

## NOTES AND CORRESPONDENCE

### Near-Global Sea Surface Temperature Anomalies as Predictors of Australian Seasonal Rainfall

WASYL DROSDOWSKY AND LYNDA E. CHAMBERS

*Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia*

24 February 2000 and 28 August 2000

#### ABSTRACT

An operational system for the prediction of Australian seasonal rainfall variations using sea surface temperature anomaly (SSTA) patterns over the Indian and Pacific Oceans is described. The SSTA patterns are represented by rotated principal components, with individual monthly values at 1- and 3-month lead times used as predictors; for example, November and January SSTAs are used to forecast March–May seasonal rainfall. The historical seasonal rainfall is also represented by rotated principal components of a gridded 1° rainfall dataset, with the principal component loadings used as weights to project the forecasts back to the original 1° grid points.

Forecasts of seasonal rainfall in two (above/below median) or three categories (terciles) are produced using linear discriminant analysis. Hindcast skill, measured by the linear error in probability space (LEPS) skill score has been assessed using cross validation. Experiments were also performed using a double or nested cross-validation procedure to select the best model or combination of predictors. The model chosen for operational seasonal forecasts uses the first two rotated SSTA components lagged by 1 and 3 months as predictors for every season and location, to maintain continuity of forecast probabilities between the overlapping 3-month seasons.

Current values of the principal component amplitudes are calculated by projecting either the Bureau of Meteorology's or the National Centers for Environmental Prediction's SST analysis onto the set of SST principal components. The hindcasts and experimental real-time forecasts over the 5-yr period from January–March 1994 to December–February 1998/99 indicate improved skill over parts of southern Australia during the autumn period using the SST-based schemes when compared with forecasts using the Southern Oscillation index alone.

#### 1. Introduction

The prediction of rainfall by statistical or empirical methods is feasible if there is a lagged relationship between the rainfall and suitable predictors. Such relationships exist in some seasons between northern and eastern Australian rainfall and the Southern Oscillation (SO; McBride and Nicholls 1983), and provided, until recently the sole basis for a Seasonal Climate Outlook issued by the National Climate Centre (NCC) of the Australian Bureau of Meteorology.

These relationships between Australian rainfall and the SO are weak, however, during the late southern summer and autumn period, when El Niño–Southern Oscillation (ENSO) events tend to decay. Over the western third of the continent relationships are weak throughout the year, except for the southwest corner in autumn. These limitations of the SO have led, in recent years, to attempts to find other independent predictors. Nich-

olls (1989) and Drosdowsky (1993a,b) have documented a major mode of winter rainfall variability, consisting of a broad band extending across the continent from the northwest of western Australia, through central Australia to the south coast. Nicholls (1989) showed that this pattern was related to the gradient in sea surface temperature (SST) between the central Indian Ocean and Indonesia. This anomaly pattern has been called the Indian Ocean Dipole. Drosdowsky (1993c) showed that the strongest precursor to this dipole pattern was found in the sea surface temperature anomaly (SSTA) in the eastern Indian Ocean near the west coast of Australia during the preceding summer and autumn.

Large-scale SSTA patterns have also been shown to be related to seasonal rainfall anomalies in other regions of the globe. Ward and Folland (1991) have linked rainfall anomalies in the Sahel and northeast Brazil with large-scale patterns of Pacific and Atlantic Ocean SST anomalies. Barnston (1994) used canonical correlation analysis (CCA) with near-global SST patterns as well as Northern Hemisphere 700-hPa geopotential height fields as predictors of North American and European precipitation and surface temperature.

The objective of the research reported here is to ex-

---

*Corresponding author address:* Wasyl Drosdowsky, BMRC, P.O. Box 1289K, Melbourne, 3001 Victoria, Australia.  
E-mail: W.Drosdowsky@bom.gov.au

amine the predictability, by empirical methods, of Australian seasonal rainfall anomalies, using near-global-scale SSTA patterns as predictors, and to describe the system used by NCC for operational seasonal forecasts. A more complete account is presented by Drosowsky and Chambers (1998).

The methodological aspects are discussed in section 2. The statistical model employed is discriminant analysis, since this method handles nonlinear relationships, and directly produces probability estimates of various rainfall categories such as above or below median, which is the format in which the seasonal climate outlook is issued.

The datasets used and the preprocessing are described in section 3. The variability of the rainfall and SST data is compressed into a small number of dominant modes using rotated principal component analysis. While various techniques such as CCA or singular value decomposition have been used to find coupled patterns in climate data (Bretherton et al. 1992; Barnston 1994), the single field principal component analysis is used here in preference to one of these combined techniques. This is done since it does not enforce an orthogonal predictor–predictand relationship, and it allows an easy extension to other predictors such as atmospheric circulation patterns, or predictands, such as maximum or minimum temperatures.

Results of hindcast experiments using terciles as the predicted quantity, and a small number of rainfall principal components as the predictands, are presented in section 4. Various predictors such as the Southern Oscillation index (SOI) and SSTA principal component amplitudes and combinations of these were tested on data for the 1950–93 period, with independent forecasts evaluated for the 5-yr period from January–March 1994 to December–February 1998/99. Conclusions are presented in section 5.

## 2. Methodology

### a. Discriminant analysis

Various statistical methods have been used in seasonal climate prediction; the most popular at present appears to be CCA that can be regarded as a generalized multiple linear regression (Wilks 1995). However, most seasonal climate forecasts are expressed in probabilistic terms and a forecast methodology that directly generates such forecasts was considered desirable to minimize subjective assessment and assignment of forecast probabilities. The probability forecasts in this study are produced using linear discriminant analysis that can be regarded as an application of Bayes theorem to invert conditional probabilities (Huberty 1994; Wilks 1995). The procedure used here is similar to that described by Ward and Folland (1991), and He and Barnston (1996).

Given a dataset consisting of observations of  $N$  potential predictor variables  $X = (X_1, X_2, \dots, X_N)$ , such

as the SOI or the SSTA principal components, and a classification variable  $t$ , which in this case will be the tercile ranking of the rainfall in the subsequent season, discriminant analysis calculates the conditional probability,  $P(t|X')$ , of the occurrence of class  $t$  given a new observation  $X'$ , for each of the three possible classes, that is, above normal or the top one-third, near normal, and below normal of the seasonal rainfall.

### b. Linear error in probability space (LEPS) skill scores

The skill of the forecast system is expressed in terms of the LEPS score that is described in detail by Potts et al. (1996). The score is derived from a general form  $S = 1 - |P_f - P_v|$ , where  $P_f$  and  $P_v$  are the cumulative probability of the forecast (or hindcast) and the verifying observation, so that the score measures the absolute error in terms of the cumulative probability of the forecast and observations. The score is normalized so that random forecasts score zero and perfect forecasts at the extremes of the distribution score higher than perfect forecasts in the middle of the distribution. The scores are also scaled so that they decrease uniformly with increasing separation between the forecast and verifying observation. The rationale for this normalization and scaling is discussed in detail by Potts et al. (1996). The extension of LEPS scores to probability forecasts is described by Ward and Folland (1991) and Potts et al. (1996).

### c. Cross-validation estimates of model skill

For a given model, we need to estimate the model parameters, and the model's expected skill on independent data. If all the available data is used to fit the model parameters, an independent estimate of the skill cannot be obtained. The available data must therefore be split into a dependent set used for model development and an independent or validation set. The traditional approach to this problem, generally referred to in the discriminant analysis literature as the "Holdout Method" (Huberty 1994), has been to split the dataset of length  $N$  into two parts; a training set of (about)  $2N/3$  to develop the model and estimate the parameters, with the remaining data used as a verification set on which independent hindcasts are performed and assessed.

In the "cross-validation" technique the skill of a forecast model is estimated from a series of independent hindcasts over all the available data. This is achieved by  $N$  repeated applications of the traditional approach, with a training set of  $N - 1$  data points and one observation in the verification set. In each application the model parameters are recalculated and a hindcast produced for the omitted observation. If the observations are strongly autocorrelated then blocks of two or more observations may need to be omitted to maintain the

independence of the hindcast and development data subsets.

In both cases it is desirable to derive the final forecast model from the full dataset, to make full use of the available data. The estimates of skill are then applicable to the particular type of model used, rather than the specific set of parameters of that model. Since the cross-validation method produces  $N$  hindcasts, this method should give a more robust estimate of the true hindcast skill than the traditional methodology for which there are typically only  $N/3$  hindcasts.

#### *d. Model selection using cross validation*

In the development of most empirical forecast systems there are generally a large number of potential predictors identified that can be used in the forecast model. Various stepwise procedures have been described in the statistical literature and implemented in statistical computer packages for selecting the optimal predictors, especially in the case of multiple linear regression. Since the fit of a regression model to the dependent data is always improved by adding more predictors (see Wilks 1995), the standard selection criteria for assessment of model fit include a penalty term involving the number of predictors used. Similar procedures can be applied to predictor selection in linear discriminant analysis, although the fit of the discriminant analysis model is not always improved by adding more predictors, unlike the case of multiple linear regression.

These stepwise selection procedures, however, do not necessarily guarantee that the best possible combination or subset from a given number of potential predictors will be found. The alternative strategy is to test all possible subsets. With two potential predictors there are four possible models; either predictor could be used alone or both used together. The fourth model would be a null forecast such as climatology, involving neither predictor. In general, with  $N$  potential predictors, and including climatology as a possible null predictor, there are  $2^N$  subsets of predictors. To obtain independent estimates of the skill of the forecast model, the process of selection or screening of potential predictors also needs to be validated. This leads to the concept of a double or nested cross-validation procedure, such as that described by Stone (1974), Michaelson (1987), and Elsner and Schmertmann (1994).

The final stage of this process is the selection of the best model using all the available data. The cross-validation estimate of the hindcast skill of this model is "in sample" since all the data have been used to select the model. The difference between this estimate and that obtained through the double cross validation gives an estimate of the artificial skill.

#### *e. Significance levels for LEPS scores*

Potts et al. (1996) give no estimate of significance levels for LEPS skill scores. For the simple case of

estimating the skill of a specified model described in section 2c above, we can make an assessment of significance levels by a Monte Carlo method. Random forecasts are generated by replacing the predictors with a sequence of random numbers of the same length and drawn from a similar distribution. As with any statistical model, the hindcast skill can be increased by increasing the number of predictors, so that the experiments were performed with different numbers of predictors. From the distribution of the resulting LEPS scores we estimate the upper 5th percentile as 3.8, 7.3, 8.6, and 10.9 for one to four predictors, respectively.

The procedure can also be used to estimate the significant threshold values to apply to the selection of the best predictor or combination of predictors from a large pool of potential predictors. In this case the threshold for acceptance increases as the number of potential predictors increases.

### **3. Data and principal component analysis**

Seasonal predictions are inherently uncertain and best expressed in probabilistic terms. The aim of this study therefore is to produce forecasts of the probability of (three monthly) seasonal rainfall being in one of three categories: above normal, near normal, and below normal at a fairly fine resolution (e.g.,  $1^\circ$  grid points) across the entire Australian continent. Producing a separate forecast at each individual grid point results in a very patchy or noisy spatial structure due to the large noise variance in the monthly or seasonal rainfall data at this scale. In most operational seasonal forecasts some form of data averaging is employed, either large-scale area averages such as for the Sahel or northeast Brazil, or data reduction using EOF techniques; as in the National Centers for Environmental Prediction (NCEP) seasonal forecasts.

The rainfall and SSTA variance has been represented by a few dominant modes using Varimax rotated S-mode principal component analysis. For both the rainfall and SST datasets standardized monthly anomalies were calculated at each grid point by removing the full dataset monthly mean and dividing by the monthly standard deviation. The resulting time series have zero mean and approximately unit variance and, hence, the correlation and covariance matrices are virtually identical.

The relationships between Australian rainfall and some simple predictors, for example, the SOI (McBride and Nicholls 1983; Drosowsky and Williams 1991) or Indian Ocean SST (Nicholls 1989; Drosowsky 1993c), are known to vary with season and location. A forecast scheme based on SSTA patterns would be expected to exhibit similar seasonal variation, so that a seasonal stratification of the datasets is necessary. Tests were performed using monthly and 2- or 3-month, nonoverlapping, seasonally averaged data, for both the rainfall and SST data. In both cases the spatial patterns (principal component loadings) produced by the monthly and

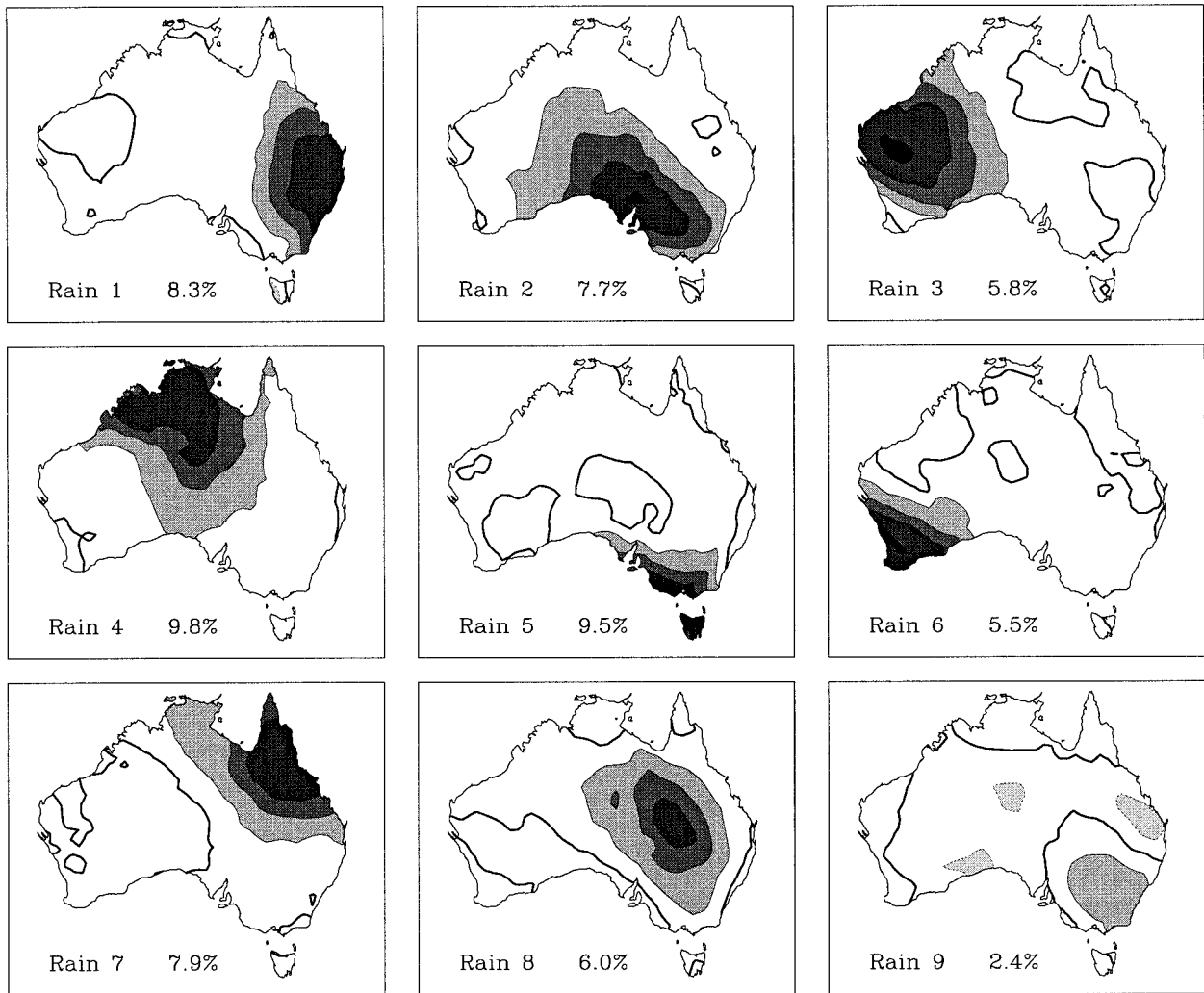


FIG. 1. Spatial pattern of loadings of the first nine gridded Australian rainfall VARIMAX rotated principal components of the standardized monthly anomalies of the dataset. Contour interval is 0.2, with zero contour heavy, negative contours dashed, and areas above +0.2 and below -0.2 shaded.

seasonal datasets were virtually identical, as were the resulting principal component time series when the monthly series were averaged into the appropriate seasons. In addition, although forecasts are currently issued for a 3-month season, the feasibility 1- and 2-month forecasts should also be examined. The principal component analysis is therefore performed on the monthly data, and seasonal values of the principal component score series are produced by averaging these monthly time series.

The application of the cross-validation technique requires the complete removal of data for the period being hindcast. Experiments with a small portion (up to 1 yr) of data removed showed that the principal component spatial patterns are very stable and the projection of the missing data is very strongly correlated with the full dataset analysis. The PC time series are therefore taken

as the primary dataset, and the analysis is not repeated during the cross-validation procedure.

#### a. Rainfall

The rainfall dataset, with monthly temporal resolution, was created by the National Climate Centre (Jones and Weymouth 1997) by analyzing all the available quality-controlled station data onto a  $\frac{1}{4}^\circ$  grid using a successive correction scheme. A reduced resolution subset (at  $1^\circ$  spatial resolution) has been used in this study.

The PC analysis is similar to that described by Drosowsky (1993a) for seasonal district rainfall anomalies for the period 1950–87. When the first nine components are rotated using the Varimax procedure (Fig. 1), the eight regions (S1–S8) found in Drosowsky (1993a) are broadly reproduced, although the relative importance is

altered. The additional component (Rain8 in Fig. 1) in the gridded dataset covers an area over the interior of eastern Australia, reflecting the change in data density; there is relatively more variance in the inland area since there are more  $1^\circ$  grid squares than rainfall districts in this region. Overall these nine components account for approximately 60% of the variance of the Australian monthly rainfall anomalies.

These rotated components form a strongly regionalized coverage of the Australian continent, and have been normalized so that the loadings represent correlations between the principal components scores and the original standardized monthly anomalies at each grid point. The squared loadings, therefore, represent the percentage of variance each component accounts for at each grid point, and can be used as weights to interpolate forecasts of the component scores back to standardized rainfall anomalies at the original  $1^\circ$  resolution grid points. The forecast system therefore needs to produce 9 individual forecasts for each of the 12 overlapping seasons. The model skill, however, is assessed against the actual rainfall at each grid point, after the forecast probabilities have been interpolated back to the grid points.

#### b. Sea surface temperature

The sea surface temperature dataset used is the U.K. Met Office Global Ice and Sea Surface Temperature dataset (GISST, version 1.1; Parker et al. 1995), from January 1949 to December 1991. Earlier work by Nicholls (1989) had suggested that there were no significant relationships between Australian rainfall and Atlantic Ocean SST. Therefore a subset covering the region from  $60^\circ\text{N}$  to  $55^\circ\text{S}$  and  $30^\circ\text{E}$  to  $70^\circ\text{W}$ , that is the Indian and Pacific Oceans only, was used and the spatial resolution of the data was reduced from a  $1^\circ \times 1^\circ$  to a  $5^\circ \times 5^\circ$  grid.

The first two unrotated principal components extracted from this data are shown in Fig. 2. The first component, which accounts for 14.9% of the variance is clearly related to ENSO. Its time series is strongly correlated with the SOI ( $-0.67$ ) and the Niño-3 index ( $+0.81$ ). The second component shows long-term trends over most of the domain, with warming in the Indian Ocean and western Pacific and cooling in the central and eastern Pacific. The remaining components (not shown) are difficult to interpret due to the orthogonality constraint on the principal components. This results in complex spatial patterns that will be difficult to relate to the rainfall in a physical manner. Hence rotation of the SSTA principal components was also performed to localize or regionalize the variance of each of the components.

Spatial patterns (the principal component loadings) and time series (the principal component scores) of the first two (SST1 and SST2) of 12 rotated components are shown in Fig. 3. SST1 is similar to the first unrotated

component, although the percentage of variance explained is lower (11.5%). It still represents the mature phase of an ENSO event, with largest (positive) anomalies in the central and eastern equatorial Pacific and in the Indian Ocean, and weaker negative loadings in the Pacific Ocean midlatitudes. As with the unrotated component, the time series is strongly correlated with the Niño-3 index and the SOI.

The second component, SST2, is located just west of the Australian continent and extends northwest to the central equatorial region. The time series of this component is strongly correlated ( $r = +0.77$ ) with the Indian Ocean index devised by Drosowsky (1993c), and shown to be a precursor to the dipole pattern documented by Nicholls (1989) and strongly related to Australian early winter rainfall.

The Indonesia–Indian Ocean dipole pattern itself does not appear in the all-months, combined Pacific and Indian ocean analysis in either the original or rotated components. The dipole accounts for a small fraction of the total variance in the SSTA, being primarily a wintertime phenomenon and occupying a relatively small spatial region compared to the size of the analysis domain, while the third and higher unrotated components have many maxima, due to the orthogonality and maximum variance constraints. Most rotation criteria, including Varimax, tend to produce components with single localized areas of high loadings rather than dipoles, with the poles of the dipoles appearing in different rotated components.

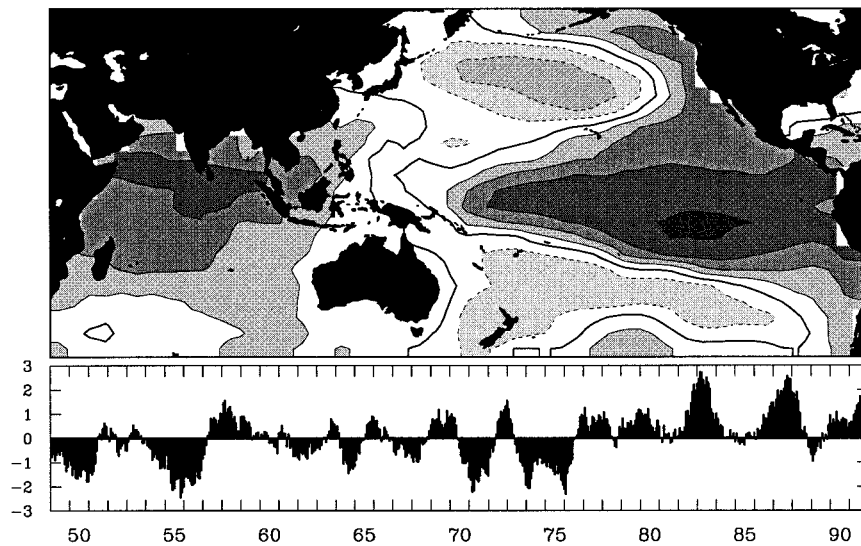
To extend the principal component time series from December 1991 to the present, two operational analyses are used. The first of these, the NCEP optimum interpolation analysis is available from 1982 to the present (Reynolds and Smith 1994), while a locally produced analysis (Bureau of Meteorology, Melbourne, Australia) is available from June 1993 (Smith 1995). Both these datasets have the same  $1^\circ \times 1^\circ$  spatial resolution as the original GISST data. The operational analyses are therefore also reduced to the  $5^\circ \times 5^\circ$  grid and projected onto the principal components using the scoring coefficients matrix.

## 4. Results

### a. SOI and Indian Ocean index

The LEPS score measures the skill of forecasts relative to a simple standard procedure, taken to be climatology or forecasts of equal probability for each of the three terciles. Another useful benchmark, and one that should be exceeded by any prospective operational scheme, is the skill possible with the original predictor used by NCC, that is, the seasonal SOI with 0-month lag. Cross-validated hindcasts were performed and assessed for the 1950–93 period using the previous seasons SOI (i.e. zero lag) as the predictor. The LEPS skill score exceeds the 5% significance level of 3.8 over parts

## PC 1 14.9%



## PC 2 6.1%

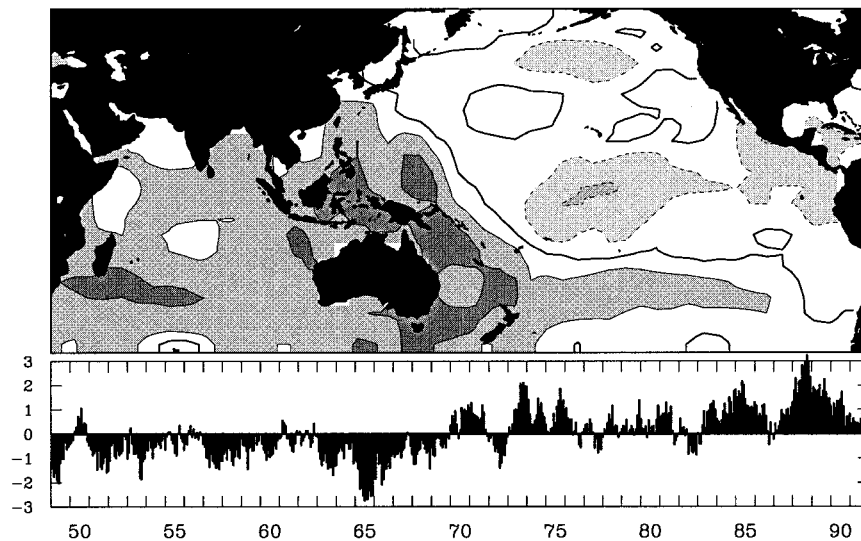


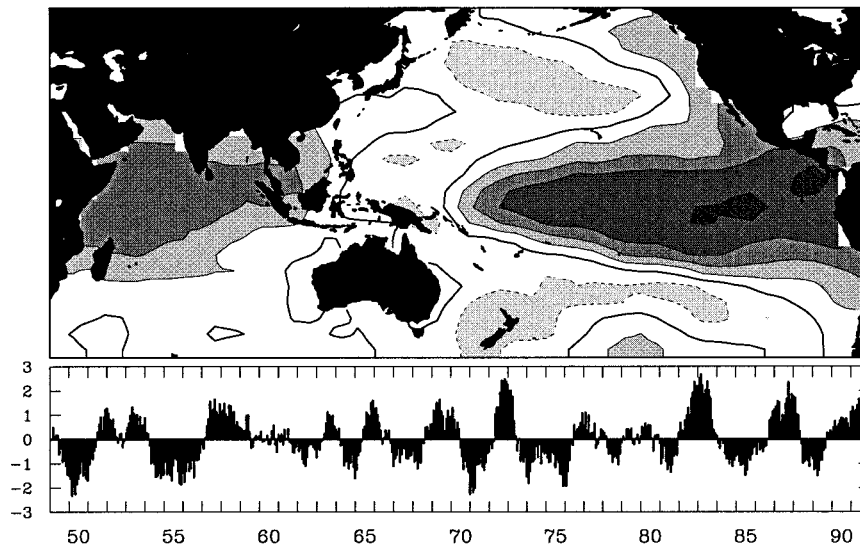
FIG. 2. Spatial pattern of loadings and associated scores (time series) of the first two unrotated principal components of the standardized monthly anomalies of the GISST dataset. Contour interval is 0.2, with zero contour heavy, negative contours dashed, and areas above +0.2 and below -0.2 shaded.

of northern and eastern Australia from winter (June–July–August) through to late summer (February–March–April). This spatial and temporal distribution of skill is very similar to the well-known seasonal rainfall–SOI correlation pattern (McBride and Nicholls 1983; Drosowsky and Williams 1991). An improved operational procedure would be to increase the forecast lead time and use the SOI lagged by one month. In this case the skill (not shown) is slightly lower overall, but with

a similar spatial coverage to the no-lag case. Using two seasonal values of the SOI, lagged at one and two (non-overlapping) seasons, leads to slightly greater skill, however, the 5% significance level is also somewhat higher (7.3) in the case of two predictors, so that the significant areas are similar.

An additional independent influence on Australian rainfall is the Indian Ocean Dipole (Nicholls 1989). A precursor signal to this dipole pattern, called the Indian

## SST 1 11.5%



## SST 2 4.3%

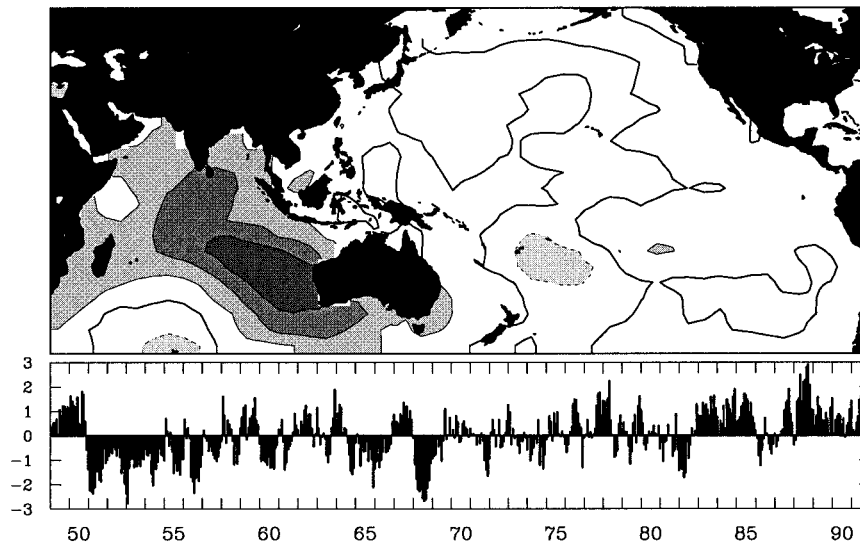


FIG. 3. Spatial pattern of loadings and associated scores (time series) of the first 2 of 12 VARIMAX rotated principal components of the standardized monthly anomalies of the GISST dataset. Contour interval is 0.2, with zero contour heavy, negative contours dashed, and areas above +0.2 and below -0.2 shaded.

Ocean index, was identified by Drosowsky (1993c). This is calculated as the SSTA averaged over the area  $22^{\circ}$ – $34^{\circ}$ S and  $90^{\circ}$ – $110^{\circ}$ E. Bimonthly values of this index in mid- to late summer were found to be correlated with early winter (May–June and June–July) rainfall over parts of southern and eastern Australia. As with the SOI, the pattern of cross-validated hindcast LEPS

skill scores (not shown) obtained by using bimonthly values of this index at 1- and 3-month lags are consistent with these earlier results. A farther small area of significant positive LEPS scores is found over the southwest of the continent during spring and early summer, a period not examined by Nicholls (1989) or Drosowsky (1993b,c).

### *b. First two rotated SST components*

The dominant mode of SST variability over the Indian–Pacific Ocean domain, SST1, corresponds to the El Niño and its time series is strongly correlated with the SOI. Its skill as a predictor of seasonal rainfall with 1-month lag (not shown) is comparable to that of the SOI with no lag, and superior to the SOI with 1-month lag. The major differences are improved skill over northern Australia, and slightly reduced skill over southeast Australia during spring, particularly September–October–November. The region with largest loadings on the second mode of SST variability, SST2, coincides to a large extent with the area used to calculate the Indian Ocean index. The seasonal and spatial variation of skill of SST2 (not shown) is similar to that of the Indian Ocean index.

Since SST1 and SST2 are only weakly correlated, using both at two lags (i.e., four predictors) results in greater overall hindcast skill (Fig. 4) than either predictor alone. This skill could possibly be increased further by using the best combination of the four predictors. This can be found by testing, on all the data, each of the 16 combinations (including climatology) possible, for each rainfall component and season, and selecting that which shows the highest skill. Applying the cross-validation procedure to this model results in an estimate of the in-sample skill, since the same data are used to find the best model and estimate the skill.

While this best combination of predictors determined from all the data should be used for forecasts on future independent data, the true skill of this procedure needs to be evaluated using the nested cross-validation technique described in section 2d, so that the selection procedure itself is validated. The spatial and seasonal distribution of skill is similar to the in-sample estimate, but at a slightly lower level overall, and suggests that the most skillful predictor combinations are reasonably robust, that is, they are selected most times with 1 yr left out. The major difference between these and Fig. 4 is a reduction in the size of areas of negative LEPS scores (not shown), and especially the magnitude of the negative scores compared with the all predictor case. This is achieved through the use of climatology as a predictor in regions where no combination of the four available predictors shows skill, since climatology should have a LEPS skill score of zero. However, the regions of positive and significant skill are largely unchanged.

### *c. Best subset of up to two predictors using full set of SST predictors at two lags*

We can extend the analysis of the previous section by including some or all of the other rotated SST principal components as predictors. As noted earlier, adding more predictors in a discriminant analysis procedure does not necessarily improve the accuracy of the fore-

casts, and may degrade the forecast skill in situations where some of the predictors have little skill. Using 10 of the SST components with two lagged values increases the number of potential predictors to 20, and the number of possible combinations to  $2^{20}$ , which is clearly too many to consider. Restricting the models to the best subset of up to two predictors, there are still over 200 possible combinations of 20 predictors. With this many models, there is now a marked difference between the in-sample dependent data skill and the “out-of-sample” independent data hindcast skill levels than was the case for just four predictors. The in-sample validation shows very high levels of skill for every season over almost the entire continent. By increasing the total pool of potential predictors (e.g., by examining other parameters besides SSTA and allowing more lagged values), and the number of predictors used for each rainfall component, the in-sample dependent data skill could in theory be raised to (almost) any desired level. This estimate is extremely optimistic, being inflated by the large number of possible predictor combinations tested. In contrast the spatial pattern of the out-of-sample hindcast skill is not significantly different from Fig. 4, and while the magnitude of the areas of positive skill does increase, the significance level remains similar due to the greater number of models and predictors examined.

### *d. Independent forecasts*

Since no data after the December–January–February 1993/94 season have been used in the development of the forecast schemes described above, predictions for seasons after this time can be regarded as true forecasts. Three of the models developed above are tested over a 5-yr period from 1994 to 1998; the SOI with two seasons as predictors (SOI2); the first two SST components at 1- and 3-months lag (SST4), and the two best from all the predictors lagged by 1- and 3-months (SST2). Seasonally stratified LEPS scores for each of these sets of forecasts are extremely noisy and do not reveal any clear picture of the relative accuracy of each scheme due to the small sample of only 5 yr in each of the seasonally stratified estimates. Aggregated LEPS scores over the 60 seasons from January–February–March 1994 to December–January–February 1998/99 are shown in Fig. 5. All three schemes show some areas of skill with LEPS scores greater than 2.5 over large parts of eastern Australia, and the far southwest corner of western Australia. The regional differences between the three schemes are most likely not significant due to sampling variations. Taking 3- or 4-yr subsets of these results in changes in detail in all three patterns. The reasons for this can be seen in Fig. 6 that shows the skill of each seasonal forecast, aggregated over all grid points in southeast Australia (south of 29°S and east of 141°E). Again all three schemes show a similar behavior, although there are periods when one scheme appears to be superior to the others. From early 1994 through to mid-1995 the



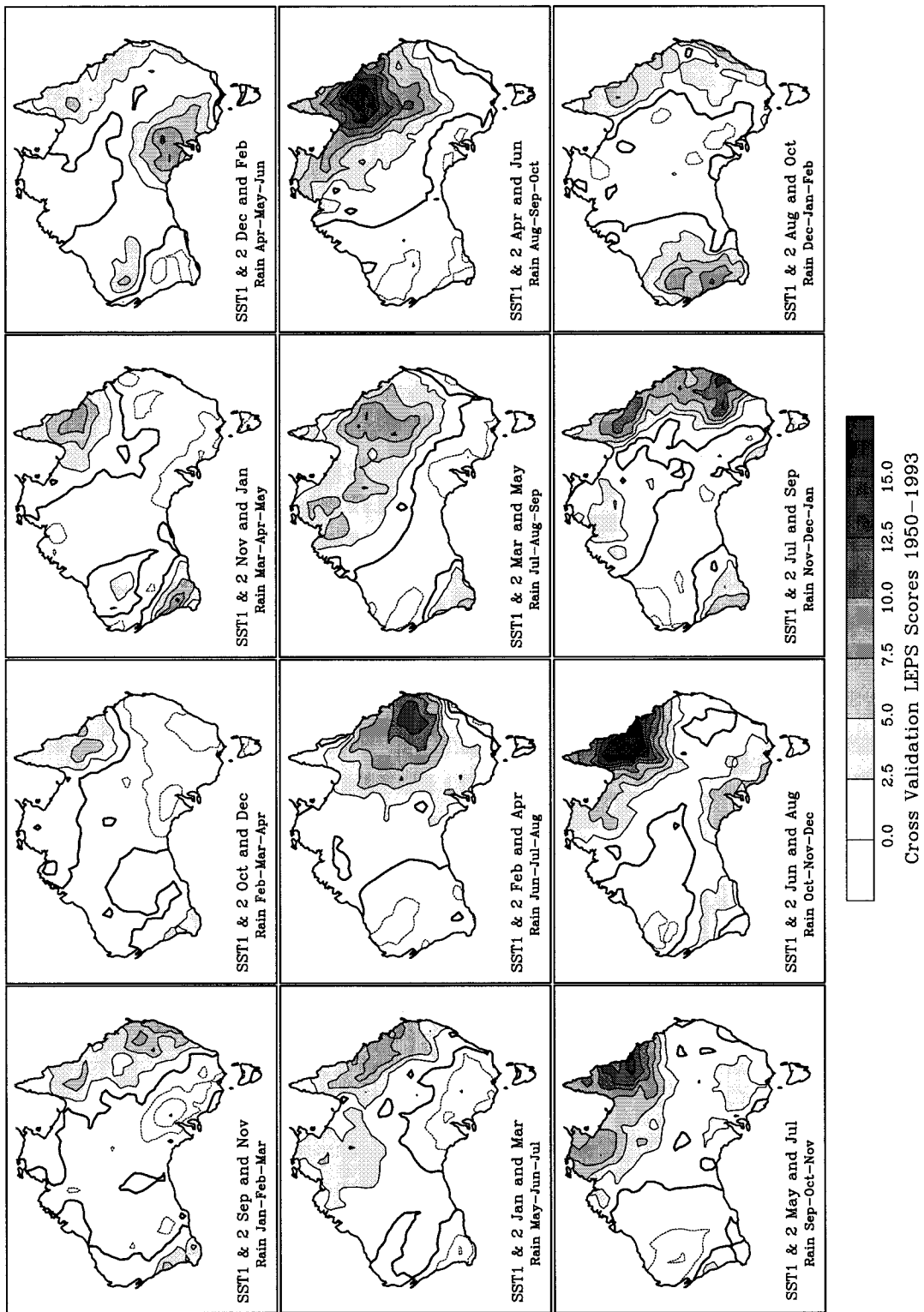


FIG. 4. Cross-validated LEPS scores for each of the 12 overlapping 3 monthly seasonal rainfall hindcasts using the first 2 rotated SST principal components, lagged by 1 and 3 months as predictors, for the period 1950-93. The 5% and 10% significance levels for a model with four predictors are estimated by Monte Carlo simulations to be about 10.8 and 7.6, respectively.

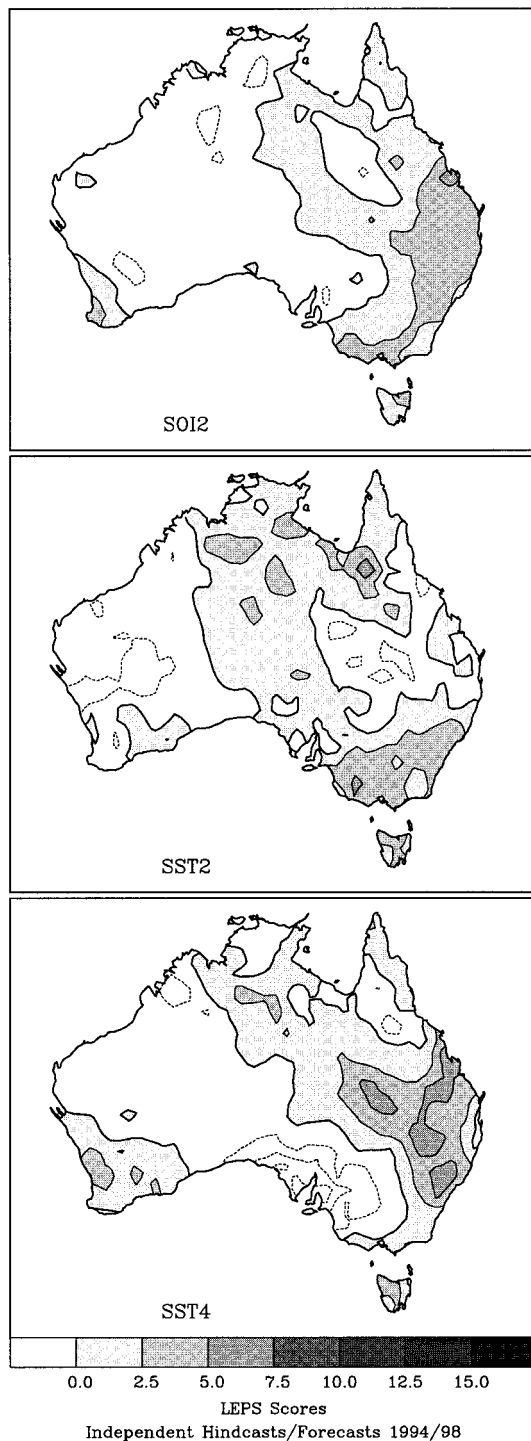


FIG. 5. LEPS skill scores aggregated over 60 independent hindcasts/forecasts for seasonal rainfall from Jan–Feb–Mar 1994 to Dec–Jan–Feb 1998/99, (a) for two seasons lagged SOI as predictors (SOI2), (b) best subset of two SST principal components chosen from the full pool of 10 at 1 and 3-month lag (SST2), and (c) the first 2 SST components lagged at 1 and 3 months (SST4).

SOI-based system outperforms both SST schemes, while the reverse occurs during 1996, although SST2 is somewhat erratic. All schemes perform well during 1997, particularly in winter during the early stages of the 1997/98 El Niño event. All perform poorly during the wetter than expected (given the intense El Niño) spring and summer of 1997/98. Finally during most of 1998 the SST-based schemes, particularly SST4 performed better than the SOI, during the development of the weak to moderate 1998/99 La Niña event. The time mean of these scores (1.73 SOI2, 1.23 SST2, and 0.88 SST4), are not significantly different from zero or each other. The higher value for SOI2 is a result of the good forecasts by this model in 1994; the 4-yr average excluding this year reverses the overall result (0.98 SOI2, 0.83 SST2, and 1.95 SST4).

Although this is a relatively small sample, the seasonal cycle in skill may be discerned, with highest LEPS scores during winter and spring, particularly with the SOI in 1994, all schemes in 1997, and the SST schemes in 1998.

## 5. Conclusions

The predictability of Australian seasonal rainfall anomalies, using near-global SSTA principal components as predictors has been examined. Extensive testing was performed to determine the best set of predictors using a cross-model validation technique (Elsner and Schmertmann 1994). Using only four predictors (two SST components each at two lags), the best subsets were found to vary considerably between adjacent rainfall regions as represented by the rainfall principal components used as predictands, and more importantly between overlapping seasons. Adding more potential predictors by using 10 components does not result in significantly improved hindcast skill. Therefore, the model chosen for operational seasonal forecasts uses the first two rotated components lagged by 1 and 3 months as predictors for every season and location, in an attempt to maintain continuity of forecast probabilities between the overlapping 3-month seasons. The physical and dynamical connections between the first two SST components and Australian rainfall are also well understood.

The forecasts are produced using discriminant analysis, which has two major desirable properties as a forecast method. First, it explicitly produces probability forecasts, although these can be easily converted to categorical forecasts if desired. Second, it is inherently a nonlinear technique for three or more predictand categories. This is desirable since the relations of many of the predictors examined, including those associated with ENSO, with atmospheric circulation, and rainfall have some nonlinear characteristics (Drosowsky and Williams 1991).

Forecast skill has been determined using the LEPS skill score. The properties of this measure are discussed in detail by Potts et al. (1996). Overall the results sug-

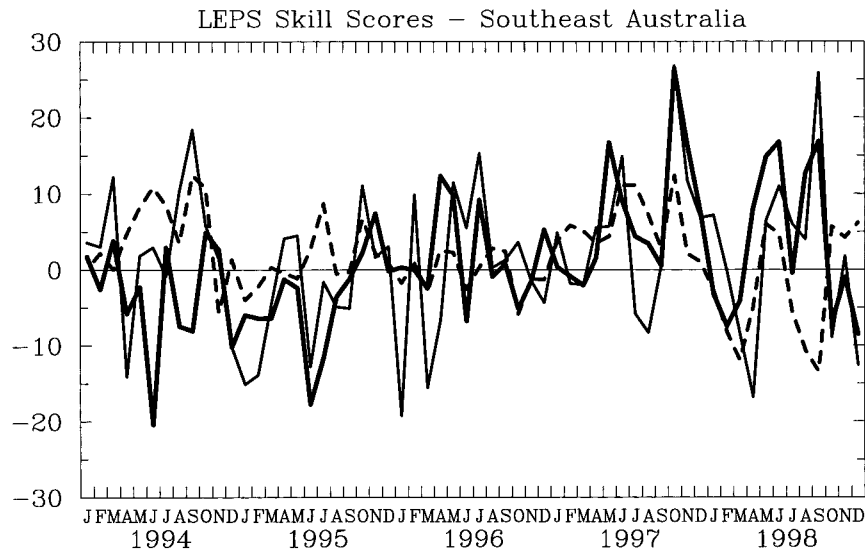


FIG. 6. LEPS skill scores aggregated over all grid points in southeast Australia (south of 29°S and east of 141°E) for each of the 60 seasons from Jan–Feb–Mar 1994 to Dec–Jan–Feb 1998/99, for the three sets of forecasts in Fig. 5: SOI2 (dashed curve), SST2 (light solid curve), and SST4 (heavy solid curve). Mean values for the three forecasts are 1.73 (SOI2), 1.23 (SST2), and 0.88 (SST4).

gest a small increase in hindcast skill of the SSTA patterns over the SOI used alone through the so-called “predictability barrier” in late summer–early autumn. At other times of the year skill levels of the SST and SOI models are comparable.

*Acknowledgments.* This research was supported in part by a Land and Water Resources Research and Development Corporation (LWRRDC) Grant, Project BOM1. Assistance and advice on various aspects of this project was provided by present and former members of the BMRC Climate Group; Neville Nicholls, Neville Smith, Carsten Frederiksen, David Jones, Richard Kleeman, Ramesh Balgovind, and Alex Kariko.

#### REFERENCES

- Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7**, 1513–1564.
- Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541–560.
- Drosowsky, W., 1993a: An analysis of Australian seasonal rainfall anomalies: 1950–1987. I: Spatial patterns. *Int. J. Climatol.*, **13**, 1–30.
- , 1993b: An analysis of Australian seasonal rainfall anomalies: 1950–1987. II: Temporal variability and teleconnection patterns. *Int. J. Climatol.*, **13**, 111–149.
- , 1993c: Potential predictability of winter rainfall over southern and eastern Australia using Indian Ocean sea-surface temperature anomalies. *Aust. Meteor. Mag.*, **42**, 1–6.
- , and M. Williams, 1991: The Southern Oscillation in the Australian region. Part I: Anomalies at the extremes of the oscillation. *J. Climate*, **4**, 619–638.
- , 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, Australia, 39 pp.
- Elsner, J. B., and C. P. Schmertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619–624.
- He, Y., and A. G. Barnston, 1996: Use of discriminant analysis for seasonal forecasts of surface climate in the United States. *Proc. Twenty-first Annual Climate Diagnostics and Prediction Workshop*, Huntsville, AL, NCEP/Climate Prediction Center, 67–70.
- Huberty, C. J., 1994: *Applied Discriminant Analysis*. Wiley, 466 pp.
- Jones, D. A., and G. Weymouth, 1997: An Australian monthly rainfall dataset. Tech. Rep. 70, Bureau of Meteorology, Melbourne, Australia, 19 pp.
- McBride, J. L., and N. Nicholls, 1983: Seasonal relationships between Australian rainfall and the Southern Oscillation. *Mon. Wea. Rev.*, **111**, 1998–2004.
- Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.*, **26**, 1589–1600.
- Nicholls, N., 1989: Sea surface temperatures and Australian winter rainfall. *J. Climate*, **2**, 965–973.
- Parker, D. E., C. K. Folland, A. C. Bevan, M. N. Ward, M. Jackson, and K. Maskell, 1995: Marine surface data for analysis of climatic fluctuations on interannual-to-century time scales. *Natural Climate Variability on Decadal-to-Century Time Scales*, National Research Council, 241–252.
- Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton, 1996: Revised “LEPS” scores for assessing Climate model simulations and long-range forecasts. *J. Climate*, **9**, 34–53.
- Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimal interpolation. *J. Climate*, **7**, 929–948.
- Smith, N. R., 1995: The BMRC ocean thermal analysis system. *Aust. Meteor. Mag.*, **44**, 93–110.
- Stone, M., 1974: Cross-validated choice and assessment of statistical predictions. *J. Roy. Stat. Soc.*, **B36**, 111–147.
- Ward, N. M., and C. K. Folland, 1991: Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea surface temperature. *Int. J. Climatol.*, **11**, 711–743.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.