Aust. Met. Mag. 30 (1982) 201-203 551.509.314

Statistical correction of thickness height prognoses

A. F. Bennett* and L. M. Leslie†

(Manuscript received December 1981; revised January 1982)

The S_1 skill scores of thickness height prognoses produced by the Australian Region Primitive Equation Model can be reduced by a statistical correction scheme. In addition to having lower skill scores, the statistically-corrected model prognoses are also subjectively better than the uncorrected model prognoses. The scheme therefore has considerable potential value for the Bureau of Meteorology's numerical prognosis system, as is illustrated by its performance in two case studies, one of which is an unusually strong cold outbreak over southeastern Australia.

Introduction

Two statistical studies of the errors in Australian region 24-hour numerical weather prediction models found that the prognosis errors were correlated with the prognoses themselves. Accordingly, it proved possible to devise simple linear schemes that predict a fraction of the errors, using the prognoses as predictors. The first study (Bennett and Leslie 1978) examined the mean squared errors in six years of daily prognoses of 500 mb geopotential height (ϕ_{500}), produced by a barotropic filtered model. It was found that the mean squared error could be reduced by about one-third, using a statistical correction scheme. The second study (Bennett and Leslie 1981) examined the S_1 skill score (Teweles and Wobus 1954) in six years of daily prognoses of mean sea level pressure (P_{MSL}), produced by the Australian Region Primitive Equation (ARPE) model. It was found that the S_1 skill score could be reduced by about 3.5 points, using a statistical correction scheme. Earlier attempts to correct prognoses statistically (for example Rinne 1971a, 1971b, 1972) produced corrected prognoses that were unrealistically smooth. Such was not the case in the two Australian region studies. The reason for this crucial improvement is that the **error** field was predicted, rather than the full field. Statistically predicted fields of either kind are invariably smooth, but a field comprised of a dynamical prognosis plus a statistical correction contains the small-scale structure predicted by the dynamics alone. As a result the Australian region corrected prognoses were subjectively, as well as objectively, better than the uncorrected prognoses.

This article is a brief report of the results of a study of the S_1 skill score of six years (1972-1977) of daily prognoses of 1000 mb to 500 mb thickness heights ($\Delta\phi \equiv \phi_{500} - \phi_{1000}$), produced by the ARPE model. It was found that the S_1 skill score could be reduced by about 3.0 points, using a statistical correction scheme. The method used in this third study was identical to that used in the study of P_{MSL} errors, and the reader is referred to Bennett and Leslie (1981) for all details. The results obtained in this study were very similar to those obtained in the earlier study, so only a representative selection is included here, as we are concerned in this article mainly with indicating the possibilities of the scheme as part of the Bureau of Meteorology's operational numerical forecast system.

*Department of Mathematics, Monash University, Clayton.
†Australian Numerical Meteorology Research Centre, Melbourne.

Results

Comparisons of ϕ^M and $\Delta\phi^{M+Sc}$

Let $\Delta\phi^M$ denote the ARPE model prognosis of $\Delta\phi$, and let $\Delta\phi^{M+Sc}$ denote the statistically-corrected model prognosis. A comparison of daily values of the S_1 skill score for $\Delta\phi^M$ and $\Delta\phi^{M+Sc}$, for the month of July 1977, is shown in Fig. 1. The monthly means were, respectively, 35.4 and 33.1 points. The reduction of 2.3 points was comprised of 0.5 points due to simple elimination of the seasonal mean error field, plus 1.8 points due to regression of the residual error field against the prognosis. It may be noted that July is in general a very good month for the ARPE model, hence the relatively modest reduction of 2.3 points. The mean reduction for the winter season of June, July and August (JJA) in all six years was 3.0 points, while the mean reduction for the summer season of December, January and February (DJF) in all six years was 2.8 points.

Monthly means for all six years of S_1 skill scores for $\Delta\phi^M$ and $\Delta\phi^{M+Sc}$ are shown in Fig. 2, along with scores for climatological forecasts (in which the 24-hour prognoses are all given by the six-year mean analysis for the month), persistence forecasts (in which the 24-hour prognoses are given by the initial analyses), and for linear predictions. The last forecasts were derived by a purely statistical procedure, in which the analysis fields were used as predictors of the full fields expected 24 hours later. Manual forecasts of $\Delta\phi$ have only been made infrequently by the Bureau of Meteorology, and no manual skill scores were available to us. However, the studies of ϕ_{500} and P_{MSL} showed that manual skill was much the same as the skill of linear predictors. Thus it is clear from Fig. 2 that statistical correction roughly doubles the impact of model prognoses on a manual forecasting system.

Fig. 1 Daily comparison of S_1 (0) for the model prognosis with S_1 (•) for the statistically-corrected model prognosis, during July 1977. The monthly mean values are 35.4 and 33.1 respectively.

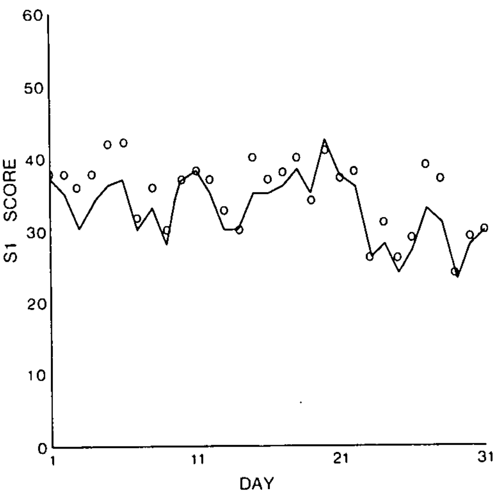
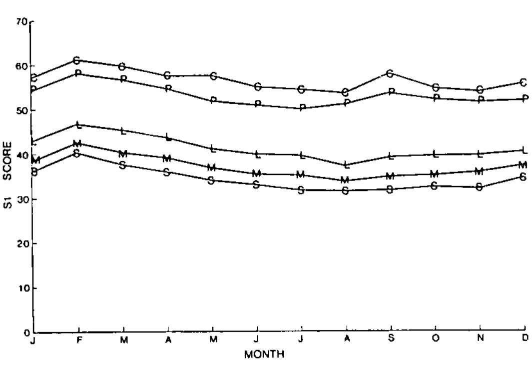


Fig. 2 Monthly mean S_1 skill score for: climatology (C), persistence (P), linear prediction (L), model (M), and model plus statistical corrections (S).



Reliability tests

The results described above were obtained by hindcasts, that is, the statistically-corrected model prognoses were compared with verification analyses included in the six-year sample that had already been used to estimate covariances. To test the reliability of the estimates, the sample was partitioned into two subsets. Subset 1 included the years 1972, 1974 and 1975, while Subset 2 included the years 1973, 1975 and 1977. The statistical correction schemes were then re-derived for each subset, thus using only 270 days of data per season rather than the full 540 days. Results are shown in Table 1, for both DJF and JJA. The first column (M) is the six-year S_1 skill score for $\Delta\phi^{M}$. The second (H) is the six-year score for $\Delta\phi^{M+Sc}$. The next four columns (M_1, H_1, M_2 and H_2) are the three-year scores for the corresponding quantities and subsets. The final column (F_{12}) is a three-year score for Δ^{M+Sc}, in which the covariances were estimated using Subset 1 but the scores were calculated using Subset 2. That is, F_{12} is the score after a genuine forecast experiment. As expected the three-year forecast skill reduction ($M_2 - F_{12} = 2.0$) was less than the six-year hindcast score skill reduction ($M-H = 2.8$) and also less than the three-year hindcast skill reduction ($M_1 - H_1 = 2.4$).

A probabilistic estimate of forecast skill reduction using covariances estimated from all six years of data may be made using the formula due to Lorenz (1977) or Davis (1976). With sufficient accuracy the formula is

$$F^2 = H^2 + M^2 J \left(\frac{\tau}{N}\right),$$

where F is the expected forecast skill, J is the number of independent predictors (in this case the number of empirical orthogonal function amplitudes used to represent the prognoses), τ is the decorrelation time scale of the thickness height field, while N is the number of days in the sample (hence (τ/N) is the number of independent events). Here $J=8, \tau \simeq 2$ and $N = 540$, so the expected $F = 38.2$. That is, the forecast skill reduction $(M-F)$ would probably be less than the hindcast skill reduction $(M-H)$ by about 0.6 points.

In conclusion, the overall score reduction for thickness height prognoses, by statistical correction, was about a point less than that for mean sea leavel pressures. We attribute this to the generally better performance of the ARPE models at higher levels in the atmosphere. Hence there was less bias to eliminate by statistical correction. Nevertheless, the reductions are of operational significance.

Table 1. S_1 Skill score for $\Delta\phi_{500-1000}$

	M	H	M_1	H_1	M_2	H_2	F_{12}
DJF	40.4	37.6	39.8	37.4	40.1	37.3	38.1
JJA	36.2	33.2	35.5	33.0	35.4	32.7	33.5

Case studies

In order to make subjective assessments of statistically-corrected prognoses, we carried out a large number of case studies. Two of these will now be discussed. The first was three consecutive days chosen at random from the period of the first global weather experiment. The days so chosen were 19, 20 and 21 May 1979. The currently operational version of the ARPE model was then run using the archived analyses valid for each of the three days, producing thickness height prognoses valid for 20, 21 and 22, May. Statistical corrections were then computed using, it should be noted, covariances estimated from data for 1972-1977, and these corrected prognoses were synthesised using the procedure described in Bennett and Leslie (1981). The uncorrected prognoses for 20 and 21 May were very good, with S_1 skill scores around 24. The corrections did reduce the scores by a further point, but the uncorrected prognoses were already subjectively indistinguishable from the corresponding verification analyses. However, the uncorrected prognosis valid for 22 May was relatively poor, and had an S_1 skill score of 40. Nevertheless, the corrections reduced the score by 4 points, and the corrected prognosis was markedly superior (see Figs 3(a) to (d)). The improvement obtained from the statistical correction comes from the three sources of increased skill typical of the scheme, namely: better forecasts of the height fields over the tropics (the uncorrected model forecasts usually tend to increase the height fields unrealistically over much of the tropics); superior movement of the trough in the Great Australian Bight; and increased skill over the poorer data areas south of the continent.

The second case study was chosen because it represented a relatively rare event, namely an unusually severe cold outbreak, that occurred on 31 May 1977. Melbourne's maximum temperature was only 8.7°C on that day. As can be seen in Fig. 4, the statistically-corrected forecast (Fig. 4(d)) was a marked improvement on the uncorrected forecast (Fig. 4(c)). In particular the thickness values over Victoria and Tasmania are much closer to the verifying analysis (Fig. 4(b)) and have predicted the intrusion of cold air more accurately than in the uncorrected forecast. The S_1 skill scores for the model and the statistically-corrected model were 31 and 28 respectively, an increase in skill of 3 points. The thickness values over the tropics have also been maintained by the corrected forecast.

Fig. 3 Respectively: (a) — analysis of $\Delta\phi_{500-1000}$ valid at 0000 GMT 21 May 1979; (b) — analysis of $\Delta\phi_{500-1000}$ valid at 0000 GMT 22 May 1979; (c) — model prognosis; and (d) — statistically-corrected model prognosis.

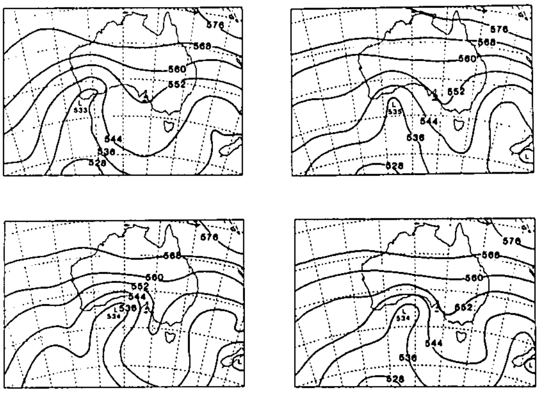
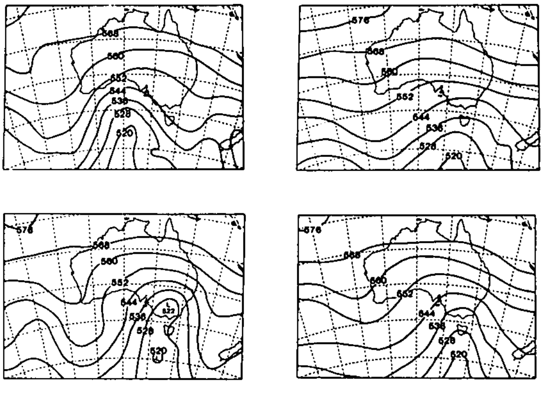


Fig. 4 As in Fig. 3 except for 0000 30 May 1979 to 0000 GMT 31 May 1979.



References

Bennett, A. F. and Leslie, L. M. 1978. Statistical correction of dynamical prognoses in the Australian region. *Mon. Weath. Rev.*, *107*, 1254-62.

Bennett, A. F. and Leslie, L. M. 1981. Statistical correction of the Australian Region Primitive Equation model. *Mon. Weath. Rev.*, *109*, 453-62.

Davis, R. E. 1976. Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.*, *6*, 249-66.

Rinne, J. 1971a. Investigation of the forecasting error of a simple barotropic model with the aid of empirical orthogonal functions. Part I. *Geophysica*, *11*, 185-213.

Rinne, J. 1971b. Investigation of the forecasting error of a simple barotropic model with the aid of empirical orthogonal functions. Part II. *Geophysica*, *11*, 215-39.

Rinne, J. 1972. Investigation of the forecasting error of a simple barotropic model with the aid of empirical orthogonal functions. Part III. *Geophysica*, *12*, 57-79.

Teweles, S. Jr. and Wobus, H. B. 1954. Verification of prognostic charts. *Bull. Am. met. Soc.*, *35*, 455-63.