Atmósfera (1994), 7, pp. 83-87

An experiment with linear regression in forecasting of spring rainfall over south Brazil

V. BRAHMANANDA RAO and K. HADA

Instituto Nacional de Pesquisas Espaciais -INPE, C. P. 151, 12201 - São José dos Campos, SP, Brasil

(Manuscript received Oct. 22, 1992; accepted in final form July 28, 1993)

RESUMEN

Se examina la posibilidad de la predicción de lluvia en primavera sobre Río Grande do Sul (Brasil) a partir de las observaciones previas del índice 50. Los resultados muestran que puede obtenerse un éxito moderado. El periodo óptimo de datos para la derivación de ecuaciones de predicción es de cerca de 11 años.

ABSTRACT

The feasibility of predicting spring rainfall over Rio Grande do Sul (Brazil) from the prior observations of SO index is examined. It is found that moderate success can be obtained. The optimum data period for deriving the prediction equations is about 11 years.

1. Introduction

More than half a century ago Mossman (1924) noted the association between the Southern Oscillation (SO) and rainfall over South Brazil, Paraguay and North Argentina as represented by level of Parana river at Rosario, Argentina. Mossan's results show that positive river level departures occur during the negative phase of the SO, when sea level pressure (SLP) is anomalously high in the region of Australia. Kousky *et al.* (1984) obtained a correlation coefficient of 0.56 between the level of Parana river as given by Mossman and the SO index (SLP of stations around Darwin, Australia) for the period 1885-1915. Aceituno (1988) and Rao and Hada (1989) confirmed the strong correlation between the SO and rainfall over South Brazil. The question that arises now is: can the SO index be used to predict seasonal rainfall over South Brazil? The purpose of the present study is to explore this possibility. The method used is similar to that developed by Kung and Sharif (1982), Nicholls (1984) and Nicholls *et al.* (1982).

Rao and Hada (1990) noted that the rainfall over South Brazil during the (spring) season, September, October and November (SON) shows highest correlation with the SO index of June, July and August (JJA). The correlation between the rainfall over South Brazil during the other seasons and the SO index is small. Thus in the present study the prediction of SON rainfall is attempted with the SO index of JJA. The forecasting method used is simple linear regression. The success of the forecast is examined by calculating the errors as the record lenghs used in deriving the regression equations is varied.

2. Data source

Rainfall data for four raingauge stations (S. L. Gonzaga, 28° 24' S, 55° 1' W; Cruz Alta, 28° 38' S, 53° 36' W; Porto Alegre 30° 04' S, 51° 13' W, and Alegrete 29° 46' S, 55° 47' W) for the period 1935-85 were obtained from Instituto Nacional de Meteorología of Brazil. At some stations a few years of data were missing. For example at Alegrete five years data were missing. However, the years of missing data at one station did not coincide with the years of missing data over other stations. Thus we could obtain 51 years of continuous data, although for a few years (about five) the number of stations used were less. Data were checked for consistency. The SO index data were obtained from climatic diagnostics Bulletin (1986).

3. Results

Figure 1 shows the SON rainfall over Alegrete plotted against the SO index for JJA for the period 1935-85. The correlation coefficient between the two series is -0.54. It could be seen that the rainfall of this season over Rio Grande do Sul can be predicted from prior observations of SO index.

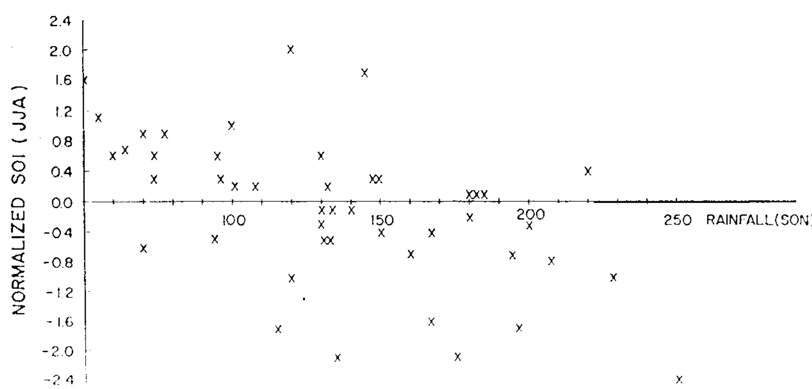

Fig. 1. Spring (September, October and November) rainfall over Alegrete (29° 46'S, 55° 47'W) and normalized SO index of June, July and August.

Next, optimum data period for deriving prediction equations is determined. This is necessary because time series of meteorological observations are not in general statistically stationary. The accuracy of long range prediction equations, derived from different record lengths is examined. The following statistics are calculated using the rainfall data of 1935-85.

$$\text{Root-mean-square error, RMSE} = \Big(\sum_{y=1961}^{1985} (R_{yn} - R_y)^2/25\Big)^{1/2}$$

$$\text{Bias, Bias} = \sum_{y=1961}^{1985} (R_{yn} - R_y)/25$$

$$\text{Absolute error, ABSE} = \sum_{y=1961}^{1985} \mid R_{yn} - R_y \mid /25$$

The rainfall data used are the average rainfall of SON for the four stations mentioned earlier. In the expressions above y is the year (1961-1985) for which the prediction was made, R_y is the observed SON rainfall in year y, and R_{yn} is the predicted SON rainfall in year y, predicted from linear regression derived on n years of data (n varied from 5 to 25) ending in year $y-1$. Thus the accuracy of the prediction equation derived from data period of 5-25 years has been tested on an identical 25 years (1961-1985) sample. In addition to the above statistics, the linear correlation coefficient r between R_y and R_{yn} using the data from 1961-1985 has also been calculated for different values of n.

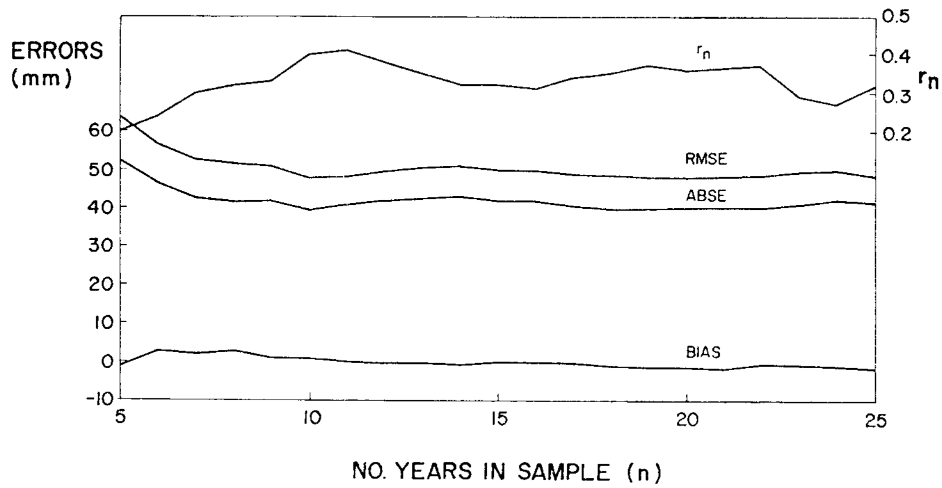

Fig. 2. Bias (BIAS), absolute error (ABSE) and Root-mean-square error (RMSE) of forecasts of South Brazil September, October and November rainfall from SO index of June, July and August for varying sample size (n) used in deriving the prediction equation. Also shown are the correlations (r) between predicted and observed rainfall. All statistics have been calculated from forecasts for the years 1961-1985, using data up to the year prior to the year for which a forecast is to be made to derive the prediction equation.

The results are shown in Figure 2. These are compared against climatological prediction. Climatological predictions were made by taking the average of 25 years immediately preceeding the year in which a forecast is made. Climatological predictions were made in this way for the period 1961-1985 and RMSE, BIAS and ABSE were calculated. These were respectively 50 mm, 0.7 mm and 43,1 mm.

The bias is generally small. The worst performance of the prediction is that due to a few years of data and the best performance is using about 11 years of data ($n = 11$). For this optimum period the RMSE and ABSE of the predictions were smaller than for the climatological prediction. Even for other values of n (except n around 13, 14) this is true. These results are similar to those obtained by Nicholls (1984), although he obtained best results in the case of Australia for n around 15.

The existence of non-stationarity in meteorological time series makes the forecast to deteriorate. This necessitates continuous updating of regression equation with new data. The necessity of updating is shown here by varying the lag between the end of the period used to derive the prediction equation and the year in which the equation is used to make prediction. Figure 3 shows the RMS error and correlation between the observed and predicted rainfall as the lag is varied. The maximum lag, d_{max} is $27 - n$. Figure 3 shows the two statistics for $n = 10$ and 15. It could be seen from Figure 3 that a deterioration in predicted performance occurs as the lag is increased. The correlation between the observed and predicted rainfall decreases from 0.4 to around 0.2 as the lag increases from one to 14. The corresponding root mean square error increases from 36 mm to 57 mm. These results indicate that the time series used are not stationary and there is a need for updating of prediction equation.

The study is repeated using the SON rainfall data of four stations separately. Results (not shown here) show essentially similar behavior, although the optimum period varied slightly.

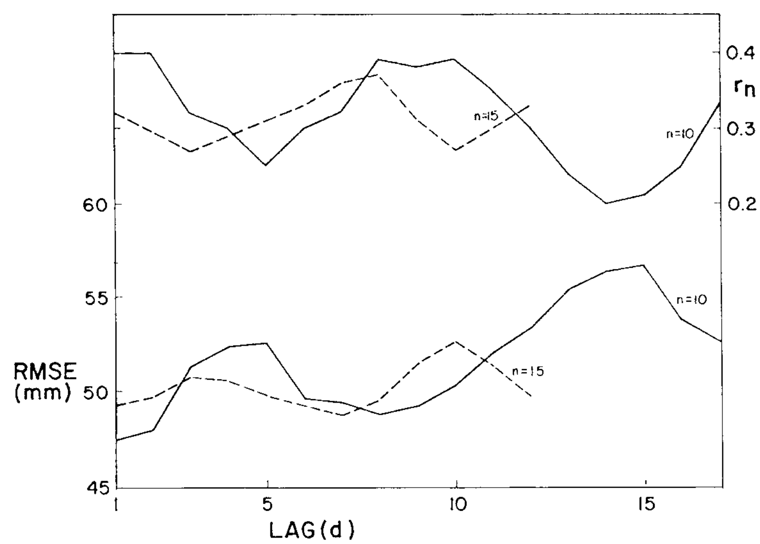

Fig. 3. Root-mean-square error of forecasts of South Brazil September, October and November rainfall from SO index of June, July and August and correlations (r) between predicted and observed rainfall. Statistics are calculated from forecasts for the years 1961-85 using 15 and 10 years data to derive the prediction equation and varying lag (d) between the end of the data period used to derive the prediction equation and the year for which a forecast is made.

4. Conclusion

The repesent study examined the feasibility of predicting SON rainfall over Rio Grande do Sul (Brazil) using the prior observation of SO index. It can be concluded that modest success can be obtained and the optimum data period for deriving the prediction equations is about 11 years. The prediction equations have to be rederived every few years to avoid deterioration of the forecast.

Acknowledgements

Thanks are due to Dr. Sergio H. Franchito for critically going through the manuscript.

REFERENCES

Aceituno, P., 1988. On the functioning of the Southern Oscillation in the South American sector. Part I. Surface Climate. *Mon. Wea. Rev.*, **116**, 505-524.

Climate Diagnostics Bulletin, 1986. Global analyses and indices. National Meteorological Center, Wahsington, DC.

Kousky, V. E., M. T. Kagano and I. F. A. Cavalcanti, 1984. A review of the Southern Oscillation: oceanic–atmospheric circulation changes and related rainfall anomalies. *Tellus*, **36**A, 490-504.

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

Mossman, R. C., 1924. Indian monsoon rainfall in relation to South American weather, 1875-1916. *Mem. Ind. Meteor. Dept* ., **23**, 157-242.

Nicholls, N., 1984. The stability of empirical long–range forecasting: A case study. *J. Clim. and Appl. Meteor.*, **23**, 143-147.

Nicholls, N., J. L. McBride and R. J. Ormerod, 1982. On predicting the onset of Austrian wet season at Darwin. *Mon. Wea. Rev.*, **110**, 14-17.

Rao, V. B. and K. Hada, 1990. Characteristics of rainfall over Brazil: Annual variations and connections with the Southern Oscillation. *Theor. Appl. Climatol.*, **42**, 81-91.