The Stability of Empirical Long-Range Forecast Techniques: A Case Study

N. NICHOLLS
Australian Numerical Meteorology Research Centre, Melbourne, Australia

(Manuscript received 31 May 1983, in final form 24 August 1983)

ABSTRACT

The stability of simple linear regression equations for the long-range prediction of Australian spring rainfall was studied. Specifically, the way in which the accuracy of the forecasts depends on the number of years of data used to derive the equations and the length of the period between the end of the data used in forecast equation derivation and the application of the equations in prediction were examined. An optimum period of data of about 15 years was found; the use of more or less data in deriving the forecast equations led to deterioration in the forecasts. The forecasts also deteriorated if the equations were used for more than a few years after the end of the period of data from which they were derived, suggesting a need for routine updating of the forecast equations. The lack of stability in the forecast equations presumably reflects nonstationarity in the data series possibly resulting from changes in the general circulation patterns. If this is so, the results might be applicable to similar statistical long-range forecast methods in other areas.

1. Introduction

Kung and Sharif (1982) formulated a long-range forecasting scheme for the Indian monsoon and *inter alia,* examined the effect of varying the number of years used in developing the predictive equations on the error in the forecasts. They concluded that the regression coefficients need to be calculated with ~16–19 years data if the forecast errors are to be minimized. Kung and Sharif also suggested that continuous updating of the regression coefficients should be an integral part of an empirical long-range forecast scheme because of the temporal variation of the patterns of the general circulation. McBride and Nicholls (1983) reached a similar conclusion, based on a study of the secular variation in the correlations between Australian rainfall and the Southern Oscillation. Rao (1976), Bell (1977) and Khandekar (1979) have also commented on the instability of empirical long-range forecast relationships.

The study described herein extends the exploratory study of Kung and Sharif (1982). It examines the stability of empirical long-range forecast equations in an area far removed from that examined by Kung and Sharif and uses considerably more data, thus allowing greater confidence in the results. By "stability" is meant the ability of a technique, derived from data in one period, to produce skillful forecasts in another period. Two aspects of stability are addressed. First, the optimum length of record needed to develop the equations is determined. Second, the possible need to continuously rederive the equations as more recent data becomes available is examined.

A case study approach is adopted in the examination of these problems. The case chosen is the prediction of September–November district average rainfall in the Darwin-Daly and South Mallee rainfall districts of Australia (see Fig. 1 for locations) from just one predictor, Darwin August station level pressure. The two areas have been chosen because several recent studies, through the successful verification of long-range forecast techniques first suggested over fifty years ago, have already established that spring rainfall in these areas can be predicted from prior Darwin pressures (Nicholls, 1981; Nicholls and Woodcock, 1981; Nicholls *et al.,* 1982). The forecast technique used throughout this paper is simple linear regression and the accuracy of forecasts obtained from applying the regression equations is examined as the record length used in deriving them and the lag between the end of this period and the period of their application in prediction, is varied.

2. Optimum data period for deriving prediction equations

If time series of meteorological observations were statistically stationary, it would be best, in deriving empirical forecast techniques using regression equations, to use as much data as possible. The more data used, the closer the equations calculated from this sample should approach the ideal predictive equations which would result from derivation with the complete population. However, meteorological time series are not, in general, stationary. They often show abrupt changes in the mean, or trends, or quasi-periodicities. Much of this nonstationarity can be attributed to changes in observing techniques, for example the changing exposure of a raingage when surrounding trees are removed or nearby open fields have large buildings erected on them. Anthropogenic effects on

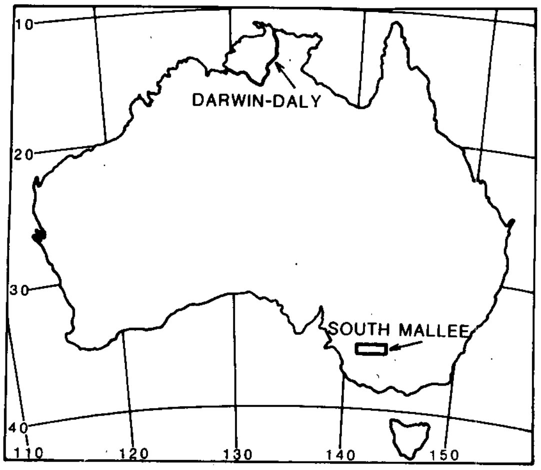

FIG. 1. Locations of Darwin-Daly and South Mallee rainfall districts.

a larger scale can also result in secular changes in meteorological time series. For instance, the continued burning of fossil fuels may increase atmospheric carbon dioxide which, in turn, might result in increased atmospheric temperatures. A third possible cause of nonstationarity in such time series is natural (i.e., not induced by man) climate change. Climate has shown considerable variation on time scales of decades or longer. Such variations could also imply non-stationarity in corresponding time series. Whatever the cause of the nonstationarity, its possible presence suggests that the use of all available data in predictive equation derivation might not be optimum. In this section, the accuracy of long-range predictive equations, derived from a variety of record lengths, is examined for the cases noted before.

Data from 1913–80 has been used and the following statistics calculated: root-mean-square error (RMSE), the bias (BIAS) and the absolute error (ABSE). The equations used to calculate these statistics are:

$$\left.\begin{aligned} \text{RMSE} &= [\sum_{y=1951}^{1980} (\hat{R}_{yn} - R_y)^2/30]^{1/2} \\ \text{BIAS} &= \sum_{y=1951}^{1980} (\hat{R}_{yn} - R_y)/30 \\ \text{ABSE} &= \sum_{y=1951}^{1980} |\hat{R}_{yn} - R_y|/30 \end{aligned}\right\}.$$

The linear correlation coefficient r between R_y and $\hat{R}_{yn}$ using data from 1951–80 has also been calculated for different values of n. Thus, the predictions have been verified on 30 years of data, from 1951–80, using a variety of measures. *All* error statistics and correlations, in this and the next section, have been calculated using the entire 30-year test period.

In the above equations, y is the year (from 1951 to 1980) for which the prediction was made, R_y is the observed September–November rainfall in year y and $\hat{R}_{yn}$ is the predicted September–November rainfall in year y, predicted from the linear regression derived on n years data (n varied from 4 to 30) ending in year $y - 1$. Thus, the accuracy of the predictions of equations derived from data periods of 4 to 30 years have been tested on an identical 30-year (1951–80) sample. The results of this exercise for the two districts are shown in Figs. 2 and 3. The results for the Darwin-Daly district (Fig. 2) are discussed first.

For each district the accuracy of the predictions can be compared with the accuracy of "climatological predictions," i.e., predicting that September–November rainfall in a specific year will equal the long-term average September–November rainfall. The long-term average has been calculated by averaging September–November rainfall for the 30 years immediately preceding the year in which a forecast is to be made. Climatological predictions of September–November rainfall were made in this way for each year from 1951 to 1980 and the RMSE, BIAS and ABSE were calculated for these forecasts using the formulas listed above. At Darwin-Daly the RMSE of the climatological predictions was 71 mm, the BIAS was −0.8 mm and the ABSE was 56 mm. For the South Mallee, the corresponding values were 48, −1.3 and 39 mm.

The bias in the forecasts for Darwin-Daly September–November rainfall is small, regardless of the record length used to derive the forecast equations, and reaches a maximum absolute value of 8 mm when 17 years are used in the sample for the derivation of the forecast equation. This bias is rather larger than the bias of the climatological predictions (−0.8 mm).

The other three statistics show different behavior. For each of them, the worst performance of the pre-

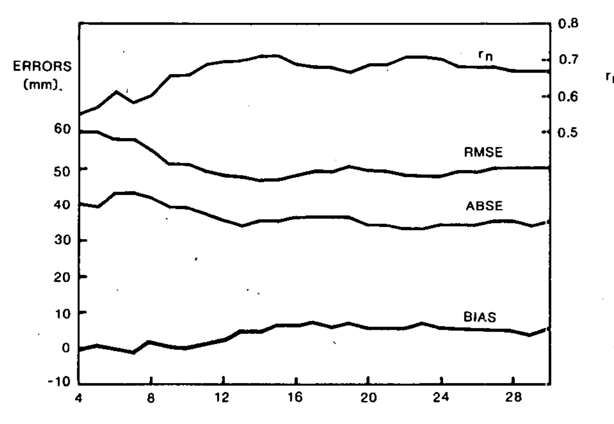

FIG. 2. Bias (BIAS), absolute error (ABSE) and root-mean-square error (RMSE) of forecasts of Darwin-Daly September–November rainfall from Darwin August pressure for varying sample sizes (n) used in deriving the prediction equation. Also shown are the correlations (r) between predicted and observed rainfall. All statistics have been calculated from forecasts for the years 1951–80, using data up to the year prior to the year for which a forecast is to be made to derive the prediction equation.

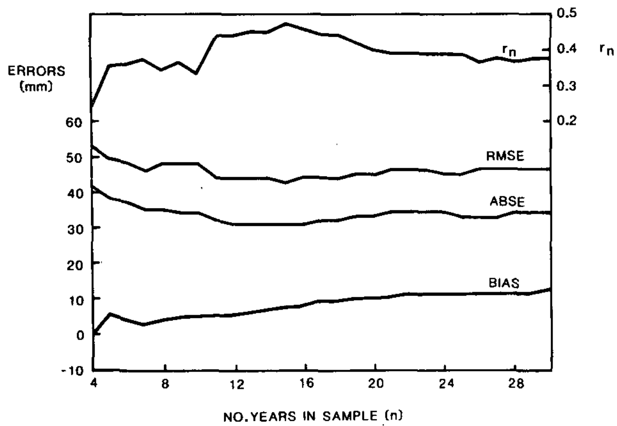

FIG. 3. As in Fig. 2 but for South Mallee.

dictions results from using only a few years data (less than eight) to derive the forecast equations. Thus, the maximum rms error of 60 mm occurs when n (the number of years used to derive the prediction equation) is 4 and drops to a minimum of 47 mm when 14–15 years data are employed. The correlation coefficient between observed and predicted rainfall is smallest (0.55) when $n = 4$ and largest (0.71) when n reaches 14. The absolute error is largest (43 mm) at $n = 7$ and falls to a minimum of 33 mm at $n = 22$. Each statistic then exhibits a gradual deterioration as n increases; thus, the correlation falls to 0.67 at $n = 30$. The rms error and the absolute error increase slightly as n increases toward 30. The differences in the statistics between when n is ~15 and when n is ~30, although suggesting the presence of nonstationarity in the data, are not large. However, the results shown in Fig. 2 do not suggest a need for using more than ~15 years of data in deriving simple linear regression long-range forecast equations, at least in this specific case, since the forecasts made for the 1951–80 period from regression equations derived on 15 years data were at least as good as the forecasts for this period made from equations derived using larger amounts of data.

The RMSE and ABSE of the predictions from Darwin pressure were smaller than for the climatological predictions, regardless of the value of n.

Corresponding results for the prediction of South Mallee September–November rainfall from August Darwin pressure are shown in Fig. 3. The South Mallee behavior is similar in many respects to the Darwin-Daly forecasts. The bias is again small, although larger than was the case in Fig. 2, and it continues to grown in magnitude as n increases. Again, the bias is larger than for the climatological predictions.

The other three statistics show similar behavior to their counterparts in Fig. 2 with an optimum sample length for deriving the predictive equation of about 15 years. Again, the performance is poorest when only a few years of data are used to derive the prediction equation. For instance, the correlation coefficient between observed and predicted rainfall is only 0.23 when $n = 4$, climbs to a maximum of 0.47 with $n = 15$ and then gradually falls to below 0.4 as n is increased further. The rms errors and absolute errors also indicate a gradual deterioration in the forecasts as n is increased above ~15. Except for small values of n (less than six), the RMSE and ABSE of the predictions from Darwin pressure are smaller than for the climatological predictions.

These results can be compared with those obtained by Kung and Sharif (1982), although they only examined the performance of their prediction equations with n varying from 10 to 19 years, and the number of forecasts available to them for verification varied from 10 (when $n = 10$) down to only one (when $n = 19$). The results shown in Figs. 2 and 3 come from verifications of 30 predictions for each value of n from 4 to 30. The same 30 years have been used in verifying the forecast equation for each value of n used. Kung and Sharif also found that their forecast equations performed worse when small amounts of data were used in their derivation.

The slight deterioration noted in Figs. 2 and 3 as n increases above 20 suggests that the meteorological time series used here are not statistically stationary. If nonstationarity does exist, Kung and Sharif might be correct in recommending the continuous updating of regression equations with new data. The next section examines, for the two regions discussed in this section, the need for such updating.

3. The need to update prediction equations

The possibility that regression equations used in long-range prediction need to be continuously updated as new data becomes available is examined here by varying the lag between the end of the period used to derive the equation and the year in which the equation is used to make a prediction. The statistics of the previous section have been rederived with the single change that $\hat{R}_{yn}$ has been replaced by $\hat{R}_{ynd}$ which is the predicted September–November rainfall in year y (from 1951–80), predicted from the linear regression derived on n years data (from 4 to 30) ending in year $y - d$, where d varies from 1 to 35. In the previous section, d was kept constant at a value of 1.

Since the data series available is limited, d is also limited so that $d_{\max} = 39 - n$. Thus, when we have 30-year samples to calculate the regression, only lags of from 1 year up to 9 years can be examined.

The statistics have been calculated, again, for the prediction of September–November rainfall for the two districts shown in Fig. 1. The results, for 15-year samples (i.e., $n = 15$) for the rms error and the correlation between observed and predicted rainfall are shown in Fig. 4 (for Darwin-Daly) and Fig. 5 (for South Mallee) for values of d from 1 to 24 i.e., lags of from 1 to 24

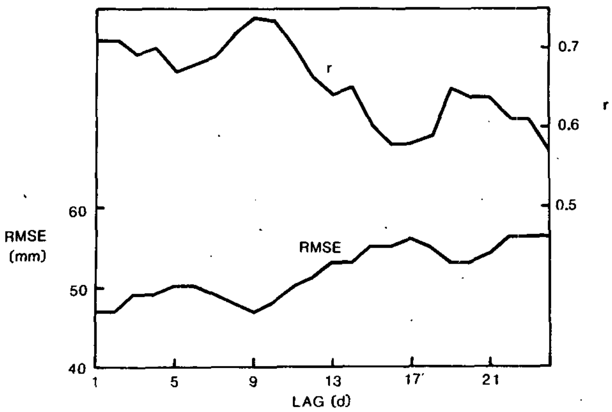

FIG. 4. Root-mean-square error (RMSE) of forecasts of Darwin-Daly September–November rainfall from Darwin August pressure and correlation (r) between predicted and observed rainfall. Statistics calculated from forecasts for the years 1951–80 using 15 years data to derive the prediction equation and varying the lag (d) between the end of the data period used to derive the prediction equation and the year for which a forecast is made.

years between the end of the data sample used to derive the equations and the year in which the equation is used to make a prediction. It is reiterated that the statistics have been calculated, for each value of d, on the same 30 years of data from 1951–80.

The results for the two areas are similar. Both show a deterioration in predictive performance as the lag between the end of the dependent sample and the year in which the equation is used increases. In Fig. 4, the correlation drops from 0.71 to less than 0.6 while the rms error increases from ~47 mm to 56 mm. The deterioration as the lag increases from one to 24 years is even greater in Fig. 5 where the correlation drops from 0.47 to about zero, with a corresponding increase in root-mean-square error from 43 mm to ~55 mm.

A similar deterioration in prediction performance is observed if values of n other than 15 are used. For instance, Figs. 6 and 7 show for the Darwin-Daly and South Mallee districts respectively, how the rms error varies with the lag for values of n of 10 and 20 years (i.e., with 10 or 20 years used to derive the prediction equation). For the 20-year case only lags of up to 19 years can be obtained, while for the 10-year samples lags of up to 29 years can be examined. In both cases, a marked increase in rms error is seen as the lag increases. The deterioration is more serious for the South Mallee case than for Darwin-Daly. In the latter case, even though increasing d leads to an increase in RMSE, the RMSE of the climatological predictions is never exceeded, even when $d = 29$ years. In the South Mallee, however, the RMSE of the predictions from Darwin pressure exceeds the RMSE of the climatological predictions when d exceeds nine or ten years.

The results illustrated in Figs. 4–7 indicate that the time series used, or the relationships between the time series, are not sufficiently stationary to enable the successful use of predictive equations derived using data in one period on data from a much earlier or later period. As new data becomes available, the prediction equations need to be updated. However, the degradation as the lag increases is fairly gradual initially (Figs. 4–7) and thus, predictive equations can probably be used for several years before requiring rederivation on the new data, at least for the time period and areas examined here.

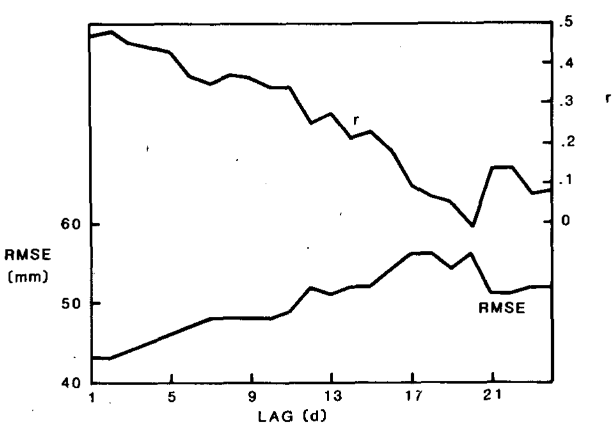

FIG. 5. As in Fig. 4 but for South Mallee.

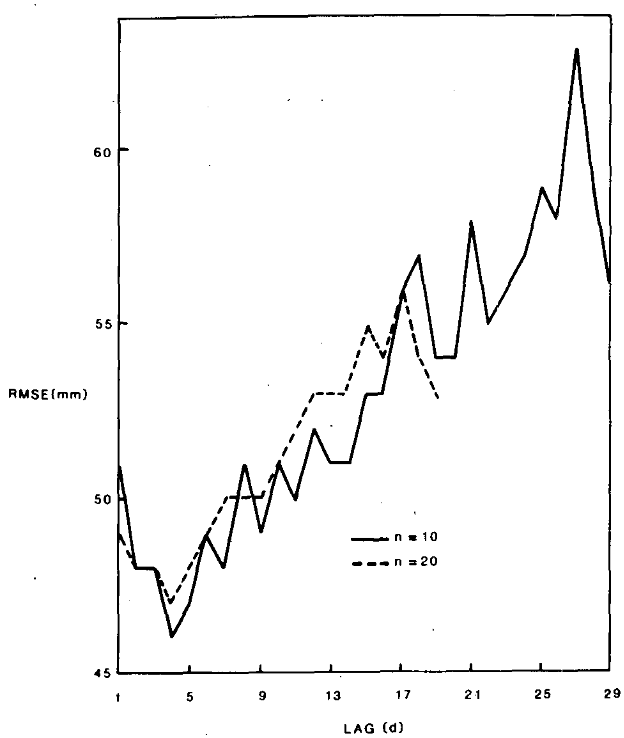

FIG. 6. As in Fig. 4 but with 10 and 20 years of data used to derive the prediction equation and not showing the correlation between forecast and observed rainfall.

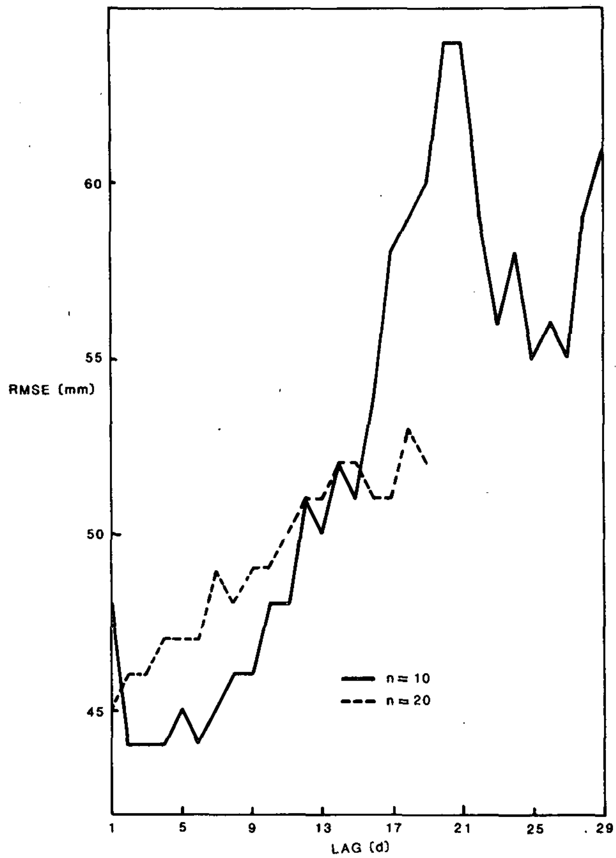

FIG. 7. As in Fig. 6 but for South Mallee.

4. Concluding remarks

The case study described above has used the previously demonstrated feasibility of long-range prediction of Australian spring rainfall to examine the stability of empirical predictive equations. The way in which the accuracy of spring rainfall forecasts for the 30-year period 1951–80 varied was studied as the number of years of data used to derive the forecast equation and the length of time between the end of this "forecast equation deriving" period and the year in which the equation was used to make a forecast, varied. Two conclusions were reached:

1) Approximately 15 years was the optimum data period for deriving the prediction equations. The use of shorter or longer periods of data led to a deterioration in the forecasts.

2) The prediction equations do not require *continuous* updating as new data becomes available but they should be rederived every few years, if deterioration of the forecasts is to be avoided.

These conclusions refer specifically to the case study examined; only further case studies, in other areas where evidence of long-range predictability exists, can conclusively determine whether they have a wider application. However, such could well be the case, as the cause of the instability noted in the forecast equations arises from nonstationarity in the time series of the data or from trends or fluctuations in the lag relationships between the time series. In turn, such nonstationarity or fluctuations may reflect a temporal variation in the patterns of the general circulation, as suggested by Kung and Sharif (1982). If so, it seems unlikely that the effects of such a shift would be restricted to the area considered in the present case study and therefore, the above conclusions might be applicable in other cases.

REFERENCES

Bell, G. J., 1977: Changes in sign of the relationship between sunspots and pressure, rainfall and the monsoons. *Weather,* **32,** 26–32.

Khandekar, M. L., 1979: Climate teleconnections from the equatorial Pacific to the Indian Monsoon—analysis and implications. *Arch. Meteor. Geophys. Bioklim.,* **A28,** 159–168.

Kung, E. C., and T. A. Sharif, 1982: Long-range forecasting of the Indian summer monsoon onset and rainfall with upper air parameters and sea surface temperature. *J. Meteor. Soc. Japan,* **60,** 672–681.

McBride, J. L., and N. Nicholls, 1983: Seasonal relationships between Australian rainfall and the Southern Oscillation. *Mon. Wea. Rev.,* **111,** 1998–2004.

Nicholls, N., 1981: Air–sea interaction and the possibility of long-range weather prediction in the Indonesian Archipelago. *Mon. Wea. Rev.,* **109,** 2435–2443.

——, and F. Woodcock, 1981: Verification of an empirical long-range weather forecasting technique. *Quart. J. Roy. Meteor. Soc.,* **107,** 973–976.

——, J. L. McBride and R. J. Ormerod, 1982: On predicting the onset of the Australian wet season at Darwin. *Mon. Wea. Rev.,* **110,** 14–17.

Rao, Y. P., 1976: Southwest Monsoon. *Meteor. Monogr., Synoptic Meteorology,* Vol. 1, Indian Meteorology Department, Poona, 354–360.