A comparison of two seasonal rainfall forecasting systems for Australia

R.J.B. Fawcett[1] and R.C. Stone[2]
[1]National Climate Centre, Bureau of Meteorology, Australia
[2]Australian Centre for Sustainable Catchments, University of Southern Queensland, Australia

(Manuscript received October 2007, revised October 2009)

Operational El Niño–Southern Oscillation (ENSO)-based statistical forecasting of seasonal rainfall has been undertaken in Australia now for the best part of two decades. This article compares the performance in recent years (1997 to 2009) of the two major governmental programs currently forecasting seasonal rainfall across Australia, through the verification of independent forecasts. These programs are run by the Australian Bureau of Meteorology and the Queensland Government's Department of Environment and Resource Management (and its predecessors). In this study, the Queensland Government's Southern Oscillation Index (SOI) phase-based forecasts are not verified directly. Instead, the SOI-phase system methodology is reconstructed using the Bureau's monthly rainfall analyses and those reconstructed forecasts verified. Verification techniques employed include linear error in probability space (LEPS2) skill scores, per cent consistent rates, and reliability statistics.

Over recent years, the Bureau forecasts and the SOI-phase forecasts have performed comparably when measured by the per cent consistent statistic, although with successes and failures in different parts of the country. Both sets of forecasts show widespread forecast skill in excess of climatology, the Bureau forecasts performing better in the west of the country and the SOI-phase forecasts performing better in the centre and east. The generally more emphatic SOI-phase forecasts translated into higher LEPS2 skill scores, but the Bureau forecasts showed greater reliability.

Introduction

There are at present two major governmental seasonal rainfall forecasting programs in Australia. The first of these programs is run by the Australian Government through the Bureau of Meteorology and commenced in 1989. The second is run by the Queensland Government (QG) through its Department of Environment and Resource Management (and its predecessors) and commenced in 1994. Both programs issue seasonal (three-month) rainfall outlooks, using the format of the (conditional) probability of exceeding the (climatological) seasonal median, employing empirical statistical schemes informed by climatological understanding of relevant mechanisms. In what follows, these two programs or systems will be referred to as the Bureau and QG systems, respectively.

Corresponding author address: R. Fawcett, National Climate Centre, Bureau of Meteorology, GPO Box 1289, Melbourne, Vic. 3001, Australia. Email: r.fawcett@bom.gov.au

The primary known driver (apart from weather noise) of inter-annual climate variability in Australia is the El Niño-Southern Oscillation (ENSO), which accordingly must be taken into account when devising seasonal forecasting systems for Australia. Both forecasting systems use an ENSO index as the primary predictor, currently in the form of a sea-surface temperature (SST) index for the Bureau system and the Southern Oscillation Index (SOI) for the QG system. The Bureau system additionally uses an SST index with substantial input from the tropical Indian Ocean, particularly those waters between southern India and Western Australia (Drosdowsky and Chambers 1998, 2001).

In assessing forecast skill, we undertake measurements of how good a forecasting method is and/or is expected to be. There are many different metrics or skill scores, with different scores for different applications. The actual forecast format is an important ingredient in the selection of an appropriate verification score. Another important issue is the distinction between validation and verification.

For the purposes of this article, *validation* is taken to indicate skill assessment of independent hindcasts. This is typically done by 'leave one out' cross-validation (see for example Drosdowsky and Chambers 1998) or a 'two-thirds/one-third split' method (or equivalent) if there is sufficient data available. Model validation techniques are used in the construction of forecast models, selection of methods and predictors, and form an extremely important part of the forecast model-building process.

Model *verification* is taken to indicate skill assessment of independent forecasts, as issued operationally or quasi-operationally. It is used to check how a forecast system is actually performing. This can be for quality control and accountability purposes (e.g. key performance indicators), but it also can be used to benchmark a forecast system against other systems. In the longer term, it may help detect distortions caused by climate change to the underlying statistical relationships between predictor and predictand assumed in the model-building process. (These ideas of validation and verification apply equally to statistical seasonal forecasting and dynamical seasonal forecasting (through the use of coupled general circulation models).)

The purpose of this article is to compare these two governmental programs in Australian seasonal rainfall forecasting from the verification (independent forecasting) perspective, performing verifications of forecasts actually issued to the general public. In the case of the QG forecasts, neither the forecast grids nor an appropriate set of homogeneous seasonal gridded rainfall analyses are available to us. This prevents us from performing a direct verification of the QG forecasts. Instead, we approach the verification of the QG forecasts indirectly, by reconstructing the QG SOI phase forecasts using the Bureau's operational monthly rainfall analyses to form seasonal analyses and verifying those reconstructions. We will refer to those forecasts as the 'reconstructed SOI phase forecasts', to distinguish them from the QG forecasts as issued operationally. Verification techniques employed in the verifications include linear error in probability space (LEPS2) skill scores (Potts et al. 1996), per cent consistent rates, and reliability data.

The seasonal forecasting systems

Both seasonal forecasting systems being considered in the present work have been published in the refereed scientific literature (Stone and Auliciems 1992; Stone et al. 1996; Drosdowsky and Chambers 1998, 2001), but for completeness their essential characteristics are given below. Because we are verifying a reconstruction of the SOI phase scheme using the Bureau's monthly rainfall analyses, something which has not yet appeared in the literature, we give additional details of this reconstruction and its hindcast model validation skill (see the Appendix).

Bureau of Meteorology

The current statistical scheme (Drosdowsky and Chambers 1998, 2001) used by the Bureau uses indices of SSTs in the tropical Pacific and Indian Oceans as predictors (see Fig. 1 in Fawcett et al. 2005, hereafter FJB05). These are the first two rotated empirical orthogonal functions (EOFs) obtained from a principal components analysis (PCA) of near-global monthly SSTs. The predictions are made using the technique of linear discriminant analysis (LDA). Over the period SON[1] 1997 to the present, there have been some changes in the choices of predictors. The current SST predictors are used at lags of one and three months, which means (for example) that January and March SST data are used to predict AMJ rainfall. Previously, three-month averages of the SOI[2] were used in a similar manner as predictors. (See FJB05 for a description of the various changes that have taken place over the period of interest.)

The Bureau's seasonal rainfall forecasts are issued to the public in two main formats: (a) the probability of the rainfall total being in each of the three climatological terciles for the season in question; and (b) the probability of the rainfall total being above the climatological median for the season in question (from which the probability of the below-median outcome can obviously be derived). For the purposes of this study (and those of the two previous Bureau verification studies FJB05 and Fawcett (2008b), henceforth F08b), the terciles forecasts are available from SON 1997 to the present, while the above-median forecasts are available from JJA 2000 to the present. This discrepancy in the two periods arises in part because in the early years of the current verification period, the Bureau placed much greater emphasis on the terciles forecasts than it did on the above-median forecasts, but by JJA 2000 that emphasis had reversed, with the above-median forecasts taking precedence. (Recently, additional forecast formats have been adopted for seasonal rainfall prediction, but the new formats are not discussed in this article – see Fawcett et al. (2009) for further details.)

The changes in the choices of predictors mentioned above means that the terciles forecasts cannot be considered to be completely homogeneous across the period SON 1997 to the present, but as this is a verification study of publicly issued forecasts, that lack of complete homogeneity will be ignored. The above-median forecasts are homogeneous in this sense across the period JJA 2000 to the present.

Both forecast formats are prepared in the form of 1°×1° grids across the country. The terciles forecasts are verified against climatological grids calculated from a 98-year (1900-1997) data-set consisting of the Bureau's 0.25°×0.25° Barnes successive correction analyses (Jones and Weymouth 1997), with the 0.25°×0.25° grids being re-gridded onto 1°×1° for the verification. The above-median forecasts are likewise verified at 1°×1° resolution, with the model climatology period 1950-1999 being used to calculate the climatology median grids.

1 The seasons are indicated as JFM = January-February-March, FMA = February-March-April, etc.

2 The Bureau's calculation of the Troup SOI (Troup 1965) was used for this purpose, derived from the 60-year base period 1933-1992. It is the standardised anomaly of the monthly MSLP difference between Tahiti and Darwin (Australia). The normalisation is carried out separately for each calendar month.

In both cases, the observational seasonal rainfall totals grids are obtained from the Bureau's operational monthly rainfall analyses (Jones and Weymouth 1997).

Queensland Government

The QG forecast system uses the statistical technique of 'stratified climatology', employing five fixed categories or phases (Stone and Auliciems 1992; Stone et al. 1996) derived from pairs of consecutive monthly values of the SOI[3]. These phases were obtained by the application of cluster analysis techniques to the results of a PCA on the pairs of monthly SOI values[4]. The five phases of the SOI are (1) consistently negative, (2) consistently positive, (3) rapidly falling, (4) rapidly rising and (5) consistently near zero.

The technique of 'stratified climatology' involves matching the current situation (through the current SOI phase, for example) with those years in the historical record with a matching phase (the stratification). A conditional probability of exceedance of a given threshold (such as the climatological median) is then calculated, and can be compared against the corresponding climatological probability calculated using all years in the historical record. The SOI phase is used at zero months lag (compare with the one-month lag for the Bureau scheme) which means (for example) that January/February SOI values are used to predict MAM rainfall.

This approach to seasonal forecasting was in part inspired by the work of Williams (1987), who noted that in order to capture important tropospheric relationships associated with the SOI, month-to-month changes of the SOI needed to be considered. In this respect, an important consideration in the derivation of the QG SOI phases was the need to capture both consistency and change in the SOI with the associated due weighting of each aspect.

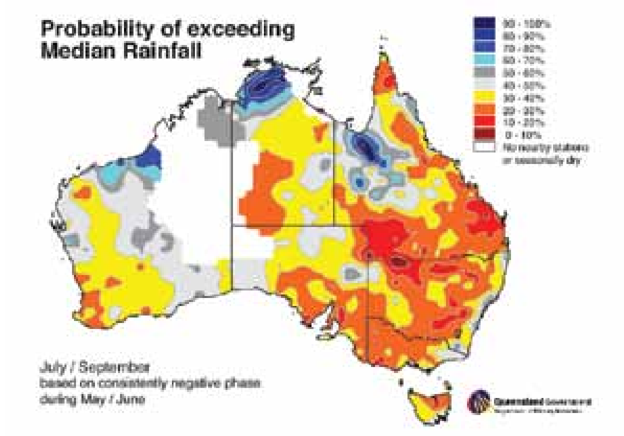
The QG system uses station rainfall data to calculate seasonal outlook probabilities at the station level directly, with the results analysed for spatial needs through kriging methods. Forecast probabilities for station locations are also available and provided through computer programs such as Australian Rainman. The principal forecast format is the probability of above-median seasonal (three-month) rainfall, although we note that Australian Rainman in effect makes available the entire conditional probability distribution at station locations.

As noted in the Introduction, neither the gridded forecasts nor gridded verifying rainfall analyses are available to us. Hence for the purposes of this study, we reconstruct the SOI phase scheme using the Bureau's operational monthly gridded rainfall analyses (Jones and Weymouth 1997). These analyses are available from 1900 to the present at 0.25° resolution.

3 The Troup SOI is also used for this purpose, but normalised with respect to the base period 1887-1989 in the QG forecast system.

4 Consecutive monthly values of the SOI are moderately correlated, with some variation in strength throughout the annual cycle. The application of PCA to such pairs yields two EOFs, the first representing (approximately) the mean SOI value, and the second (again, approximately) the difference. The amplitudes of these EOFs are temporally uncorrelated as a result of the PCA, and standardised prior to the cluster analysis.

Fig. 1 Official QG SOI phase 1 (consistently negative) for JAS 2006. White areas indicate data voids and/or seasonally dry areas.



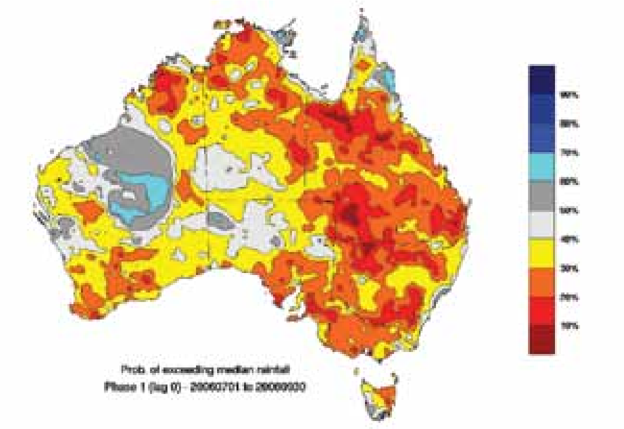
Therefore, for the rest of this article (unless otherwise stated), by the SOI phase scheme we mean its reconstruction using the Bureau's rainfall analyses. The reconstruction is performed at 0.25° resolution, apart from the calculation of reliability statistics which is done at 1° resolution.

Figure 1 shows an example of the official QG outlook for JAS 2006, based on the SOI phase for May/June 2006 being in the consistently negative phase. It is obtained from the Long Paddock website (www.longpaddock.qld.gov.au). Previous years with this May/June phase include 1888, 1896, 1905, 1911, 1912, 1914, 1940, 1946, 1972, 1977, 1987, 1993, 1994, 1997 and 2002. El Niño years are, not surprisingly, well represented in this set.

The reconstructed SOI phase forecasts are calculated using SOI phase data obtained from the QG's Long Paddock website[5], based on the Troup SOI (Troup 1965) calculated with the base period 1887 to 1989. The forecasts are reconstructed using a climatology period from 1900 to the year prior to the forecast period, and the SOI phase is used at lag 0 months. For example, all autumns from 1900 to 2005 are used to forecast and verify the autumn 2006 season (by comparing the autumn 2006 rainfall with the climatological autumn median calculated over the period 1900 to 2005). A consequence of this is that a phase N forecast (N = 1,…,5) issued in 2006 (say) will differ slightly from the analogous phase N forecast issued in 2005 (assuming, of course, similar behaviour in the SOI), because of the additional year's data. This means that our reconstruction of the SOI phase scheme differs slightly from the operational QG scheme which does not routinely update the climatology period from year to year. Figure 2 shows the reconstruction of JAS 2006 forecast using the Bureau's monthly rainfall analyses. The climatology period used in the reconstruction is 1900 to 2005.

5 http://www.longpaddock.qld.gov.au/SeasonalClimateOutlook/SouthernOscillationIndex/SOIDataFiles/MonthlySOIPhase1887-1989Base.txt

Fig. 2 Reconstruction of the SOI phase 1 (consistently negative) forecast for JAS 2006, based on the Bureau's rainfall analyses from 1900 to 2005.



The larger scale features of Figs 1 and 2 are in reasonable agreement (for example, the increased chances of below-median rainfall over southern Queensland, New South Wales, Victoria and southeast South Australia), but there are some substantial discrepancies at the regional scale (e.g. the Top End of the Northern Territory and the southern Gulf of Carpentaria). Possible sources of these discrepancies include the different rainfall analysis techniques and different base periods used in the computations.

The QG has been issuing seasonal outlooks in this form (i.e. probability of above-median outcomes) since 1994, but for purposes of comparison with the Bureau forecasts, we have reconstructed the above-median seasonal rainfall SOI phase forecasts from SON 1997 to the present. Analogous terciles forecasts have been prepared by the QG since at least mid-2005 but, since these forecasts have been given much less prominence in terms of public issuance, we have elected not to reconstruct them for verification purposes.

Verification techniques

Following FJB05 and F08b, we use the per cent consistent rate and the LEPS2 skill score (Potts et al. 1996) as the skill metric used to verify the forecasts, supplemented by the calculation of reliability statistics. Per cent consistent rates and LEPS2 skill scores are used to score the forecasts – the details below are taken from Walsh et al. (2001) and FJB05. The per cent consistent rates measure the fraction of times the forecast probabilities swung from climatology (i.e. the forecast in which the categories are considered to be equally likely) in the direction subsequently observed. Both these methods treat the two (above/below median) and three (tercile) categories together, rather than showing preference for one category over the other/others.

Above-median forecast format

For above-median forecasts, let q be the forecast probability, expressed as a fraction of 1. Let r be the observed seasonal rainfall and r_m be the climatological median seasonal rainfall against which the forecast is verified. The per cent consistent rate arises in the context of 2×2 contingency tables (e.g. Mason 2003), where it is called the proportion correct, the per cent consistent rate merely being the proportion correct expressed as a percentage rather than as a fraction of 1). If ($q > ½$ and $r > r_m$) or ($q < ½$ and $r < r_m$), then the contribution of the forecast to the proportion correct is 1. Conversely, if ($q > ½$ and $r < r_m$) or ($q < ½$ and $r > r_m$), then the contribution of the forecast to the proportion correct is 0. (Special arrangements can be made for the cases where $q = ½$ and/or $r = r_m$, for example by a *pro rata* distribution of the contribution.) The proportion correct is then the average of these contributions across the set of forecasts, and will be converted into a percentage for mapping purposes. The base rate for comparison purposes is ½ or 50 per cent: per cent consistent rates above 50 per cent can be considered skilful relative to climatology.

For the LEPS2 scoring system, if $r > r_m$, then the contribution by the forecast to the LEPS2 skill score is $2q - 1$, while if $r < r_m$, then the contribution to the LEPS2 skill score is $1 - 2q$. (Again, special arrangements can be made for the cases where $q = ½$ and/or $r = r_m$.) The LEPS2 skill score is the average of these contributions across the set of forecasts. Like the proportion correct, it will be converted into a percentage for mapping purposes. The base rate for comparison purposes is zero per cent: LEPS2 skill scores above zero per cent can be considered skilful relative to climatology.

Terciles forecast format

For terciles forecasts, let p_1, p_2 and p_3 be the probabilities that the seasonal rainfall will be in the climatological terciles 1, 2 and 3, respectively, these probabilities being expressed as fractions of one. Obviously, $p_1 + p_2 + p_3 = 1$. Let r_{12} and r_{23} be the seasonal rainfall totals representing the climatological 33.33…rd and 66.66…th percentiles, respectively, against which the forecast is verified. If ($p_1 > \max(p_2,p_3)$ and $r < r_{12}$) or ($p_2 > \max(p_1,p_3)$ and $r_{12} < r < r_{23}$) or ($p_3 > \max(p_1,p_2)$ and $r > r_{23}$), then the contribution of the forecast to the proportion correct is one, otherwise zero. (Again, special arrangements can be made in the special cases $r = r_{12}$, $r = r_{23}$, the two larger probabilities are equal, or all three probabilities are equal.) The proportion correct is then the average of these contributions across the set of forecasts, and will be converted into a percentage for mapping purposes. The base rate for comparison purposes is one-third or 33.33… per cent: per cent consistent rates above 33.33… per cent can be considered skilful relatively to climatology. (An above-median forecast format per cent consistent rate r, expressed as a percentage, can be transformed into a skill score $2(r - 50)$, while a terciles forecast format per cent consistent rate s, also expressed as a percentage, can be transformed into a skill score $(3s - 100)/2$.)

Following Walsh et al. (2001) and FJB05, if $r < r_{12}$ then the forecast is given the LEPS2 score $s = {}^8/_{27}\, p_1 - {}^1/_{27}\, p_2 - {}^7/_{27}\, p_3$, if $r_{12} < r < r_{23}$ then the forecast is given the LEPS2 score $s = -{}^1/_{27}\, p_1 + {}^2/_{27}\, p_2 - {}^1/_{27}\, p_3$, and if $r > r_{23}$ then the forecast is given the LEPS2 score $s = -{}^7/_{27}\, p_1 - {}^1/_{27}\, p_2 + {}^8/_{27}\, p_3$. These scores range from $-{}^7/_{27}$ to $+{}^8/_{27}$, but can be multiplied by

$^{27}/_8$ to give a scaled LEPS2 score which ranges from $-^7/_8$ to +1. For a sequence $\{s_1,\ldots,s_n\}$ of these LEPS2 scores, calculate two additional sequences, the first being $\{u_1,\ldots,u_n\}$ which is the sequence of maximum possible LEPS2 scores given the observed outcomes, and the second being $\{l_1,\ldots,l_n\}$ which is the sequence of minimum possible LEPS2 scores given the observed outcomes. This implies that $l_i \le s_i \le u_i$ for $i = 1,\ldots,n$. If the outcome for the ith forecast is a tercile 1 or tercile 3 outcome, then $u_i = {}^8/_{27}$ and $l_i = -^7/_{27}$, whereas if the outcome for the ith forecast is a tercile 2 outcome, then $u_i = {}^2/_{27}$ and $l_i = -^1/_{27}$. The LEPS2 skill score for the sequence of forecasts is calculated as

$$\text{LEPS2 SKILL} = \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} u_i}$$

if the mean LEPS2 score is non-negative, and as

$$\text{LEPS2 SKILL} = \frac{\sum_{i=1}^{n} s_i}{-\sum_{i=1}^{n} l_i}$$

otherwise. As in the two-category case, the LEPS2 skill score can range from –1 to +1, but it will be converted into a percentage for mapping purposes. The base rate for comparison purposes is zero per cent: LEPS2 skill scores above zero per cent can be considered skilful relative to climatology.

These two skill metrics are supplemented by the calculation of reliability statistics (Hartmann et al. 2002). This involves conflating forecasts and the associated verifying observations across many different grid-points into a group and stratifying that group according to forecast probability. The intent is to assess for a given forecast probability of an above-median outcome (say) the rate at which that outcome actually occurs, and as such it represents a check that the forecasts are correctly calibrated to the observations rather than being a measure of forecast skill. While this technique can of course be applied in principle to a single grid-point, the number of forecasts available in the present study is very much smaller than would be necessary to make such an attempt meaningful. In an effort to minimise sampling variability and to give a whole-of-model estimate, the conflation has been performed across all Australian grid-points and all issued forecasts. Reliability statistics are calculated for the above-median forecasts only.

Results

Figure 3(a) shows the per cent consistent rates for the Bureau terciles forecasts (SON 1997 to MJJ 2009; 141 forecasts), while Fig. 3(b) shows the reduced forecast set of JJA 2000 to MJJ 2009 (108 forecasts). The climatological base rate for comparison is 33.33… per cent. Red shades indicate results better than climatology (in effect, positive skill), with blue and white shades results worse than climatology (in effect, negative skill). Figure 4 shows the per cent consistent rates for the Bureau above-median forecasts (JJA 2000 to MJJ 2009; 108 forecasts), with a climatological base rate for comparison of 50 per cent. The contours in Figs 3(a) and 3(b) have been chosen to be equivalent to those in Fig. 4, under the standard method of transformation to skill scores described in the previous section.

Fig. 3(a) Bureau per cent consistent rates (terciles), all forecasts from SON 1997 to MJJ 2009.

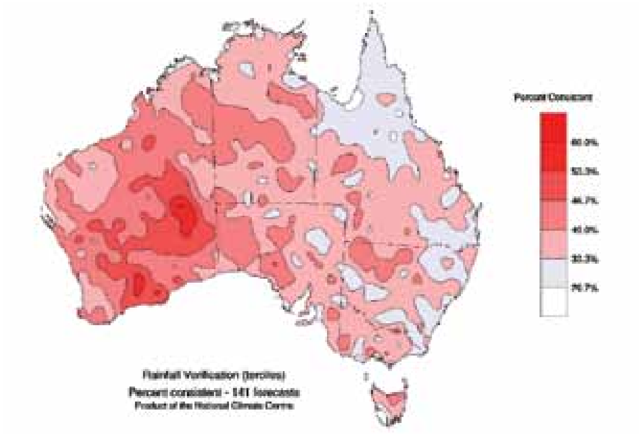


Fig. 3(b) Bureau per cent consistent rates (terciles), all forecasts from JJA 2000 to MJJ 2009.

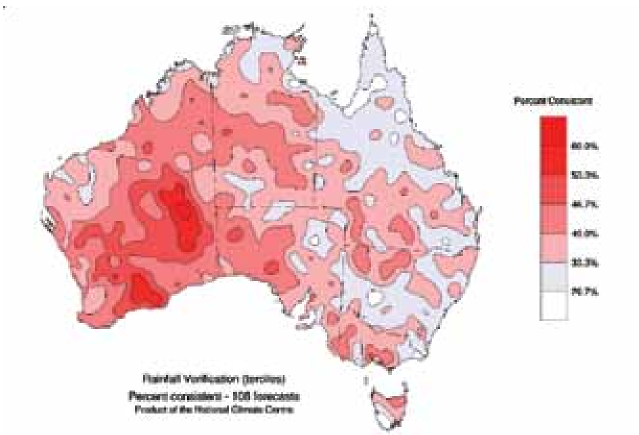
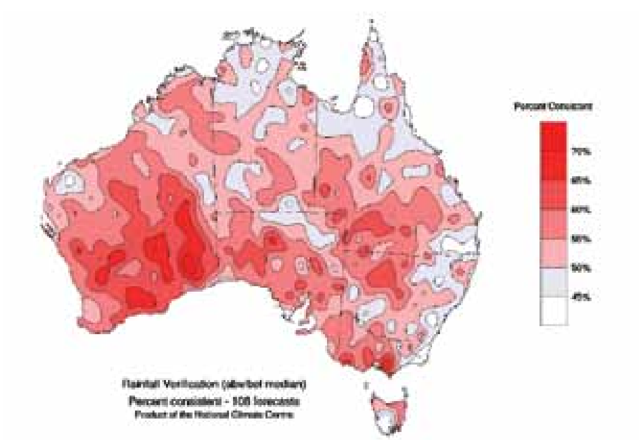


Fig. 4 Bureau per cent consistent rates (above median), all forecasts from JJA 2000 to MJJ 2009.



In spite of the variations in verification periods, forecast formats and verification climatologies, there is a lot of consistency between Figs 3 and 4. Verification skill has principally been exhibited in Western Australia, contrary to that expected from the hindcast validation study, which shows hindcast skill in Queensland and northern parts of the Northern Territory (see Fig. 18 in FJB05).

In contrast, the skill of the reconstructed SOI phase forecasts (see Fig. 5(a), which shows the per cent consistent rate for all above-median forecasts from SON 1997 to MJJ 2009, and Fig. 5(b), which shows the equivalent results for the restriction to the period JJA 2000 to MJJ 2009) has been principally in the central (SA and the NT) and eastern (Qld, NSW and Vic.) parts of the country. These are, of course, the parts of the country with the more consistent ENSO rainfall impacts. The reconstructed SOI phase system's success in the far west is all the more interesting. Neither system has performed particularly well across Tasmania. The grids used to prepare Figs 5(a) and 5(b) have been lightly smoothed to assist in the mapping process.

Figures 6 to 8 show the LEPS2 skill scores corresponding to the per cent consistent rates of Figs 3 to 5. As with Figs 5(a) and 5(b), the grids used to prepare Figs 8(a) and 8(b) have been lightly smoothed to assist in the mapping process. Theoretical considerations lead us to the conclusion that the LEPS2 skill scores for tercile forecasts can be considered as being approximately on the same scale as those for above-median forecasts (Fawcett 2008a). Accordingly, the same contour choices have been used, unlike those for the per cent consistent statistic. The various spatial distributions of the LEPS2 skill score correspond reasonably closely to those of the per cent consistent rates.

Figure 9 shows (verification) reliability data, accumulated across all Australian grid-points on a 1°×1° grid, for the Bureau's above-median forecasts between JJA 2000 and MJJ 2009. The distribution of forecast probabilities is also shown, at the bottom of the plot. In the construction of this figure, the forecast probabilities are sorted into single percentage bins. Figure 10 shows the corresponding results for the reconstructed SOI phase forecasts (above-median forecasts, all

Fig. 5(a) Reconstructed SOI phase forecast per cent consistent rates (above median), all forecasts from SON 1997 to MJJ 2009.

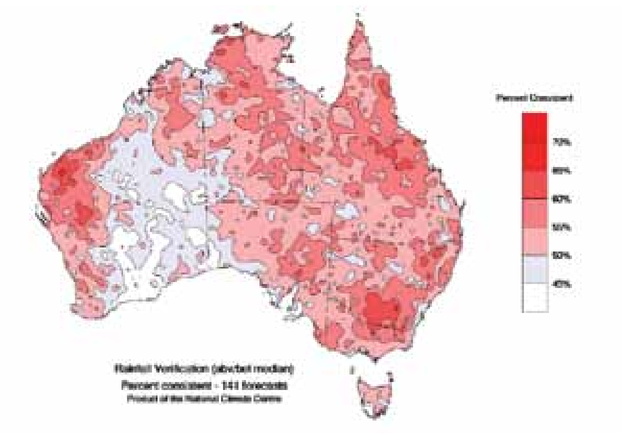


Fig. 5(b) Reconstructed SOI phase forecast per cent consistent rates (above median), all forecasts from JJA 2000 to MJJ 2009.

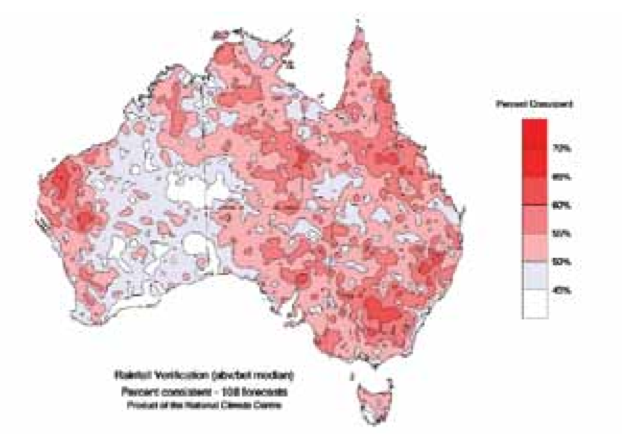


Fig. 6(a) Bureau LEPS2 skill scores (terciles), all forecasts from SON 1997 to MJJ 2009.

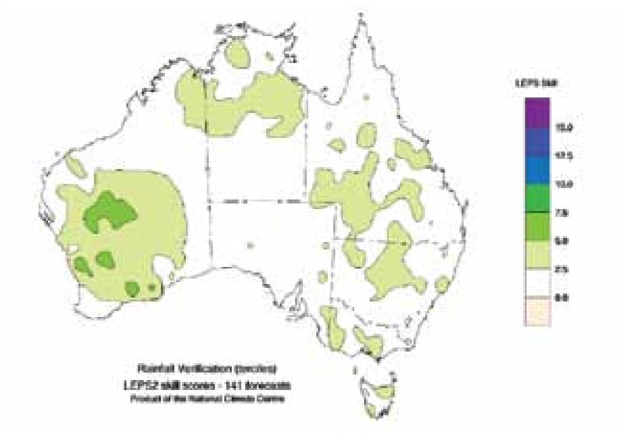


Fig. 6(b) Bureau LEPS2 skill scores (terciles), all forecasts from JJA 2000 to MJJ 2009.

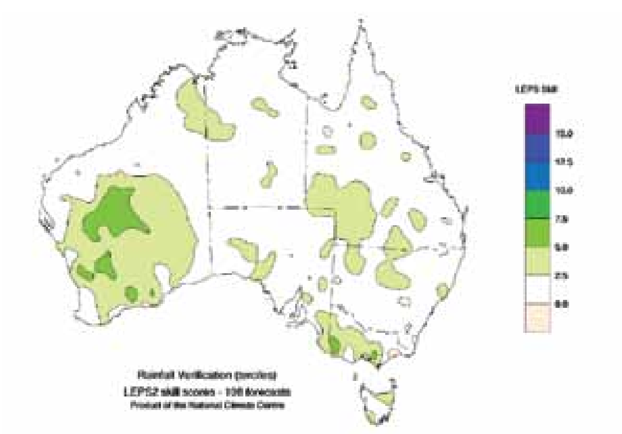


Fig. 7 Bureau LEPS skill scores (above median), all forecasts from JJA 2000 to MJJ 2009.

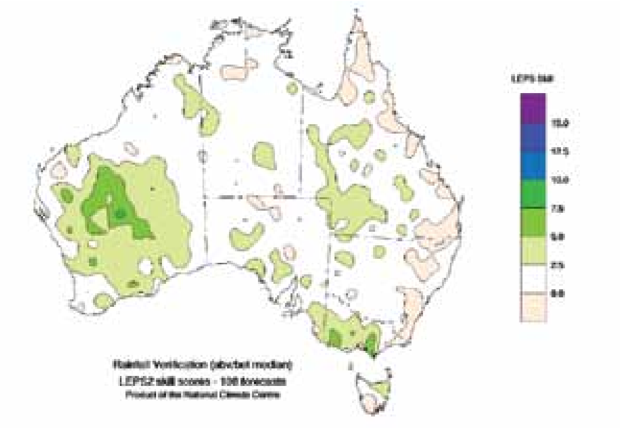


Fig. 8(a) Reconstructed SOI phase forecast LEPS2 skill scores (above median), all forecasts from SON 1997 to MJJ 2009.

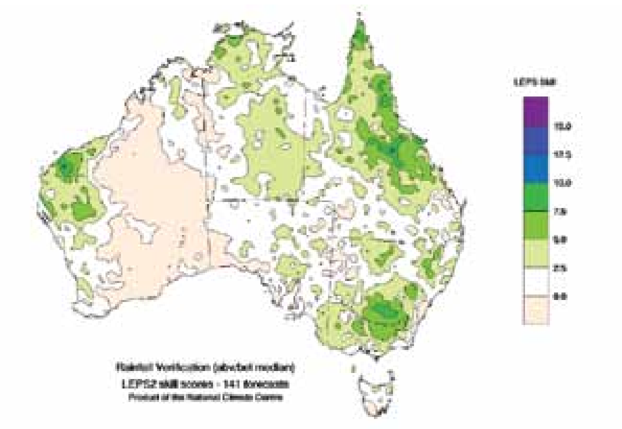
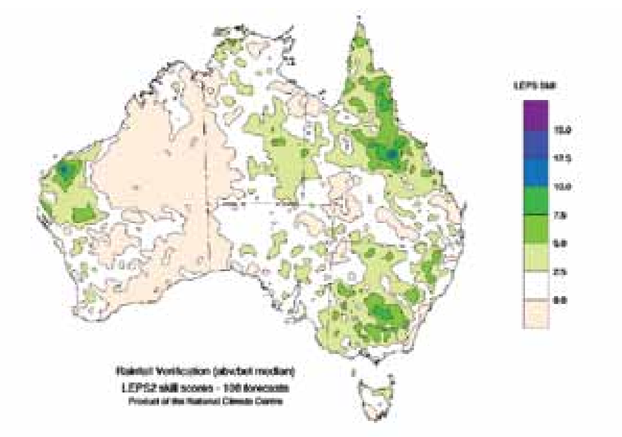


Fig. 8(b) Reconstructed SOI phase forecast LEPS2 skill scores (above median), all forecasts from JJA 2000 to MJJ 2009.



forecasts between SON 1997 and MJJ 2009). In both figures, the line of perfect reliability ($y = x$) is shown as a reference line.

These two figures show that at the national scale, the reconstructed SOI phase forecasts are typically much more emphatic (through having a much higher mean absolute departure from climatology) than the Bureau forecasts; the histogram of Bureau forecast probabilities is relatively narrow, while the reconstructed SOI phase forecasts histogram is much wider. This explains, at least in part, why the LEPS skill scores in Fig. 8 are generally higher than those in Figs 6 and 7. (Note that the bins corresponding to forecast probabilities of 49 per cent and 51 per cent are empty in the reconstructed SOI phase forecast set, because the various SOI phases do not contain enough years in them to generate those forecast probability values.) The greater forecast variance of the reconstructed SOI phase forecasts makes them (and likewise the QG SOI phase

Fig. 9 Bureau reliability data (above-median forecasts), all forecasts from JJA 2000 to MJJ 2009 and all Australian grid-points. The histogram of forecast probabilities is also shown. The straight line indicates perfect reliability.

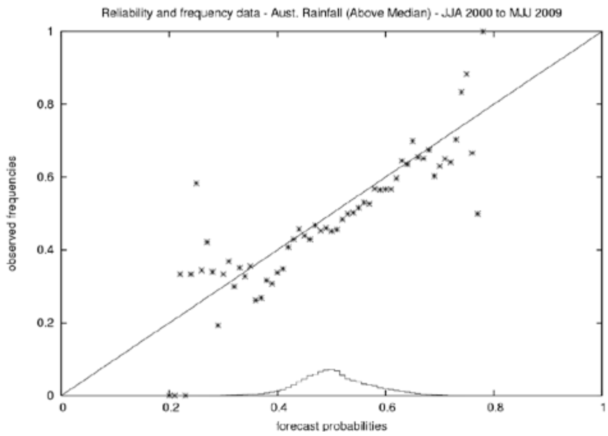
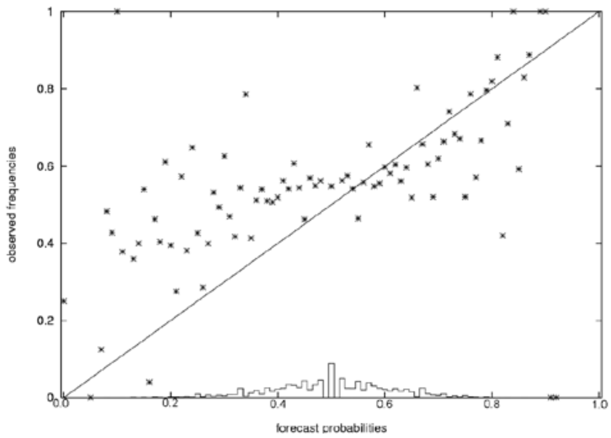


Fig. 10 Reconstructed SOI phase forecasts reliability data (above-median forecasts), all forecasts from SON 1997 to MJJ 2009 and all Australian grid-points on the 1°×1° grid lattice. The histogram of forecast probabilities is also shown. The straight line indicates perfect reliability.



forecasts that we are attempting to reconstruct) potentially more useful to the general public, but at a cost of reduced reliability. For reconstructed SOI phase forecast probabilities below 50 per cent, above median outcomes are typically occurring more frequently than forecast. Forecasts above 50 per cent proved more reliable in the reconstructions, on the other hand. When restricted to Queensland (not shown), the region of most predictability for the SOI phase scheme, the reliability of the reconstructed SOI phase forecasts was somewhat improved over the national picture, but still with considerably more scatter than the corresponding Bureau forecasts over the shorter period. The Bureau forecasts (Fig. 9) show much better reliability than the reconstructed SOI phase forecasts (Fig. 10).

Concluding remarks

In this article, we have compared verifications of the Bureau of Meteorology's official seasonal rainfall forecasts with those of a reconstruction of the Queensland Government's SOI phase (stratified climatology) seasonal rainfall forecasts. Both systems use statistical forecasting techniques and are based wholly or in part on teleconnections which exist at lag between Australian seasonal rainfall and broadscale atmospheric/oceanic circulation indices.

In general terms, they show similar levels of skill when independent forecasts are assessed over a reasonable length of time, although the reconstructed SOI phase system shows higher LEPS2 skill scores, at least in part arising out of more emphatic forecasts, while the Bureau system shows better reliability characteristics. The skill level demonstrated for seasonal rainfall forecasting is only moderate; while clearly better than climatological and randomly guessed forecasts, it is less than that achievable for Australian seasonal (three-month) temperature forecasting (FJB05 and F08b), but arguably of greater economic importance. (Hill et al. (2000) have reported on the economic value of SOI-based seasonal forecasts to Canadian and USA wheat producers, while Chen et al. (2002) have investigated the benefits which flow from use of the QG forecasting system specifically in the USA context.) We note that the QG system, with its selection of analogue years (used in the construction of the stratified climatology) is much more useful from an agricultural modelling perspective than the Bureau system. This is because the analogue years can provide daily and/or weekly climate data, which feed into crop and pasture models. In time, that use of high temporal resolution historical climate data might be supplemented, or indeed superseded, by the use of data from dynamical (coupled general circulation) seasonal forecasting models.

Acknowledgments

The authors wish to thank Neil Plummer and Grant Beard for providing helpful comments on a preliminary version of this article.

Appendix

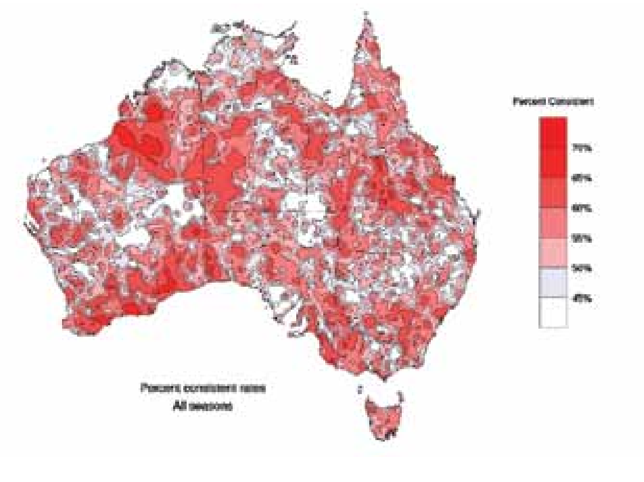
Hindcast skill estimates for the reconstructed SOI phase forecasts

As described in the main text, we have attempted in this study to reconstruct the QG SOI phase forecasts using the Bureau's monthly gridded rainfall analyses. In so far as the QG scheme and our reconstruction use different monthly/ seasonal rainfall analyses and different climatology periods, their results differ. Therefore, we present in this Appendix a model validation or hindcast skill estimate for the reconstructed SOI phase forecasts, as this has not previously appeared in the peer-reviewed literature.

For the purpose of generating a single hindcast skill estimate for the reconstructed SOI phase forecasts, the base period of 1900 to 1999 has been used (1900/01 to 1999/2000 for seasons NDJ and DJF). The commonly used approach of single cross-validation (e.g. Drosdowsky and Chambers 1998) is used to obtain a set of independent hindcasts. For each year, that year's data are removed, and the hindcast and climatological median for verification purposes generated using the remaining 99 years' data. (The SOI phases themselves are taken as given or fixed in this respect, and have not been subjected to cross-validation.)

Figure A1 shows the per cent consistent rate for all twelve seasons together (1200 hindcasts comprising the seasons JFM 1900 to DJF 1999/2000) for the reconstructed SOI phase system. The hindcasts and their verifications were performed using 0.25° resolution seasonal rainfall grids (as before, obtained from the Bureau's operational monthly rainfall analyses), but the resulting per cent consistent rates shown in the figure have been lightly smoothed to assist the mapping process. Figure A1 may be compared with Fig. 18 in FJB05 which shows the corresponding per cent consistent model validation rates for the Bureau system (598 forecasts comprising the seasons JFM 1950 to OND 1999). It indicates that the SOI phase scheme yields widespread skill in excess of climatology.

Fig. A1 Cross-validated hindcast skill estimate (1,200 hindcasts, all seasons together) for the reconstruction of the QG SOI phase system on the Bureau's operational monthly rainfall analyses.



References

Chen, C-C., McCarl, B. and Hill, H. 2002. Agricultural value of ENSO information under alternative phase. *Climatic Change, 54*, 305-25.

Drosdowsky, W. and Chambers, L.E. 1998. Near global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *BMRC Research Report No. 65,* Bur. Met., Australia (www.bom.gov.au/bmrc/clfor/cfstaff/wld/RESREP65/rr65.htm).

Drosdowsky, W. and Chambers, L.E. 2001. Near global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *Jnl Climate, 14,* 1677-87.

Fawcett, R.J.B. 2008a. Verification techniques and simple theoretical forecast models. *Weath. Forecasting, 23,* 1049-68.

Fawcett, R.J.B. 2008b. Verification of the Bureau of Meteorology's seasonal forecasts 2003-2005. *Aust. Met. Mag., 57*, 273-8.

Fawcett, R., Beard, G. and Jones, D. 2009. Extending the Bureau of Meteorology's seasonal forecast products. *Bull. Aust. Met. Oceanogr. Soc., 22*, 70-7.

Fawcett, R.J.B., Jones, D.A. and Beard, G.S. 2005. A verification of publicly issued seasonal forecasts issued by the Australian Bureau of Meteorology: 1998-2003. *Aust. Met. Mag., 54*, 1-13.

Hartmann, H.C., Pagano, T.C., Sorooshian, S. and Bales, R. 2002. Confidence builders: evaluating seasonal climate forecasts from user perspectives. *Bull. Am. Met. Soc., 84*, 683-98.

Hill, H.S.J., Park, J., Mjelde, J.W., Rosenthal, W., Love, H.A. and Fuller, S.W. 2000. Comparing the value of Southern Oscillation Index-based climate forecast methods for Canadian and US wheat producers. *Agric. Forest Meteor., 100*, 261-72.

Jones, D.A. and Weymouth, G. 1997. An Australian monthly rainfall data set. *Tech. Report No. 70,* Bur. Met., Australia.

Mason, I.B. 2003. Binary events. In *Forecast verification: a practitioner's guide*. Jolliffe, I.T. and Stephenson, D.B. (Eds.), Wiley.

Potts, J.M., Folland, C.K., Jolliffe I.T. and Sexton, D. 1996. Revised "LEPS" scores for assessing climate model simulations and long-range forecasts. *Jnl Climate, 9,* 34-53.

Stone, R.C. and Auliciems, A. 1992. SOI phase relationships with rainfall in eastern Australia. *Int. J. Climatol., 12,* 625-36.

Stone, R.C., Hammer, G.L. and Marcussen, T. 1996. Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. *Nature, 384*, 252-5.

Troup, A.J. 1965. The Southern Oscillation. *Q. Jl R. Met. Soc., 91*, 490-506.

Walsh, K.J.E., Bettio, L., Power, S., Fawcett, R. and Pahalad, J. 2001. Extended seasonal prediction of precipitation in Fiji. *Aust. Met. Mag., 50*, 195-203.

Williams, M. 1987. Relations between the Southern Oscillation and the troposphere over Australia. *BMRC Research Report No. 6*, Bur. Met., Australia.