Aust. Met. Mag. 44 (1995) 219–224

Optimal linear combination of seasonal forecasts

Trevor Casey
National Climate Centre, Bureau of Meteorology, Australia
(Manuscript received February 1995; revised May 1995)

This paper illustrates the improvement in skill obtained in seasonal rainfall outlooks for the Australian region from the optimal combination of partially independent statistical prediction schemes. The linear combination technique was applied to two Southern Oscillation Index (SOI) based seasonal rainfall forecast techniques currently used in the National Climate Centre; a linear discriminant analysis (LDA) technique and a SOI phase analogue technique. Optimal weighting of the individual probability predictions of each prediction scheme is achieved using a least-squares variational reduction of the forecast errors with respect to the weighting factors. The combination of the two prediction schemes shows a significant improvement in forecast skill over a twenty-year evaluation period from 1972 to 1991. Mean square errors (half-Brier scores) of 0.213 for the SOI phase technique and 0.221 for the LDA model were reduced to a half-Brier score of 0.197 for the combination over the evaluation period.

Introduction

There is an increasing recognition that consensus probability forecasts, produced by the optimal combination of separate prediction techniques, provide, on average, a higher forecast skill over the longer term than the skills of the individual prediction schemes which constitute the consensus. The improvement in skill occurs for two reasons. First, when the individual predictions are not perfectly correlated, one scheme may be a better predictor for some locations or time of year and contain predictive skill that is not contained in the other. Similarly, the other scheme may be a better predictor for other times and locations, and the combination therefore yields better skill when averaged overall. Secondly, the individual predictions are weighted according to their relative skill, and are given more weight when and where they display greater skill.

The use of the optimal combination technique has long been advocated in econometric forecasting and in engineering applications, but has been slow in finding acceptance in statistical meteorological forecasting. The reason for this appears to be that statistical estimates of meteorological parameters in short and long-term forecasting have been largely overshadowed in recent years by a reliance on numerical prediction models. There is a growing body of evidence, however, which indicates that stochastic models can achieve skill comparable to that of dynamical numerical models or provide improved skill when used in combination with numerical model output, particularly for climatological time-scale forecasts. For example, Passi (1975) used a combination of unbiased estimates between correlated variables to yield a better single estimate of observational meteorological parameters, and Thompson (1977) showed the improvement in accuracy obtained by the combination of independent (imperfectly correlated) forecasts. More recently, Fraedrich and Leslie (1987) have extended the use of linear combinations to probabilistic prediction variables and demonstrated the improvement to both short and long-term forecasts when used with numerical weather prediction model output statistics and stochastic Markov chains.

Corresponding author address: Mr T. Casey, National Climate Centre, Bureau of Meteorology, GPO Box 1289K, Melbourne, Vic 3001, Australia.

In this paper, it is first shown that a substantial improvement in skill in a simple climatological forecast of seasonal rainfall probabilities is obtained for an optimal linear combination of climatology and persistence. The method is then applied to the combination of two separate statistical techniques used operationally in the National Climate Centre for producing probabilistic seasonal rainfall outlooks. The first of these is a SOI phase analogue technique which composites the frequency distribution of (three month) seasonal rainfall commencing in the month following defined phases of the low-pass filtered historical SOI record (Zhang and Casey 1992). The second is a linear discriminant analysis (LDA) technique, similar to that described by Ward and Folland (1991), which is applied to categorical rainfall distributions associated with variations of the SOI.

Finally, the derived probabilities from the combination of the two seasonal rainfall prediction techniques were further optimally combined with persistence probability values obtained from the rainfall record from 1913 to 1971. Even though the predictive probability skill of the persistence forecast does not have the same level of skill as the other predictive schemes, a further marginal increase in skill was nevertheless obtained for the multiple combination over a 20-year test period from 1972 to 1991. The results indicate that the successive combination of individual forecast techniques, where each is at least partially independent of the other techniques, results in a better consensus forecast.

Method

The minimisation of errors by least squares variational analysis, first developed by Gauss (1809), has found widespread use in many fields including meteorology. Gauss's theorem and Bayesian probability considerations provide a theoretical basis for the optimal combination of independent probabilistic prediction schemes to produce consensus forecasts with greater average skill than that of any of the individual schemes which make up the consensus. Details of the theoretical development of the optimal linear combination technique appear in Passi (1975), Thompson (1977) and Fraedrich and Leslie (1987). A summary of the essential formulae used in this paper is presented in the Appendix.

Combination of climatology and persistence

The technique is first applied to the combination of two fundamental climatological predictors, persistence and climatology, to demonstrate the increase in overall skill achieved by their optimal linear combination. Three-month total rainfall for all Australian meteorological districts was divided into three categories; above average, average and below average, based on the 30 and 70 percentile values. Here, the climatological forecast probability for a given three-month period in a given category is simply taken as the time-averaged binary count of the observed occurrences in each category over the period of the historical record 1913 to 1993.

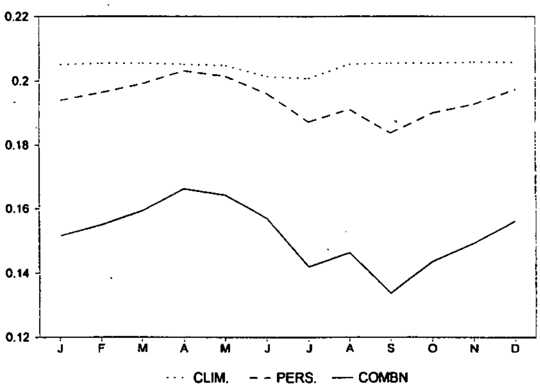
The persistence probabilities of three-month rainfall in each rainfall category for each month of the year were similarly calculated as the time-average of the binary count of the number of occasions when the subsequent three-month rainfall occurred in the same category as in the preceding three months. Weights for the optimal combination were calculated as in Eqn A9 and half-Brier scores as in Eqn A10, over the period from 1913 to 1993.

Figure 1 shows a plot of the mean square errors for each month of the year, for the following three-month forecast produced by climatology, persistence and the optimal combination of these two. These scores have been averaged for all rainfall categories, and averaged over all Australian districts. Averaged for all months in the year, half-Brier scores for climatology were 0.204, and for persistence, 0.194. Half-Brier scores for the combination average at 0.152 which is about a 25 per cent increase in skill over pure climatological expectation alone.

The skill of the empirically derived climatological probability is generally consistent all year round and shows little variation in skill level when averaged over the whole country, whereas persistence scores show greater skill in the latter half of the calendar year for forecasts commencing in the months from June to November. Individual category scores in the above, below and average rainfall categories show a similar relationship.

It could be argued that the evaluation has not been performed on independent data, and that the climatological and persistence values are only known *a posteriori*. However, a basic assumption has been made in this work, that over a sufficiently long period of time, the discrete sample values of the means, variances and covariances of both the climatological and persistence time-series converge towards infinite population statistics and are stationary. For periods of around thirty years or longer, sample statistics approach reasonably close to longer term equilibrium

Fig. 1 Mean square errors (half-Brier scores) for the combination of climatology and persistence forecast probabilities averaged for the three categories of below average, average and above average rainfall for the period 1913 to 1993. The scores shown in each month are the average over all Australian rainfall districts.



values. It is considered, therefore, that the use of the entire record to assess skill levels is appropriate in this instance.

Combination of SOI phase analogue and linear discriminant analysis predictions

The SOI phase analogue method provides the probabilities of three-month rainfall in the categories of above, below and about average rainfall. Rainfall categories for each Australian rainfall district are defined by the 30 and 70 percentile rainfall values for every three-month period through the year. The method is based on the historical frequency of the rainfall in each category, commencing in the month following defined phases in the low-pass filtered SOI record. Probabilities for the three-month outlook period are derived from the identified phase of the moving average SOI in the preceding month. Probabilities in each category are expressed as a fraction of the total probability space of one for all categories.

The linear discriminant analysis technique also extracts the different probabilities in each rainfall category for three-month total rainfall and, in the form employed in this work, uses the three-month moving average SOI calculated in the preceding month as predictor. This technique utilises the shift in the relationship between the SOI and the assumed normal rainfall probability distribution functions in each of the different categories of rainfall to determine the respective probabilities. These are then also expressed as fractions of the pooled category probability space of one.

Individual probability predictions from both models were made each month over an independent evaluation period from 1972 to 1991 for each category of below average, average and above average rainfall. Optimum weights for the combination were calculated as in Eqn A4 using the training period ensemble averages over the period from 1913 to 1971. Mean square prediction errors for each scheme were calculated as in Eqn A6 using the observed values over the evaluation period 1972 to 1991. Mean square errors of the optimum combination were similarly calculated over the evaluation period as in Eqn A5.

Figure 2(a) shows a plot of the individual prediction scheme half-Brier scores and a plot of the half-Brier scores of the optimal combination, averaged over all categories and all rainfall districts for the independent evaluation period from 1972 to 1991. In this example, the skill scores in each rainfall district have been weighted by the relative area of the district expressed as a proportion of the total area of Australia. This was done to reduce the bias that might be introduced into the composite average by an increased density of observations in regions which show greater skill.

Both prediction schemes are comparable in skill over the evaluation period, having half-Brier scores of 0.221 for the LDA method and 0.213 for the phase method averaged for all districts over all months of the year. Mean square errors for the combination average 0.197 for all categories, about a ten per cent improvement in skill, which illustrates the effectiveness of the linear combination technique.

Skill scores for the eastern States only, also spatially weighted by the relative areas of the rainfall districts as a proportion of the total area of the eastern States, are plotted in Fig.2(b). These show more than a doubling of the skill found for Australia as a whole, and it is immediately apparent that both the phase analogue and LDA predictive schemes show a much greater skill over eastern Australia than for the western and central States.

Both predictive schemes and the combination all show reduced skill during the austral autumn months from April to June. It can be seen from Figs 2(b) and 2(c), that the LDA method, in its realisation here using the moving average SOI, has a broader spread of reduced skill across the autumn period. The LDA method also has a wider variability in its prediction skill in some rainfall categories, as can be seen in Fig. 2(c), which shows a plot of the respective half-Brier scores for the category of below average rainfall. The more vari-

Fig. 2(a) Mean square errors for the combination of two SOI-based prediction schemes, a phase analogue and a linear discriminant analysis method, averaged for the three categories of below average, average and above average rainfall for the period 1972 to 1991. The scores shown in each month are the average over all Australian rainfall districts weighted by the relative area of each district to the total area of Australia.

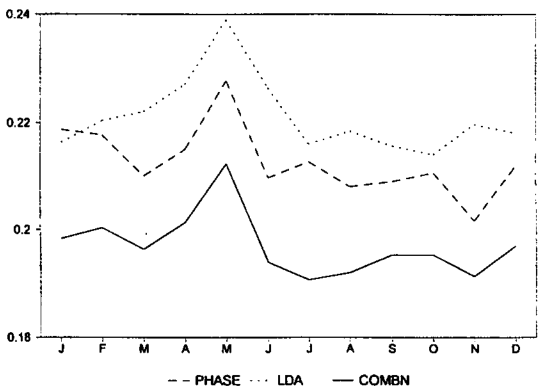


Fig. 2(b) As in Fig. 2(a), but for eastern Australian States only. Scores are weighted by relative area of district to total area of eastern Australia.

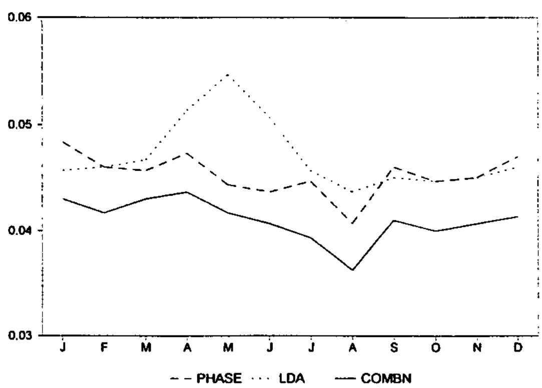
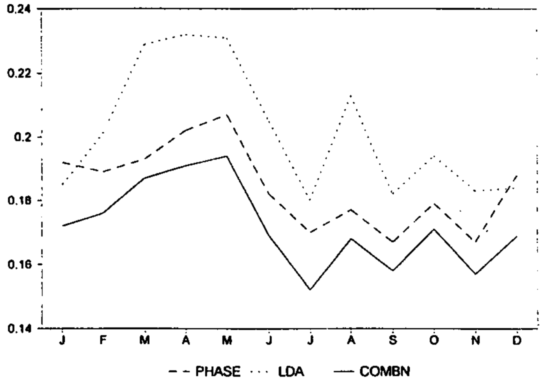


Fig. 2(c) As in Fig. 2(a), but for the category of below average rainfall only.



able error scores of the LDA model appear to be due to the non-normal distribution of the district rainfall data, particularly the small numbers of very high rainfalls in the above average category. To give the LDA method due credit, however, in practice the model is used operationally with other predictors, additional to the simple three-month moving average SOI. First differences, or the trend in the moving average SOI, is also used, with some increase in overall skill, but is not shown here for the sake of brevity. The main purpose of the comparison is not to evaluate the individual prediction schemes, but to demonstrate the effectiveness of the linear combination technique.

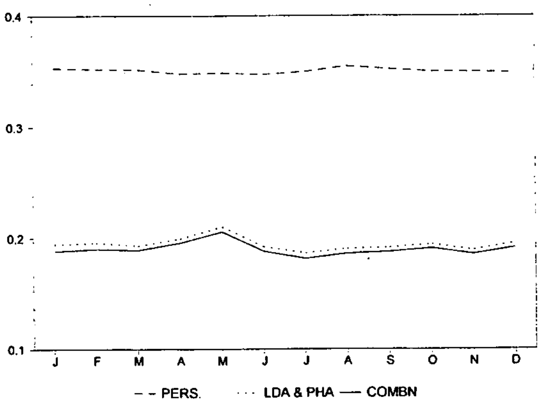
Multiple combinations

The consensus forecast obtained from the combination of the SOI phase analogue and LDA prediction techniques was further combined with persistence probability values, calculated from the training period from 1913 to 1971, to investigate the effect of multiple combinations on the skill levels. Persistence probability forecasts do not have skill comparable to the skill shown by either of the SOI-based predictive schemes, and it would be reasonably expected that the inclusion of the persistence probabilities should provide little further gain in skill. The question is, does the inclusion of a scheme with lower overall skill degrade the skill of the combined consensus forecast?

The combined SOI phase analogue and LDA prediction probabilities were calculated for the evaluation period from 1972 to 1991 based on the training set from 1913 to 1971 as before. Probability predictions obtained from persistence values calculated over the same training data from 1913 to 1917 were also produced for the test period 1972 to 1991. Optimum weights for the linear combination of the persistence predictions and the combined SOI schemes were calculated as in Eqn A4. Probability predictions from the multiple combination were then scored against the observations over the evaluation period.

Figure 3 shows plots of the half-Brier scores for the combination of the LDA and phase analogue SOI prediction schemes, for the persistence forecasts and for the multiple combination of the combined SOI schemes and persistence over the evaluation period. The half-Brier scores for the additional combination of the persistence forecast with the combined SOI prediction schemes show a marginal incrèase in skill over the combined SOI schemes alone. Importantly, no reduction in skill occurs as a result of combining a scheme of lower overall skill with an existing

Fig. 3 Mean square errors for successive combination of persistence probabilities for period 1913 to 1971 with combination of two SOI-based prediction schemes for test period 1972–1991. Averaged in each month for all districts and all categories. Scores are weighted by relative district areas as before.



scheme. This is because the weightings accorded to the scheme with lower skill are very small except in cases (i.e. locations or times) where the other predictive schemes also have a lower skill.

Discussion

The increase in skill shown from the optimal linear combination of partially independent predictive schemes is considerable and commends the use of consensus forecasts where the individual prediction schemes have something to contribute to the overall skill. The inclusion of additional predictive schemes in successive combination with existing predictive techniques does not produce a reduction in skill, even if the scheme to be included has lower general skill than the original consensus. The inclusion of a scheme with lower overall skill in the consensus combination may, however, provide baseline skill value at certain locations or times when there is little signal in the other predictive schemes.

Very few assumptions have been made about the statistical nature of the forecast system other than that the errors are normally distributed. The dynamics and form of the statistical relationship between, say, the SOI predictors and rainfall in Australia are particular to the individual predictive schemes and are immaterial to the combination technique. It is implicit, however, that the variance regression is of a Gauss-Markov type, and it is clear from Passi (1975) that it is desirable to use as long a period as possible to calculate optimum weights. The assumption of normally distributed error variance is premised on the basis that discrete realisations approach the continuous population statistics for large sample size. This is especially true for binary or delta function representations of the observational data. Shorter term realisations may be subject to oscillations in the auto-correlation functions and may therefore be unrepresentative of the assumed normal probability distributions. Sample sizes should be a minimum of around 30 years for monthly or seasonal data. In practical application, it is suggested that optimum weights should be calculated on the entire available record, unless there is good evidence to indicate that a definite shift in the statistical distributions has occurred.

The optimal linear combination of probabilistic forecasts, when used for a number of separate prediction techniques, is in essence a deconstructed form of the more generalised canonical correlation analysis (CCA) method which utilises the major eigenvalues of a multiple predictor covariance matrix to maximise the relationship between predictor variables. The optimal linear combination technique offers an effective and simpler alternative to CCA as a consensus forecasting method using multiple predictive schemes. With canonical correlation, it is often difficult to provide a physical interpretation of the maximum contributions to the skill, and the technique also suffers from potential invertibility and instability problems of the covariance matrix. Contributions to the overall skill by the individual predictive schemes are readily seen with the linear combination technique and computation is relatively easy.

Acknowledgments

The author wishes to express thanks to Professor Klaus Fraedrich for originally bringing the optimal linear combination technique to his notice, and for his helpful suggestions and review of an earlier draft. Thanks also to Bob Seaman for his review and comments which improved the manuscript, and to Mr Paul Nydam, who kindly provided the LDA model probability test data.

Appendix

Theory of optimal linear combination of independent probability forecasts

Consider two prediction variables ϕ_1 and ϕ_2, denoting the probability of an event in a given category given by two independent (imperfectly correlated) prediction schemes. ϕ_1 and ϕ_2 may be probabilistic variables ($0 \leqslant \phi_i \leqslant 1$), or binary ($\phi_i = 1$ or 0).

From Bayesian considerations, for the linear combination

$$\phi_* = a\phi_1 + b\phi_2 \qquad \ldots \text{A1}$$

to occupy the same categorical probability space as the individual probability predictions requires that $a + b = 1$.

If the observed value at each time-step in a series of forecasts is denoted by the binary variable δ ($\delta = 1$ or 0), the mean square error (half-Brier score) of the deviations of the predictions from observations for the time-average of the ensemble forecast is given by

$$\begin{aligned} B^* &= \langle(\delta - \phi^*)^2\rangle \\ &= \langle\delta^2\rangle - 2\langle\delta\phi_2\rangle + \langle\phi_2{}^2\rangle + 2a(\langle\delta\phi_2\rangle - \\ &\quad \langle\phi_2{}^2\rangle - \langle\delta\phi_1\rangle + \langle\phi_1\phi_2\rangle) \\ &\quad + a^2(\langle\phi_1{}^2\rangle + \langle\phi_2{}^2\rangle - 2\langle\phi_1\phi_2\rangle) \qquad \ldots \text{A2} \end{aligned}$$

where the angle brackets denote the ensemble time average and noting that $b = 1 - a$.

Gauss's least square method of variational analysis with respect to the weights a and b minimises the bias and error variance of the combination probability forecast.

$$\begin{aligned} \frac{dB^*}{da} &= \langle\delta\phi_2\rangle - \langle\delta\phi_1\rangle - \langle\phi_2{}^2\rangle + \langle\phi_1\phi_2\rangle \\ &\quad + a(\langle\phi_1{}^2\rangle + \langle\phi_2{}^2\rangle - 2\langle\phi_1\phi_2\rangle) = 0 \end{aligned} \qquad \ldots \text{A3}$$

which solving for the weight a gives

$$a = \frac{\langle\delta\phi_1\rangle - \langle\delta\phi_2\rangle + \langle\phi_2{}^2\rangle - \langle\phi_1\phi_2\rangle}{\langle\phi_1{}^2\rangle + \langle\phi_2{}^2\rangle - 2\langle\phi_1\phi_2\rangle} \qquad \ldots \text{A4}$$

The mean square error of the optimum combination ϕ^* is then

$$B^* = \langle(\delta - \phi_2)^2\rangle - \frac{(\langle\delta\phi_1\rangle - \langle\delta\phi_2\rangle + \langle\phi_2{}^2\rangle - \langle\phi_1\phi_2\rangle)^2}{\langle\phi_1{}^2\rangle + \langle\phi_2{}^2\rangle - 2\langle\phi_1\phi_2\rangle} \qquad \ldots \text{A5}$$

It can be seen that the whole expression on the RHS is smaller than the first term $\langle(\delta - \phi_2)^2\rangle$, which is the mean square error of the prediction scheme for θ_2. Similarly, it can be shown for minimisation with respect to the weight, b, that the optimum combination has a smaller mean square error than the prediction scheme for ϕ_1. Mean square prediction errors for each individual prediction scheme are given by

$$B_i = (\delta - \phi_i)^2\rangle = \langle\delta^2\rangle - 2\langle\delta\phi_i\rangle + \langle\phi_i{}^2\rangle \qquad \ldots \text{A6}$$

It is noted here that, for binary observation data, $\delta = 1$ or 0, the mean and second moments are equal, i.e.

$$\begin{aligned} \phi_1 = \langle\delta\rangle = \langle\delta^2\rangle = \phi_c \\ \langle\phi_1{}^2\rangle = \phi_c{}^2 \end{aligned} \qquad \ldots \text{A7}$$

where ϕ_c is the climatological probability of rainfall occurring in the given category.

$$\phi_{c2} = \langle\phi_2\rangle = \langle\phi_2{}^2\rangle$$

and

$$\langle\delta\phi_2\rangle = \phi_{c2}\langle\delta\rangle = \phi_c\phi_{c2} \qquad \ldots \text{A8}$$

The optimum linear combination $\phi^* = a\phi_1 + (1 - a)\phi_2$ can be found by substituting the corresponding autovariance and covariance terms from A7 and A8 into A4 which yields for the weight a,

$$a = \frac{\phi_c{}^2 + \phi_{c2} - 2\phi_c\phi_{c2}}{\delta_c + \phi_{c2} - 2\phi_c\phi_{c2}} \qquad \ldots \text{A9}$$

The mean square error of the optimum combination is

$$B^* = B_2 - \frac{(\phi_c{}^2 + \phi_{c2} - 2\phi_c\phi_{c2})^2}{\phi_c + \phi_{c2} - 2\phi_c\phi_{c2}} \qquad \ldots \text{A10}$$

In the case where both predictive schemes are statistically independent, meaning they are totally uncorrelated, i.e. $\langle\phi_1\phi_2\rangle = 0$, and where the optimum estimate is unbiased, it can be shown (e.g. TASC, 1974, pp. 5–6) that Eqn 4 reduces to

$$a = \frac{\sigma_2{}^2}{\sigma_1{}^2 + \sigma_2{}^2}$$

where σ_1 and σ_2 are the error variances of each scheme. Similarly,

$$b = \frac{\sigma_1{}^2}{\sigma_1{}^2 + \sigma_2{}^2}$$

so that the weights are then assigned according to the relative weights of the respective error variances. In other words, each scheme is weighted according to its skill relative to the other. For predictive schemes that are partially or perfectly correlated, the ratio of the weighting is reduced proportionately by the degree of correlation.

References

Fraedrich, K.. and Leslie, L.M. 1987. Combining predictive schemes in short-term forecasting. *Mon. Weath. Rev., 115*, 1640–4.

Gauss, C.F. 1809. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium.* Perthes and Besser, Hamburg.

Passi, R.M. 1975. Statistical estimation of meteorological parameters with correlated observations. *Mon. Weath. Rev., 103*, 521–7.

The Analytical Sciences Corporation 1974. In *Applied Optimal Estimation.* Ed. A. Gelb, The MIT Press, Cambridge, 374 pp.

Thompson, P.D. 1977. How to improve accuracy by combining independent forecasts. *Mon. Weath. Rev., 105*, 228–9.

Ward, M.N. and Folland, C.K. 1991. Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. *Int. J. Clim., 11*, 711–43.

Zhang, X.-G. and Casey, T.M. 1992. Long-term variations in the Southern Oscillation and relationships with Australian rainfall. *Aust. Met. Mag., 40*, 211–25.