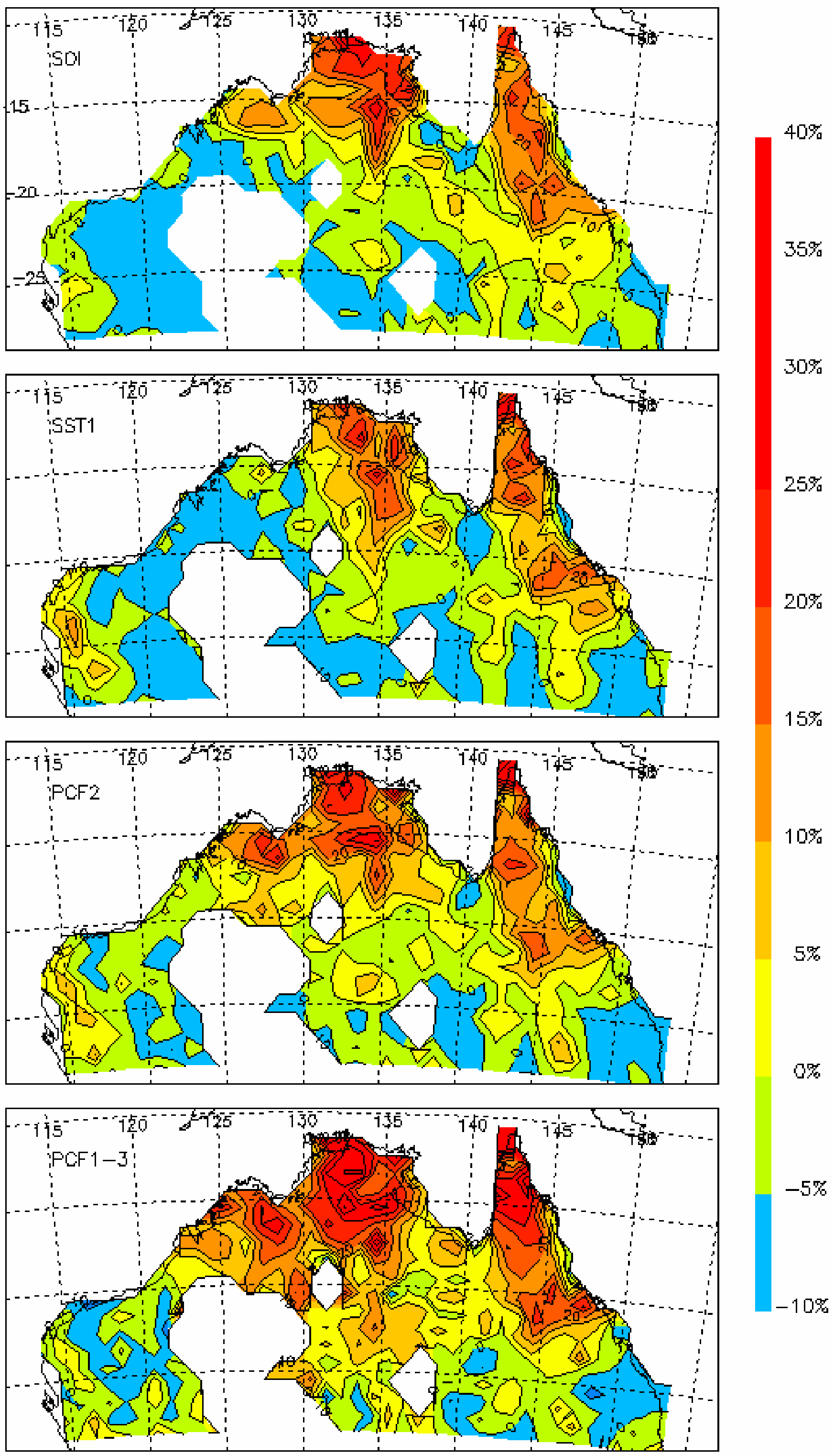

Figure 13:Comparison of Brier skill score for onset forecast, as a percentage improvement over climatology using July-August values of (a) SOI, (b) SST1, (c) PCF2, and (d) PCFs 1-3 as predictors.

Table 2: Area-averaged Brier skill scores over north Australia of hindcasts of onset with different predictors for July-August.

	Climatology	Persistence
SOI	5.62%	47.32%
SST1	4.80%	46.81%
PCF2	7.17%	48.74%
PCF1-3	10.04%	50.07%

7. SUMMARY

L07 developed a statistical prediction scheme using logistical regression for forecasting the probability of a late onset of the wet season in north Australia with the SOI as a sole predictor. This work improves upon that of L07 by using the first three seasonally-dependent EOFs of the Indo-Pacific SSTs as predictors. The grazing industry has indicated that a forecast of the duration of the wet season can also be very useful, thus this forecast technique is extended to also predict duration.

The improvement in skill of the forecasts comes from using more appropriate predictors. The predictors were selected with ENSO in mind, because previous work has shown that ENSO has a strong influence on north Australian rainfall. The essence of ENSO is captured by the first three EOFs generated from May-August SST anomalies in the Indo-Pacific. The July-August averaged PCF1-3, as associated time series to the EOF1-3, are selected as predictors for both wet season onset and duration.

Indeed, a linear relationship between the onset and PCF1-3 exists and is demonstrated by statistically significant correlations at the 5% level over considerable areas in north Australia for both PCF2 and PCF3. Similar statistically significant results are produced when repeated with correlations of duration and PCF1-3. As predictors, PCF2 and PCF3, contribute greatly to the forecast skill. PCF1 is included as a predictor, despite having only statistically significant correlations in limited areas, because of it's large explained variance (31.4%) and it's representation of ENSO variability.

PCF1-3 are input as predictors into the logistic regression model. The model is generated with cross-validated hindcasts. A forecast of the onset is presented as a probability of the onset occurring later than a specific threshold date. A forecast of the duration is presented as a probability of the wet season being longer than a threshold duration length. Due to the limited record length and noise, forecasts are smoothed in space and time following the reasoning and procedures outlined in L07.

The skill of the forecasts is evaluated using Brier skill scores in comparison to climatology and persistence. Due to the high dependence of duration on onset, the forecasts of onset and duration look very similar to each other. The Top End and Cape York areas display the most skill, where some areas are up to 40% better than climatology for onset and 60% better than climatology for duration. Previously, the Gulf of Carpentaria was an area with poor forecast skill; it had forecast skill less than climatology using predictors analyzed in L07 (SOI and SST1). However, using PCF1-3 as predictors, a skilful forecast of up to 10% improvement over climatology is now possible for Gulf of Carpentaria region.

The forecast of duration is affected by the Indo-Pacific warming trend in SST, indicated by PCF1. This trend is evidenced by two distinct areas in the study region: the west predicts increasingly longer wet season over the 59 year record, and the east has an increasingly shorter wet season.

This forecast prediction scheme was originally designed so that it could be used operationally for the agricultural industry. With this improvement in the skill of the onset forecast, and additional information of forecast of duration, these forecasts could be even more useful to the agricultural industry.

Acknowledgements. This research was supported in part by the Managing Climate Variability Program of Land and Water Australia, through the multi-institution collaborative project QPI62. Thank you to Holger Meinke and Alexis Donald for their input and comments with this project. We thank Eun-Pa Lim and Brad Murphy for their internal reviews. And, thanks also to Robert Fawcett for his efforts in producing and providing the gridded rainfall dataset.

8. REFERENCES

Australian Bureau of Meteorology cited 2007: Trend maps –Australian climate variability and change, trend in summer total rainfall 1950-2007. [Available online at http://www.bom.gov.au/cgi-bin/silo/reg/cli_chg/trendmaps.cgi]

Hoerling, M.P., A. Kumar and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate*, **10**, 1769-1786.

Lo, F., M.C. Wheeler, H. Meinke and A. Donald, 2007: Probabilistic forecasts of the onset of the North Australian wet season. *Mon. Wea. Rev*., **135**, no.10, 3506-3520.

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

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

Nicholls, N., 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.

Trenberth, K.E., P.D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J.A. Renwick, M. Rusticucci, B. Soden and P. Zhai, 2007: Observations: surface and atmospheric climate change. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, D., M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

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

Wang, G., H.H. Hendon, 2007: Sensitivity of Australian rainfall to inter-El Niño variations. *J. Climate*, **30**, 4211-4226.

The Centre for Australian Weather and Climate Research is a partnership between CSIRO and the Bureau of Meteorology.