Comparing Phase Based Seasonal Climate Forecasting Methods for Sugarcane Growing Regions

Everingham, Y.L.[1]

[1] School of Mathematics, Physics and Information Technology,
James Cook University, Queensland
Email: yvette.everingham@jcu.edu.au

Keywords: *Seasonal Prediction, Sugar, Comparison, Skill, Test*

EXTENDED ABSTRACT

Climate forecasting systems that group years on the basis of a climate forecasting index like the Southern Oscillation Index (SOI) or sea surface temperatures (SSTs) are quite simple to explain to industry personnel. Phase systems identify a subset of years (analogues) that have the same phase for a particular month. Industries can then investigate how the response of interest varied historically by the SOI or SST phase and self-validate the system. This is possible because industry members will remember the big wet and big dry years. Phase systems also allow industry personnel to visualise distributional shifts in rainfall and other responses (e.g. yield) between the different phases. These components spark a great deal of interest and enthusiasm at case study meetings. The simplicity of phase systems contributes to increased understanding of the forecasting approach, and highlights both the strengths and limitations associated with seasonal climate forecasting. Given that climate forecasts are not a perfect science, it is important that industries understand the risks and probability concepts so they can better integrate forecasts into a decision-making framework.

The Australian sugar industry has predominantly used the five-phase SOI climate forecasting system as its benchmark in recent years. The purpose of this paper is to compare the performance of the benchmark system with other phase-based climate forecasting systems. Three-phase and nine-phase SST forecasting systems and a three-phase SOI system formed part of the investigation. An assessment is made across the sugarcane growing regions and across the calendar year, simultaneously. This is done for seven sugar growing regions that collectively produce approximately 90% of Australia's sugar. A methodology that enables a fair comparison of the systems is presented. This methodology caters for the different number of phases with each forecasting system. We consider three performance measures: P-values of (i) the Kruskal-Wallis (KW) test statistic, (ii) a linear error in probability space (LEPS) skill score and (iii) a relative operating characteric (ROC) skill score for above and below median rainfall. P-values are used to overcome obstacles associated with the different numbers of phases. This is important since, by chance alone, it is easier to get a higher or better categorical LEPS score for systems that have more phases.

Results can vary with the performance measure. If ROC- and LEPS-based performance measures were preferred, then the three-phase SST system produced a higher number of significant results across the regions and three-month rolling periods. If performance measures that reflect the degree of distributional shifts or discriminatory ability between phases are preferred, then the five-phase SOI system produced the highest number of significant fields. Taking into consideration dependencies and auto-correlations associated with the response measurements across the calendar year and across coastal regions which essentially differ in latitudinal positioning, it is important to assess the likelihood that the number of significant fields could have occurred purely by chance.

Whilst a methodology for comparing different phase systems, where the number of phases varies from system to system is presented, the dilemma as to which performance measures to base decisions remains. Users must carefully consider which performance measures are most appropriate for their investigation.

1. INTRODUCTION

Natural swings between wet and dry years pose many challenges for agricultural industries. Certain farming operations and planning activities are more suited to wet conditions, whilst alternative operations are more suited to dry conditions. Perfect knowledge of whether the season ahead was going to be wetter or drier than average would greatly assist decisions and operations that are impacted by climate. Unfortunately, there is no such crystal ball and perfect knowledge about the climate to be experienced during the season ahead is unattainable. There are however, seasonal climate forecasts which provide probabilistic information on the likelihood that the season ahead will be wet, or will be dry. Once farmers and industry personnel have grasped the concepts of integrating probabilistic information about climate forecasting as part of a decision-making framework, then opportunities and benefits associated by better anticipating future climate conditions may be realised.

There exists a diverse range of forecasting systems that provide probabilistic information about the future state of the climate. Climate forecasting systems based on phases simplify the process of communicating climate forecasting technologies to industry. Phase systems vary in the number of phases, analogue years (*i.e.*, years with the same phase at a particular time) and the index from which the phases are derived. Some phase systems include the five-phase Southern Oscillation Index (SOI) (Stone and Auliciems, 1992), a three-phase SST system, the nine-phase SST system (Drosdowsky, 2002) and a three-phase SOI system. It is worthwhile noting that rigid definitions exist for defining phases for the nine-phase SST system and the five phase SOI system, but definitions vary slightly in the research community regarding how phases in the three-phase SST and three-phase SOI systems are constructed.

The Australian sugar industry has paid attention to the five phase forecasting system (c.f. Everingham *et al.*, 2002). The objective of this paper is to compare the forecasting performance of this system with three-phase SOI and SST systems and a nine-phase SST system. This paper outlines different investigations that were conducted in pursuit of identifying if one system might be more advantageous than another system from a statistical perspective. Whilst the assessment has focussed on sugarcane growing regions in Australia, the methodology is transferable to other industries.

2. RAINFALL DATA

A rainfall index for each rolling three monthly calendar period was computed individually for each major sugarcane growing region along the eastern coast of Australia. The seven regions represented in this study were (from north to south) Cairns (CNS), Mourilyan (MLN), Lucinda (LUC), Townsville (TVL), Mackay (MCK), Bundaberg (BUN) and north eastern New South Wales (NSW). The calendar or temporal periods were January to March (JFM), February to April (FMA), etc, through to December to February (DJF). Rainfall data from each region were obtained from the nearest high-quality official weather station to each of several mills in each region (Jones and Everingham, 2005). For each region and period, a principal component analysis was performed to provide a linearly-weighted rainfall index composed of mill rainfall data within each region. High (low) values of this regional rainfall index correspond to high (low) rainfall amounts across the mills. The index derived from the principal component analysis avoided the need to consider separate rainfall indices from the different weather stations within a region when a high proportion of variability was captured by the leading component.

3. CLIMATE FORECASTING SYSTEMS

This section describes four phase based systems that satisfy the definition of reliability as described in Murphy (1993) and Potgieter *et al.,* (2003).

3.1 Three Phase SOI System

The three-phase SOI system investigated here was based on monthly mean SOI values (available from http://www.longpaddock.qld.gov.au/SeasonalClimateOutlook/SouthernOscillationIndex) to define monthly El Niño, La Niña and neutral phases. An El Niño phase for a particular month was defined if the previous six-month running average of SOI values was less than −5. A La Niña month was defined if the running average of SOI values was greater than +5. Neutral conditions were specified if the six-month running average of SOI values was between −5 and +5 (inclusively). As an example, the October 1998 phase would be La Niña if the average of the monthly SOI values between April 1998 and September 1998, inclusively was greater than +5. The chance of rainfall exceeding the median for a particular period and region was estimated using relative frequencies from historical data. For example, if there were 20 La Niña phases for a particular month, and 15 of those years resulted in an above median rainfall index for the period immediately following the phase month, the probability of receiving an above median rainfall index would have been 0.75. This stratified climatological approach that uses relative frequencies to estimate the probability of an above or below median rainfall index was applied to each forecasting system discussed in this paper.

3.2 Three-Phase SST System

The three-phase SST system investigated here was based on the Niño 3.4 region sea surface temperature anomaly according to the dataset of Smith and Reynolds (2003) to define monthly El Niño, La Niña and neutral phases. An El Niño phase month was declared if the running average of SST anomalies for the previous three months was greater than +0.5°C. Conversely, a La Niña month was defined if the previous three month running average of SST anomalies was less than −0.5°C. Neutral conditions existed when the previous three month running average of SST anomalies was between −0.5°C and 0.5°C, inclusively. For example the October 1998 phase would have been La Niña if the average of the monthly SST anomalies between July 1998 and September 1998, inclusively was less than −0.5°C.

3.3 Five-Phase SOI System

The five-phase SOI system (Stone *et al.*, 1992, Stone *et al.*, 1996) is based on phases evident within the SOI which track the changes in the SOI on a month-to-month basis. The five phases are referred to as: consistently negative (CN), consistently positive (CP), rapidly falling (RF), rapidly rising (RR), and near zero (NZ). Each month is assigned one of these five phases.

3.4 Nine Phase SOI System

The nine-phase SST system (Drosdowsky, 2002) uses two different indices representing the first two rotated principal components from the Pacific and Indian Ocean SST anomaly patterns. The SST1 index derives principally from the Pacific Ocean, while the SST2 index derives principally from the Indian Ocean. The phases were defined by placing every monthly value from each index into one of three equally likely terciles: cold SST, neutral SST and warm SST. The three terciles of SST anomalies for each index/time series were combined to give a three-by-three classification, resulting in the nine phases as shown in Table 1.

Table 1. Phases from the nine-phase SST system.

		Pacific		
		Cold	Normal	Warm
Indian	Cold	1	2	3
	Normal	4	5	6
	Warm	7	8	9

3. PERFORMANCE MEASURES

3.1 The Kruskal-Wallis Test

Demonstrating large shifts in rainfall distributional patterns is an important factor when dealing with industry. The Kruskal-Wallis (KW) test (Conover, 1980), is a nonparametric alternative to a one-way analysis of variance that can assess the distributional shifts (Stone *et al.*, 2000) of a forecast system. The null hypothesis states no difference in distributions between different treatments (phases). The test statistic, H, for the KW hypothesis test is given as

$$H = \left(\frac{12}{n(n+1)} \sum_{k=1}^{K} \frac{R_k^{\,2}}{n_k} \right) - 3(n+1) \,,$$

where n is the total number of cases (years), n_k is the number of years from phase k and R_k is the sum of the ranked observations from phase k, where k varies from 1 to K with K being the total number of phases in the system. For example, K is equal to five if there are five phases. The test statistic H follows a Chi-Squared distribution based on $K-1$ degrees of freedom.

3.2 Linear Error in Probability Space Score

Linear error in probability space, (LEPS) scores (Potts *et al.*, 1996) range from −100% (worst possible score) to +100% (best possible score). A LEPS score for categorical forecasts was computed by multiplying the probability of belonging to each category with a series of weights (see Potts *et al.*, 1996) that are dependent on the category that eventuates. This was done for each forecast. The expected scores for each forecast were summed. A summed expected score that is positive is divided by the best set of scores that could be achieved from a perfect forecast system based on the observed data. For a negative score sum, the sum of expected scores is divided by the worst possible forecast system for the observed data. For example, if forecasting above or below median rainfall, LEPS scores close to zero would be produced if the forecast system continually gave a forecast close to climatology (*i.e.*, 50% chance of above median rainfall). Higher, positive percentages of LEPS scores are preferred.

3.3 Relative Operating Characteristic Score

Relative operating characteristic (ROC) scores represent the area beneath a curve that graphs the hit-rate versus the false-alarm rate for a series of warning thresholds (Mason and Graham, 1999). ROC scores range from 0 (worst possible score) to 1 (best possible score). A ROC score of 0.5 is produced from a forecasting system when the hit-rate is equivalent to the false-alarm rate, thus scores above 0.5 are preferred. A ROC score is computed for above median rainfall forecasts, and another score is computed for below median rainfall forecasts.

3.4 P-values and Null Distributions

P-values provide a way to quantitatively compare the quality of forecast systems (Maia *et al*., 2007). We have computed P-values of the performance measures. P-values are appropriate because the null distribution of performance measures can be influenced by the number of phases and potentially the number of analogues. This is especially true for LEPS scores. The Kruskal-Wallis test automatically factors in the number of phases when computing the degrees of freedom of the test.

P-values measure the probability of obtaining a performance measure more extreme than the one observed. In the case of the *H*-statistic, the P-value can easily be computed because the null distribution conforms to a Chi-squared distribution. For the other performance measures, the null distribution has been computed by using permutation methods (Good, 1997) where the P-value is the proportion of values in the null distribution greater than or equal to the actual performance measure. P-values close to zero (*e.g*., less than 0.1) are preferred since low values suggest any forecasting ability is unlikely a consequence of chance.

3. METHOD

Figure 1 outlines the methodology implemented in this paper to compare the different forecasting approaches using the P-values of the selected performance measures. The first step involves computing the rainfall indices for each three-month rolling period of the calendar year for each location using the approach described in Section 2. This produces 84 (7×12) rainfall indices in total. The second step selects and computes the performance measure for each point on the regional by temporal field. Step 3 computes the P-value for the selected performance measure for each of the 84 fields. It is important to recognise that the number of significant fields could be artificially inflated owing to dependencies in the field created by regional and temporal combinations. Using the approach of Wilks (1995, section 5.4.2) we have checked if the scores were artificially high by computing the global field significance. This involved counting the number of points in the 7×12 grid (or field) that had a P-value less than or equal to some specified value, e.g. α=0.10 (Step 4). The global significance was then computed (Step 5). Smaller values of the global significance indicate the observed number of significant fields is unlikely due to chance. If required, users may consider a stricter cut-off criterion. Once the global field significance has been computed the same routine is applied for the new performance measure (Step 6).

4. RESULTS AND DISCUSSION

Figure 2 shows the distributional shifts in September to November rainfall indices at Mourilyan for phases from (a) the three-phase SST system and (b) the five-phase SOI system (P=0.00 see Table 2). The lower and upper edge of the boxes show the 25th and 75th percentiles respectively and the horizontal line in the box coincides with the median. Lines from the upper (lower) edge of the box extend to the largest (smallest) rainfall index that is not deemed to be extreme and the floating horizontal lines identify extreme cases that are more than 1.5 times the inter-quartile range (IQR) above(below) the upper(lower) edge of the box. Industry members concerned about rainfall interruptions to the harvest will note that higher September to November rainfall indices at Mourilyan occur more often when the August SST phase is in a La Niña state (Figure 3a) and/or the August SOI phase is consistently negative or rapidly rising (Figure 3b).

Table 2 shows there were 48 P-values associated with the KW test statistic that were less than or equal to 0.10 for the five-phase SOI system. The number of significant fields for the remaining performance measures and forecasting systems is presented in Table 3. Also shown in Table 3 are the results which indicate if the number of significant fields satisfies global field significance. The small global P-values suggests that the number of significant fields is in most cases not overly inflated by field dependencies.

The number of significant fields varies with performance measure and climate forecasting system. If ROC- and LEPS-based performance measures were preferred, then the three-phase SST system produced a higher number of significant results across the field. If performance measures that reflect the degree of distributional shifts between phases were preferred, then the five-phase SOI system produced the highest number of significant results closely followed by the three-phase SST system. The nine-phase SST system produced the least number of significant results for all performance measures.

Clearly the type of performance measure will influence the decision regarding the optimal climate forecasting approach, and so it is important to consider the fundamental characteristics of the measures when making this judgement. We refer the reader to Potgieter *et al.* (2003) and Maia *et al.* (2007) for a more detailed discussion about this topic.

5. CONCLUDING REMARKS

This paper outlined a process for assessing the climate-forecasting capability of different systems for a sugarcane-growing industry that spans the north eastern Australian coast. An overall assessment was performed for seven sugarcane growing regions across twelve three-monthly calendar periods. The performance of the benchmark five-phase SOI system could be considered sound if the performance measure was based on the KW or LEPS measures, but not if the decision was based on ROC performance measures. The three-phase SST system performed well across all performance measures whilst the nine-phase SST system performed least favourably. If it is important to perform well across a range of performance measures then this would raise the obvious question as to if the three-phase SST system might be more appropriate for the Australian sugar industry. Clearly, the advantages and disadvantages of changing forecasting systems would need to be carefully considered.

Whilst this paper has recognised problematic issues that arise when comparing different forecasting systems based on a different number of phases, the challenge to identify a suitable performance measure is dependent on application. We have considered three performance measures which have been used for illustrative purposes. There exists a wide range of alternative performance measures in the literature. It is also worth noting that the number of significant fields has a sampling variability associated with it and confidence limits could be superimposed to reflect this variability.

In this paper the field has been defined on the basis of latitudinal position and calendar period. We have not considered a longitudinal component because longitudinal variation of the locations in this study is relatively small compared to latitudinal variation. A field might more commonly consist of an arrangement of latitudinal and longitudinal coordinates. The approach outlined in this paper can easily cater for this modification as well as extensions to higher dimensional grids defined by a combination of latitude, longitude and calendar period as an example. The approach in this paper however is not suitable for assessing forecasting capability for targetted decisions which are specific to location, response and lead-time. In such circumstances we refer the reader to Maia *et al.* (2007) where the magnitude of the P-value can provide a mechanism for making this judgement. Caution should however be applied when testing highly specific circumstances so as to avoid pitfalls associated with so-called "peak-picking" that can be caused by jumping from one forecasting system to another. To contrast, in some instances, it can be appropriate to identify a statistically sound all year round industry wide forecasting system to facilitate large scale communication processes. With this goal in mind, a methodology to achieve this task was presented.

We conclude by emphasizing that the comparisons performed in this paper have been based on identical time periods. This is important because owing to decadal and multidecadal climate variability, climate forecasting performance measures can be influenced by low frequency climate oscillations such as the interdecadal Pacific oscillation. We refer the reader to Power *et al.* (1999) for more details on this topic.

ACKNOWLEDGEMENTS

This research was sponsored by the Managing Climate Variability Program, the Sugar Research and Development Corporation and James Cook University. Kathryn Jones is thanked for her contribution toward this research.

Table 2. Resultant P-values for the KW procedure for the field composed of 7 regions and 12 temporal periods. Phases were generated by the five-phase SOI system.

	CNS	MLN	LUC	TSV	MKY	BUN	NSW
JFM	**0.09**	**0.00**	0.25	**0.04**	0.17	**0.05**	**0.07**
FMA	0.41	0.51	0.70	0.44	0.62	0.72	0.72
MAM	0.49	0.68	0.86	**0.05**	0.26	0.64	**0.02**
AMJ	0.29	**0.10**	0.32	0.87	0.38	0.12	**0.09**
MJJ	0.26	0.17	0.58	0.96	0.82	0.61	**0.07**
JJA	0.18	0.22	0.85	0.20	0.74	0.18	**0.10**
JAS	0.45	0.54	**0.01**	**0.00**	**0.02**	**0.00**	0.76
ASO	**0.01**	**0.01**	**0.00**	**0.01**	**0.00**	**0.07**	0.67
SON	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	**0.05**	**0.07**
OND	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.03**	0.16
NDJ	**0.01**	**0.01**	**0.01**	**0.00**	**0.00**	**0.00**	**0.02**
DJF	**0.01**	**0.00**	**0.00**	**0.04**	**0.05**	**0.03**	**0.06**

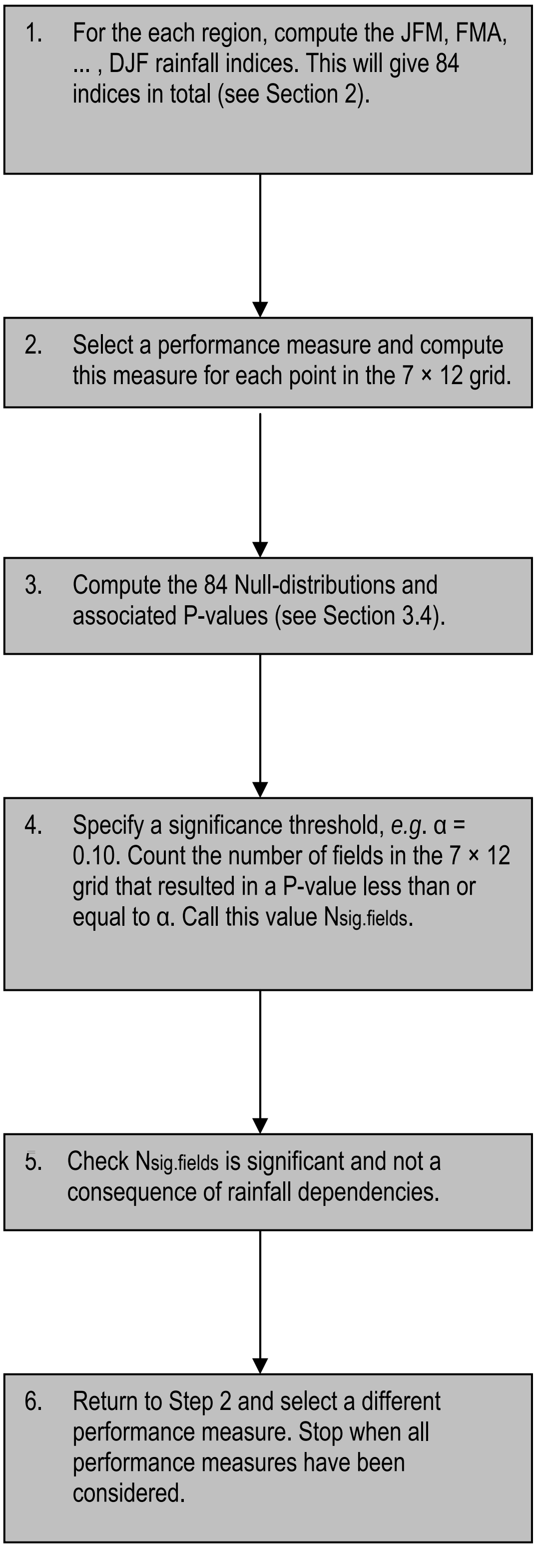
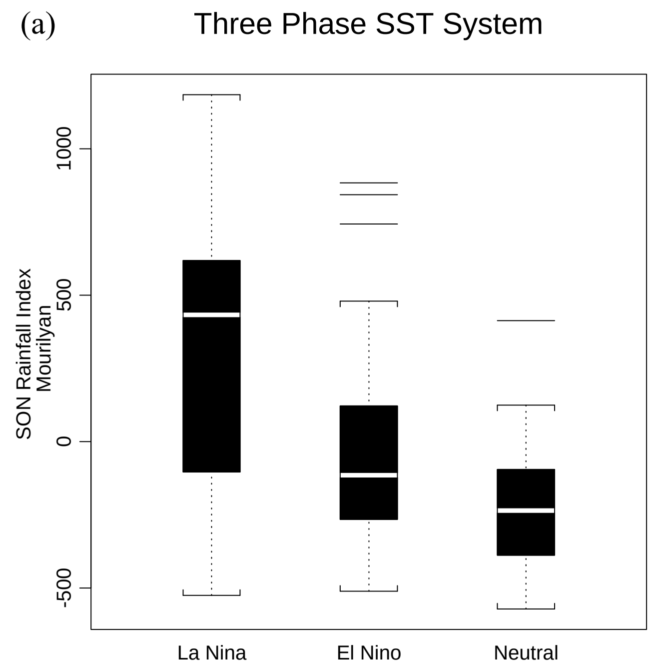


Figure 1. Flowchart of research methodology applied to each forecasting system.

Table 3. The number of significant fields in the 7x12 grid and the associated global P-value for field significance.

Measure	Forecast system	Number of Significant Fields	Global P-value
KW	3-Phase SOI	38	0.000
	3-Phase SST	43	0.000
	5-Phase SOI	48	0.000
	9-Phase SST	30	0.000
LEPS	3-Phase SOI	33	0.000
	3-Phase SST	35	0.000
	5-Phase SOI	33	0.000
	9-Phase SST	28	0.000
ROCS (above)	3-Phase SOI	23	0.001
	3-Phase SST	35	0.000
	5-Phase SOI	24	0.006
	9-Phase SST	22	0.000
ROCS (below)	3-Phase SOI	24	0.000
	3-Phase SST	33	0.000
	5-Phase SOI	15	0.038
	9-Phase SST	9	0.002

(a)



(b)

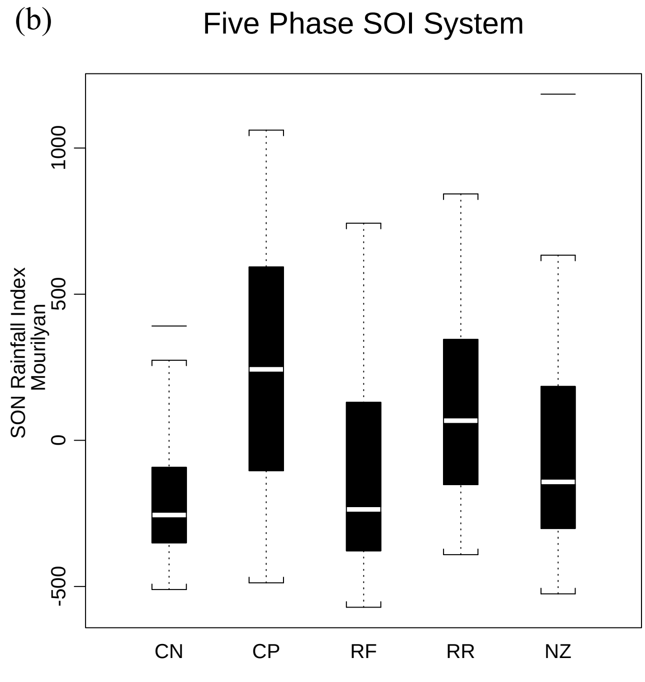


Figure 2. Shifts in distributions for the September to November rainfall principal component index at Mourilyan based on (a) SST phases and (b) SOI phases at the end of August.

6. REFERENCES

Conover, W.J. (1980), Practical Nonparametric Statistics. 3rd edn. John Wiley and Sons.

Drosdowsky, W. (2002), SST phases and Australian Rainfall. *Australian Meteorological Magazine*, **51**(1), 1-12.

Everingham, Y.L., R.C. Muchow, R.C. Stone, N.G. Inman-Bamber, A. Singles and C.N. Bezuidenhout (2002), Enhanced risk management and decision-making capability across the sugar industry value chain based on seasonal climate forecasts. *Agricultural Systems*, **74**, 459-477.

Good P.I. (1997), Permutation tests: a practical guide to resampling methods for testing hypotheses 3rd edition. Springer-Verlag New York Inc: New York.

Jones, K.E., and Y.L. Everingham (2005), Can ENSO combined with low-frequency SST signals enhance or suppress rainfall in Australian sugar-growing regions? *Proceedings of the International Congress on Modelling and Simulation*, December, Melbourne, Victoria, December 2005 pp. 1660-1666.

Maia, A. H. N., H. Meinke, S. Lennox and R. Stone (2007), Inferential, nonparametric statistics to assess the quality of probabilistic forecast systems. *Monthly Weather Review*, **135**(2), 351-362.

Mason, S.J., and N.E. Graham (1999), Conditional probabilities, relative operating characteristics, and relative operating levels. *Weather and Forecasting*, **14**, 713-725.

Murphy, A.H. (1993), What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, **8**, 281–293.

Potgieter, A.B., Y.L. Everingham and G.L. Hammer (2003), On measuring quality of a probabilistic commodity forecast for a system that incorporates seasonal climate forecasts. *International Journal of Climatology*, **23**, 1195-1210.

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. *Journal of Climate*, **9**, 34-53.

Power, S., T. Casey, C. Folland, A. Colman, V. Mehta (1999), Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics,* **15**, 319-324.

Smith, T.M., and R.W. Reynolds (2003), Extended reconstruction of global sea surface temperatures based on COADS data (1854-1997). *Journal of Climate*, **16**, 1495-1510.

Stone, R.C. and A. Auliciems (1992), SOI phase relationships with rainfall in eastern Australia. *International Journal of Climatology*, **12**, 625-636.

Stone, R.C., G. Hammer and T. Marcussen (1996), Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. *Nature*, **384**, 252-255.

Stone, R.C., I. Smith, and P. McIntosh (2000), Statistical methods for deriving seasonal climate forecasts from GCM's. Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems, G. L. Hammer, N. Nichols, and C. Mitchell, Eds., Kluwer Academic, 135-147.

Wilks, D.S. (1995), Statistical Methods in the Atmospheric Sciences. Academic Press, London.