Prediction of Northeast Brazil Rainfall Anomalies

STEFAN HASTENRATH

Department of Meteorology, University of Wisconsin, Madison, Wisconsin

(Manuscript received 27 September 1989, in final form 27 February 1990)

ABSTRACT

Expanding on earlier research, empirically based methods are developed for predicting March–September rainfall in northern Northeast Brazil. From a network of 27 raingage stations in Brazil's Nordeste, new rainfall index series are constructed for March–September (MS) and October–January (OJ). Data input to stepwise multiple regression models further include January values of the Tahiti minus Darwin pressure index, an index of sea surface temperature (SST) in the equatorial Pacific (PWT), and indices of the fields of meridional (v) and zonal (u) wind components and of SST in the tropical Atlantic between 30°N and 30°S (AFV, AFU, and AFT, respectively). Empirical orthogonal function (EOF) analysis of the v, u, and SST fields was used in constructing the latter three indices. Throughout the study, a sharp distinction is kept between a "dependent" dataset (1921–42 and 1948–57) used as a training period and an "independent" portion of the record (1958–87) reserved for prediction experiments.

Preseason rainfall (OJ) is by far the most powerful predictor, allowing 52% of the interannual MS variance in an independent dataset to be forecast. From experiments with unrestricted data input, a model with the predictors OJ, AFV, and PWT shows the best performance, capturing 71% of the MS variance in the independent dataset. Experiments on the optimal length of the training period suggest that one of 20–30 years is adequate. Updating offers no advantage over a fixed training period. For the purposes of operational application, the maintenance and timely processing of raingage measurements is of paramount importance. The next important task, demanding considerably greater resources, would be the real-time monitoring of the surface wind field over the tropical Atlantic.

1. Introduction

Climate prediction has been recognized as an important objective in recent research initiatives at the national and international levels (National Climate Program Office 1988, p. 2, 19–26; World Meteorological Organization 1980, p. 42). Investigations into the mechanisms of tropical rainfall anomalies created the basis for the development of tropical climate prediction methods in the 1980s (review in Hastenrath 1986, 1988a, p. 330–352). The droughts of Northeast Brazil (Nordeste) represent a particular "target of opportunity" because they are unusually well defined in terms of the large-scale circulation setting and possess an extraordinary economic and societal impact. Work at the University of Wisconsin in the mid-1970s (Hastenrath and Heller 1977) elucidated the general circulation causes of rainfall anomalies in northern Northeast Brazil. Building on this diagnostic research, the predictability of these anomalies was demonstrated, and an empirically based forecasting method was presented (Hastenrath et al. 1984).

Important tasks remain: (i) As elsewhere in the tropics, the raingage network in the Nordeste has progressively decayed over the past two decades. Accordingly, it has become necessary to identify the reduced number of stations still functioning regularly. From these, new rainfall indices can be compiled that are expected to be available in the future. (ii) Our previous work (Hastenrath and Heller 1977; Hastenrath et al. 1984) was based on data to 1972. It is desirable to evaluate the surface ship observations and land station records since then, with a view to improving the diagnostic and prognostic relationships.

The present endeavor has two major goals. The first is the design of a "poor man's" forecasting system, which would have moderate performance but would draw only on information that could now be compiled with adequate lead time. The second line of inquiry is directed at the intrinsic predictability of Nordeste rainfall from all diagnostically meaningful atmospheric and oceanic information, even though only part of this is presently available timely enough for purposes of operational climate prediction. An account of the main results is provided.

2. Background

The annual cycle of circulation and climate in the Brazil–tropical Atlantic sector and the mechanisms of rainfall anomalies in the Nordeste have been discussed extensively before (Hastenrath and Heller 1977; Has-

Corresponding author address: Prof. Stefan Hastenrath, Department of Meteorology, University of Wisconsin, 1225 West Dayton Street, Madison, WI 53706.

tenrath et al. 1984; Hastenrath 1988a, p. 293–300), and a brief summary must suffice here. The interior of northern Northeast Brazil (Fig. 1) is a semiarid region with annual precipitation of 300–800 mm and a long dry season centered around June–September. Sparse rains may start from October onward, but the rainy season proper is narrowly concentrated around March–April, and precipitation tapers off drastically from May onward.

At least three factors in the general circulation setting favor precipitation around March–April: (i) At that time of year, the near-equatorial trough and embedded confluence and intertropical convergence zone (ITCZ) over the adjacent tropical Atlantic reach their southernmost position and thus a location closest to the Nordeste (Hastenrath and Lamb 1977, charts 2–25); (ii) the waters of the equatorial South Atlantic are warmest, which serves to enhance the instability and possibly the moisture content of the boundary-layer flow directed into the interior of the Nordeste; and (iii) throughout the year, the sea surface temperature (SST) pattern features warmest waters in a band across the tropical North Atlantic contrasting with colder conditions south of the equator. This juxtaposition of warm and cold surface waters may serve to induce a thermal circulation with subsidence over the Nordeste, which would hamper precipitation. In the latter part of the boreal winter half-year, however, the interhemispheric SST gradients grow weakest, thus mitigating against the aforementioned thermally induced vertical circulation.

The circulation mechanisms associated with rainfall anomalies in Northeast Brazil can be understood largely as enhancements or reductions of the average annual cycle. Thus, drought years are characterized by an anomalously far northerly position of the ITCZ, reduced northeast trades and accelerated cross-equatorial flow from the Southern Hemisphere, and anomalously warm surface waters in a zonal band across the tropical North Atlantic contrasting with negative SST anomalies south of the equator. Moreover, droughts tend to coincide with the low phase of the Southern Oscillation (SO), defined by anomalously low/high pressure at Tahiti/Darwin. This association of Nordeste rainfall anomalies with the SO appears related to the fact that during boreal winter of the low/high SO phase, the near-equatorial trough tends to be displaced northward/southward (Hastenrath et al. 1987).

Drawing on these diagnostic findings, promising predictor candidates were selected and compiled in time series form by Hastenrath et al. (1984). These included the zonal and meridional wind components over limited areas of the equatorial Atlantic, SST in domains of the tropical North Atlantic and tropical South Atlantic, an index of the Southern Oscillation, and preseason rainfall in the Nordeste itself. The last can be taken as a basic indicator of the preseason latitudinal position of the ITCZ (Hastenrath and Heller

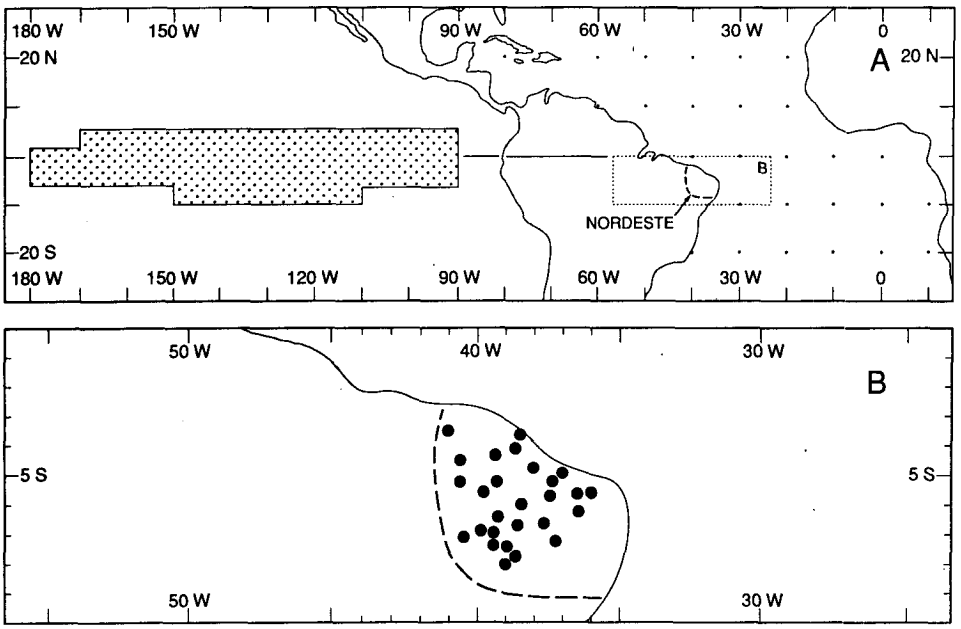

FIG. 1. Orientation maps. (a) Dot grid in the Atlantic indicates the 10° × 10° areas for which ship observations of sea surface temperature and wind were compiled. Dot raster in the equatorial Pacific marks the domain (2°–6°N, 90°–170°W; 2°N–6°S, 90°–180°W; 6°–10°S, 110°–150°W) for which the sea surface temperature index PWT was compiled. Broken line encloses the domain of the 27 raingage stations used for the northern Nordeste. Dotted-line rectangle bounds the area represented in enlarged scale in (b). (b) Heavy dots indicate the location of the 27 raingage stations.

1977; Hastenrath et al. 1984; Hastenrath 1988a, p. 293–300). The aforementioned elements served as input to a stepwise multiple regression scheme, using the years 1921–40 and 1946–56. The resultant regression model was then used in a predictive mode on an independent dataset for 1958–72. Results indicated that more than half of the interannual rainfall variance in Brazil's Nordeste can be predicted from antecedent anomalies in the large-scale circulation. Natural extensions to that study are the evaluation of the climatic record beyond 1972, improvement of prognostic relationships, and work toward operational application.

3. Observations and basic data processing

The observational basis for the present study consists of rainfall records in the Nordeste, ship observations in the tropical Atlantic and Pacific, and the surface pressure series at Tahiti and Darwin.

Of the 40 rainfall stations used in our earlier work (Hastenrath and Heller 1977), several ceased to function altogether or no longer report regularly. In collaboration with colleagues at the Fundação Cearense de Meteorologia (FUNCEME), and at the Superintendencia do Desenvolvimento do Nordeste (SUDENE), we examined the performance of the raingage network over the years 1984–88 and identified 27 stations (Fig. 1) that still report regularly. These are contained in a limited portion of the greater Nordeste, which traditionally has possessed a relatively dense raingage network. It remains to be seen whether an adequate number of long term, reliable, and continuously operating stations can be found for the neighboring parts of the Nordeste. From the aforementioned ensemble of stations new rainfall indices were compiled as described in our earlier work (Hastenrath and Heller 1977; Hastenrath et al. 1984). For each station, rainfall was summed over the intervals October–January, October–February, March–April, and March–September, respectively. For each station and "interval," time series of normalized departures were then compiled, using as reference the mean and standard deviation of the period 1912–57. For each year, the normalized departures were summed over all 27 stations to yield time series of "all-station average normalized departure" for the "intervals" October–January, October–February, March–April, and March–September, respectively. Of these, the "intervals" October–January (OJ) and March–September (MS) are used here. The interval OJ is a measure of the preseason precipitation, while MS represents an index of rainfall from the start of the core of the rainy season. The interval MS rather than the March–April index MA is here used as predictand, because it encompasses the greater part of the season's rainfall relevant for agricultural production, and because our earlier work (Hastenrath et al. 1984) has shown it to be more predictable. This may be a consequence of enhanced data stability due to the summation over a greater part of the year than MA.

The January pressure difference of Tahiti (18°S, 150°W) minus Darwin (12°S, 31°W) during 1921–87 is used as an index (SOI) of the SO. Experiments with an SOI for November–December–January showed no improvement in the correlation with Nordeste rainfall, presumably because the enhanced data stability expected from the larger sampling period may be offset by the inclusion of less recent information.

Ship observations in the tropical oceans have been a mainstay of our work since the early 1970s (Hastenrath and Heller 1977; Hastenrath and Lamb 1977; Hastenrath et al. 1984, 1987). While we updated our archives to 1983, the COADS (Comprehensive Ocean–Atmosphere Data Set) tapes (Oort et al. 1987) have since become available. It is expected that these will be regularly updated, so that they are appropriate in future work. Therefore, the COADS compilations for January are used in the present study. From the original $2° \times 2°$ spatial resolution, the departures (from the mean of 1921–42, 1948–57) of SST and the meridional (v) and zonal (u) wind components were compacted into $10° \times 10°$ block averages of the tropical Atlantic between 30°N and 30°S (Fig. 1) for the period 1912–87. However, the years 1943–47 with sparse data are excluded from the regression and prediction experiments (section 4). A further compaction of the data is achieved by empirical orthogonal function (EOF) analysis (Norusis 1988, pp. 123–163), leading to the v, u, and SST index series AFV, AFU, and AFT, to be described in section 4.

Also based on the COADS data, a January SST index (PWT) for a portion of the equatorial Pacific (see Fig. 1) was obtained, following the procedure described by Wright (1984).

Time series of these indices are displayed in Fig. 2.

4. Methods

The techniques used in this study include correlation, empirical orthogonal function (EOF) analysis, and stepwise multiple regression (Statware Inc. 1986). Linear correlation serves to ascertain the associations of circulation parameters among themselves and with Nordeste rainfall. EOF analysis is used to produce index series from ship observations in the tropical Atlantic. These two operations form the preparations necessary for the subsequent stepwise multiple regression analysis. It is essential to distinguish between a "dependent" dataset or "training period," used to develop the prediction models, and an "independent" dataset or "forecasting period" reserved for the prediction experiments.

Empirical orthogonal function analysis was applied to extract information from the fields of meridional and zonal wind components in the tropical Atlantic in a way effective for the purposes of regression and prediction. The general characteristics of departures in the

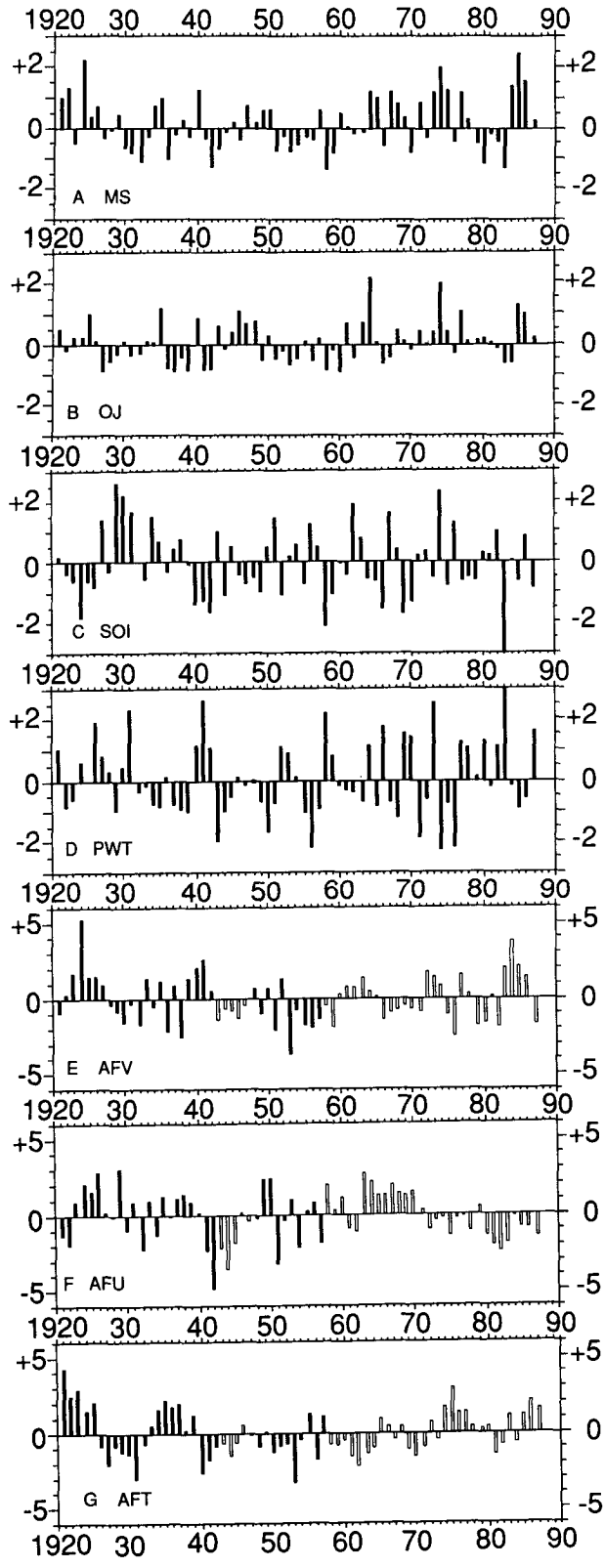

FIG. 2. Time series of elements used in regression and prediction: (a) Index of March–September rainfall in Nordeste MS (in 10^2, dimensionless); (b) Index of October–January rainfall in Nordeste OJ (in 10^2, dimensionless); (c) January pressure difference Tahiti minus Darwin SOI (normalized, dimensionless); (d) January sea surface temperature anomaly in equatorial Pacific (domain shown in Fig. 1) PWT (normalized, dimensionless); (e) Index of January meridional wind component over tropical Atlantic (−1st −4th +5th EOF) AFV (dimensionless); (f) Index of January zonal wind component over tropical Atlantic (+1st +2nd +4th EOF) AFU (dimensionless); (g) Index of January sea surface temperature in tropical Atlantic (−1st +3rd −5th EOF) AFT (dimensionless). In the series of AFV, AFU, and AFT (panels e, f, g) solid columns (1921–42 and 1948–57) indicate values obtained directly from EOF analysis and blank columns indicate years (1943–47 and 1958–87) calculated from the EOF analyses for 1921–42 and 1948–57 and the observed anomaly fields of v, u, SST in the tropical Atlantic.

wind and SST fields associated with rainfall anomalies in Northeast Brazil were described in section 2. In our earlier work (Hastenrath et al. 1984) we compiled regressor time series of v, u, and SST for critical but fixed domains. However, details of anomaly patterns associated with Nordeste rainfall variability may vary over the years. EOF analysis of the entire "dependent" data period and the tropical Atlantic as a whole is intended to mitigate against such pattern variations. For the first five EOF's of each element, correlation coefficients with MS were calculated. For each element, the three EOF's possessing the highest correlation with MS and each explaining at least 6% of the variance were multiplied by the sign of the correlation (with MS) and added to yield indices of v, u, and SST anomaly patterns, namely AFV = (−1st −4th +5th EOF), AFU = (+1st +2nd +4th EOF), and AFT = (−1st +3rd −5th EOF). This procedure was adopted in the belief that not merely the largest EOF is relevant for the present purposes, and that combining the three EOF's into a single index aids in limiting the number of regressors.

The combined factor score patterns of these indices (not shown here) are of interest in light of the discussion in section 2. AFV features negative factor scores over much of the tropical Atlantic, with particularly large negative values in the equatorial zone. This is consistent with the enhanced northeast trades and reduced cross-equatorial flow characteristic of abundant Nordeste rainy seasons. AFU possesses negative factor scores over part of the tropical North Atlantic and the Caribbean, consistent with the accelerated northeast trades described in section 2 as typical of copious Nordeste rainfall. Finally, AFT has particularly large negative factor scores in the tropical North Atlantic, also broadly consistent with the SST anomaly patterns described earlier.

Whereas the EOF analysis was limited to the "dependent" data period 1921–42 and 1948–57, values of the indices AFV, AFU, AFT for the years 1958–87, as well as 1943–47, were created as follows. For the EOF's selected as described above, the factor scores obtained for 1921–42 and 1948–57 were multiplied by the standardized observed anomalies of the years 1958–87 and 1943–47, so as to obtain the EOF patterns for each of these years. The EOF equivalents thus obtained were then combined as described above for the "dependent"

data period 1921–42 and 1948–57. The "reconstructed" values of the years 1958–87 are used for prediction and are marked in Figs. 2 and 3 by open symbols.

For the purposes of a special experiment (section 10), EOF analysis of the SST field was also performed over the entire record 1921–42 and 1948–87. As described above for AFT, an index AFΘ = (−1st +3rd −5th EOF) was constructed. The factor score pattern of this index resembles that of AFT. However, the 1958–87 portion of the series cannot be considered "independent," because the SST observations of this interval were included in the time period subjected to EOF analysis.

The application of stepwise multiple regression to climate prediction has been described in Hastenrath (1988b). That account is in part repeated here. Time series of the various circulation parameters presented in Fig. 2 and Table 1 serve as input to the stepwise multiple regression scheme. The general regression equations are

$$\mathrm{MS} = b_0 + b_1X_1 + b_2X_2 + \cdots b_nX, \qquad (1)$$

where MS is the predictand, b_i the coefficients of the regression model, X_i the regressors (variables in Fig. 2 and Table 1), and n the number of regressors employed in the regression model. A significance level α is specified where the retained regression coefficients are accepted as being different from zero.

A regression model is constructed on a "dependent" dataset, and the formula in the form of Eq. (1) thus obtained is then used to predict MS from an "independent" set of X_i values. In an operational setting, early portions of the record must be used to develop predictions of later years. In accordance with this practical reality, the time spans 1921–42 and 1948–57 form the dependent dataset and are used here to determine the regression coefficients, while the years 1958–87 form the independent dataset and are to be predicted.

Four statistics are used to measure the forecast skill: the correlation coefficient (CORR) between the forecast (MS′) and observed (MS), the root-mean-square error (RMSE), the absolute error (ABSE), and the bias (BIAS). The forecast record spans the 30 years 1958–87, and the last three statistics are compiled as follows:

TABLE 1. Matrices of correlation coefficients (in hundredths) between elements plotted in Fig. 2 for periods (a) 1921–42 and 1948–57, (b) 1958–87. Elements are

MS	Index of March–September rainfall in Nordeste
OJ	Index of October–January rainfall index in Nordeste
SOI	January pressure difference Tahiti minus Darwin
PWT	January sea surface temperature anomaly in equatorial Pacific (domain 2°–6°N, 90°–170°W; 2°N–6°S, 90°–180°W; 6°–10°S, 110°–150°W)
AFV	Index of January meridional wind component over tropical Atlantic
AFU	Index of January zonal wind component over tropical Atlantic
AFT	Index of January sea surface temperature in tropical Atlantic (EOF analysis makes a distinction between "dependent" and "independent" portions of record)
AFΘ	Index of January sea surface temperature in tropical Atlantic (EOF analysis makes no distinction between "dependent" and "independent" portions of record)

One and two asterisks indicate significance at the 5% and 1% levels, respectively

a. Period 1921–42 and 1948–57

	MS	OJ	SOI	PWT	AFV	AFU	AFT	AFΘ
MS	100							
OJ	+57**	100						
SOI	−21	−12	100					
PWT	−11	−07	−32	100				
AFV	+49**	+30	−52**	+32	100			
AFU	+39*	+23	+07	−18	+19	100		
AFT	+37	+30	−19	+34	+17	−00	100	
AFΘ	+37*	+19	−35	+01	+52**	+23	+58**	100

b. Period 1958–87

	MS	OJ	SOI	PWT	AFV	AFU	AFT	AFΘ
MS	100							
OJ	+61**	100						
SOI	+38*	+27	100					
PWT	−48**	−30	−74**	100				
AFV	+43**	+21	−05	+03	100			
AFU	+07	+03	−04	−10	−04	100		
AFT	+35	+25	−04	−15	+01	−29	100	
AFΘ	+66**	+48**	+10	−15	+23	−28	+65**	100

$$\text{RMSE} = [\sum^{n} (\text{MS}' - \text{MS})^2/n]^{1/2}$$

$$\text{ABSE} = \sum^{n} |\text{MS}' - \text{MS}|/n$$

$$\text{BIAS} = \sum^{n} (\text{MS}' - \text{MS})/n$$

where the summation extends over the n forecast years. Overall it was found that forecast performance is better for the latter two decades of the total independent record 1958–87. This may be due both to increased intrinsic predictability as well as to better quality control of observations in recent decades. Accordingly, the aforementioned four measures of forecast performance are evaluated separately over the intervals 1958–87 (N = 30 years) and 1968–87 (N = 20 years).

Regarding the significance testing of correlation coefficients, it should be noted that geophysical time series are not, as a rule, serially independent. Therefore, Quenouille's (1952, p. 168) method was used to account for the reduction of the effective number of degrees of freedom due to persistence, based on the lag autocorrelation of the time series.

5. Year-to-year variability of circulation and rainfall

Figure 2 illustrates the interannual variation of Nordeste rainfall and of selected elements indicative of the large-scale circulation, while Table 1 presents matrices of correlation between these variables. Of interest here is Table 1a, pertaining to the "dependent" part of the record (1921–42, 1948–57).

Figure 2a (MS) exhibits large year-to-year changes of March–September rainfall, the catastrophic drought years 1958 and 1983 being particularly conspicuous. Comparison with Fig. 2b (JO) reveals a remarkably parallel behavior of the preseason (October–January) precipitation. Table 1a shows a highly significant positive correlation between these two index series. The remainder of Fig. 2 and Table 1a illustrates concomitant variations in other pertinent indices. Inasmuch as Nordeste droughts tend to coincide with El Niño events (see section 2), the inverse relationship of MS with PWT appears plausible, but not that with SOI; however, both correlations are small. Here MS possesses much stronger correlations with the Atlantic circulation indices AFV, AFU, and AFT. As discussed in section 2, these associations are consistent with the general circulation anomalies known to characterize extreme climatic events in Northeast Brazil. Most promising for the construction of regression models are elements possessing not only high correlations with MS but also small correlations with other regressors.

Of further interest for prediction purposes is the temporal stability of relationships. This stability is illustrated in Fig. 3 by using 21-year sliding correlations and in Tables 1a and b by correlation coefficients for the "dependent" and "independent" portions of the record. OJ, and to a lesser extent AFV and AFT, stand out for their remarkable stability over much of this century.

6. Prediction from limited observations

Table 2 summarizes all prediction models discussed in this and the following section, but details are presented only for the most interesting of these in Tables 3 and 4 and Figs. 4–6.

An important practical requirement for operational climate prediction is the availability of input observations on a quasi-real-time basis. Of the six predictor candidates shown in Figs. 2–3 and Table 1, OJ, SOI, PWT, and AFT could now be obtained shortly after the end of January. Therefore, the present section explores the potential of seasonal rainfall prediction for the Nordeste based solely on this information. Pertinent to this section are the regression models 1 to 4 (Tables 2–4). In model 1, OJ was provided as input and accepted at a significance level α = 5%. The variance of MS explained by the regression model for the dependent portion of the record (1921–42, 1948–57) is 34%. Applying model 1 in a predictive mode on the independent portion of the record allows one to explain 37% of the observed MS variance for 1958–87 and 52% for 1968–87 (Table 2). Other measures of forecast performance are summarized in Table 4; in particular, the bias is small. Figure 4 compares a scatter plot of regressed versus observed rainfall for the training period with a plot of forecast versus observed values for the forecast interval; the fit is even better for the latter. This may reflect both intrinsically closer relationships and better quality control of observations in the course of recent decades.

In model 2 OJ and SOI serve as input. Here OJ enters again at α = 5%, but for SOI to enter α had to be relaxed to 37%. The variance explained by the regression on the dependent dataset is 36% or a little larger than for model 1. For the independent portion of the record, however, the variance of observed MS explained by the forecast is only 30% for the 1958–87 period and 48% for 1968–87. Comparison of models 1 and 2 illustrates the disadvantage of including weak predictors: while these may enhance the model fit for the dependent dataset, they may worsen forecast performance.

In model 3, OJ, PWT, and AFT were offered as input. Here OJ and AFT entered at α values of 5% and 15%, respectively, but PWT was not accepted even when relaxing α to 90%. The variance explained by the regression in the dependent dataset is 40%, a further increase over models 1 and 2. The prediction for the independent part of the record for 1958–87 shows some improvement, explaining 41% of the observed variance for that period, and 50% of the variance observed during 1968–87. Scatter plots of model 3 for the training and forecast periods are displayed in Fig. 5. As for

TABLE 2. Summary of regression models. Shown in bold print are models 1, 3, and 6, for which scatter diagrams are displayed in Figs. 4–6. Regressors are OJ (dimensionless), SOI (10^{-1} mb), PWT (10^{-2} °C), AFV (dimensionless), AFU (dimensionless), AFT (dimensionless), and AFΘ (dimensionless), as defined in Table 1 and in text. VAR rgrs = percentage variance (of MS) explained by regression model; VAR fcst = percentage variance explained by forecast. Here α is confidence level at which the regressors were accepted by model; a dash indicates that element was not provided as input.

Model no.	Training period	Entrance level α [%]							VAR rgrs [%]	VAR fcst [%]	
		OJ	SOI	PWT	AFV	AFU	AFT	AFΘ		1958–87	1968–87
a. Fixed training period											
1	**1921–42, 1948–57**	**5**	—	—	—	—	—	—	**34**	**37**	**52**
2	1921–42, 1948–57	5	37	—	—	—	—	—	36	30	48
3	**1921–42, 1948–57**	**5**	—		—	—	**15**	—	**40**	**41**	**50**
4	1921–42, 1948–57	5	—	70	—	—	—	—	35	40	53
5	1921–42, 1948–57	5			5			—	47	45	66
6	**1921–42, 1948–57**	**5**		**15**	**5**			—	**51**	**56**	**71**
7	1921–42, 1948–57	5			5	25	25	—	56	50	66
b. Optimal length of training period											
8	1923–42, 1948–57 to 1938–42, 1948–57	5	—	—	—	—	—	—	33–57	37	53
9	1923–42, 1948–57 to 1938–42, 1948–57	5	—	20–60	20–60	—	—	—	55–62	51–58	51–71
c. Updating											
10	1921–42, 1948–57 to 1955–86	5	—	—	—	—	—	—	31–46	33	48
11	1921–42, 1948–57 to 1955–1986	5		5–30	5–40				49–63	50	61
d. Pseudoprognosis											
12	1921–42, 1948–57	—	—	—	—	—	—	5	14	44	42
13	1921–42, 1948–57	—	—	—	—	—	5	—	14	12	12

model 1 (Fig. 4), the fit of forecast versus observed values in the independent dataset is even better than that of regressed versus observed values in the training period. As indicated above, both intrinsically closer relationships and better quality control of observations in recent decades may contribute to this improvement.

Model 4 is an experiment of incorporating PWT without competition by regressors other than OJ. This could be accomplished only by relaxing α to 70%. However, this model offers no improvement over model 3 (Table 2).

Of interest in the following, therefore, are models 1 and 3 (Figs. 4 and 5), which are superior to the other two.

7. Prediction with unrestricted data input

This section is concerned with the predictability of Nordeste rainfall regardless of the timely availability of the predictors. Accordingly, OJ, SOI, PWT, AFT, AFV, and AFU (Fig. 2, Table 1) all serve as input to the regression experiments. Pertinent here are regression models 5–7 (Table 2).

TABLE 3. Coefficients a and b_n of regression models based on Eq. (1). Shown in bold print are models 1, 3, and 6, for which scatter diagrams are displayed in Figs. 4–6. Regressors are as defined in previous tables and text.

Model no.	a	Coefficients b_n						
		OJ	SOI	PWT	AFV	AFU	AFT	AFΘ
1	**+18.77**	**+0.80**						
3	**+18.29**	**+0.81**					**−10.22**	
5	+20.19	+0.76			+14.32			
6	**+19.28**	**+0.71**		**−0.23**	**+17.82**			

TABLE 4. Forecast performance. Shown in bold print are models 1, 3, and 6, for which scatter diagrams are displayed in Figs. 4–6. Predictors are as defined previously. CORR = correlation coefficient between predicted and observed rainfall, in hundredths (one and two asterisks indicate significance at the 5% and 1% levels, respectively); RMSE = root-mean-square error; ABSE = absolute error; BIAS = bias.

Model no.	Forecast period	CORR	RMSE	ABSE	BIAS
1	**1958–87**	**+61****	**76**	**63**	**+4**
	1968–87	**+72****	**74**	**58**	**+0**
3	**1958–87**	**+64****	**73**	**60**	**+3**
	1968–87	**+71****	**72**	**55**	**+4**
5	1958–87	+67**	70	58	−4
	1968–87	+81**	65	52	+2
6	**1958–87**	**+75****	**63**	**54**	**−1**
	1968–87	**+84****	**58**	**49**	**−4**
12	1958–87	+66**	81	67	−13
	1968–87	+65**	88	74	−15
13	1958–87	+35	94	78	−26
	1968–87	+35	100	83	−31

In model 5 both OJ and AFV were admitted at a significance level $\alpha = 5\%$. The MS variance explained by model 5 in the dependent dataset is 47%, larger than for models 1–4. Forecasts with the independent dataset explain 45% of the variance for 1958–87 and 66% for 1968–87. The other measures of forecast performance summarized in Table 4 likewise show an overall improvement over models 1 and 3.

For model 6, OJ and AFV entered at $\alpha = 5\%$ as in model 5, but in addition PWT was admitted at a significance level $\alpha = 15\%$. The variance explained by the model in the dependent dataset is 51%, an increase over model 4. Forecasts with the independent dataset explain 56% of the variance for 1958–87 and 71% for 1968–87, a distinct improvement over model 5. The other measures of forecast performance summarized in Table 4 are also somewhat more favorable than for model 5. Scatter plots for the training and forecast periods are presented in Fig. 6. As for models 1 and 3 (Figs. 4 and 5), the fit of forecast versus observed values in the independent portion of the record is better than that of regressed versus observed values in the training period, presumably for reasons similar to those proposed in section 6.

In model 7, OJ and AFV were admitted at $\alpha = 5\%$, and AFU and AFT at $\alpha = 25\%$. The variance explained by the model in the dependent dataset is 56%, higher than for any of the preceding models. However, the inclusion of the weak regressors AFU and AFT resulted in a forecast performance distinctly inferior to model 6 (Table 2).

From the discussion in this section and the evidence summarized in Table 2, model 6 stands out as the most powerful method for predicting Nordeste rainfall. Information on the (meridional component of the) wind field over the tropical Atlantic (AFV) and SST conditions in the equatorial Pacific (PWT) substantially enhances the forecast performance compared to model 1 depending solely on preseason rainfall (OJ).

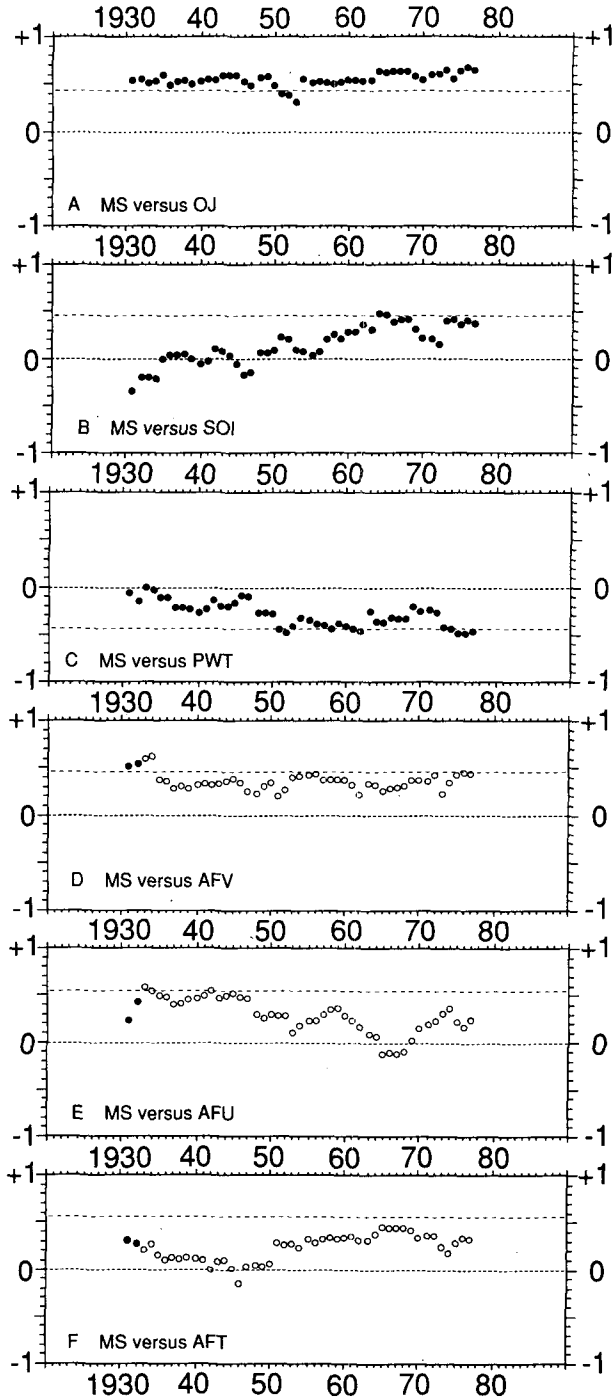

FIG. 3. Twenty-one year sliding mean correlations between March–September rainfall in Northeast Brazil (MS) and indicated elements: (a) OJ; (b) SOI; (c) PWT; (d) AFV; (e) AFU; (f) AFT. For the series AFV, AFU, AFT (panels e, f, g), open circles indicate time intervals containing at least one of the years (1943–47, 1958–87) shown in Fig. 2 by blank columns. Broken lines indicate, approximately, the 5% significance level.

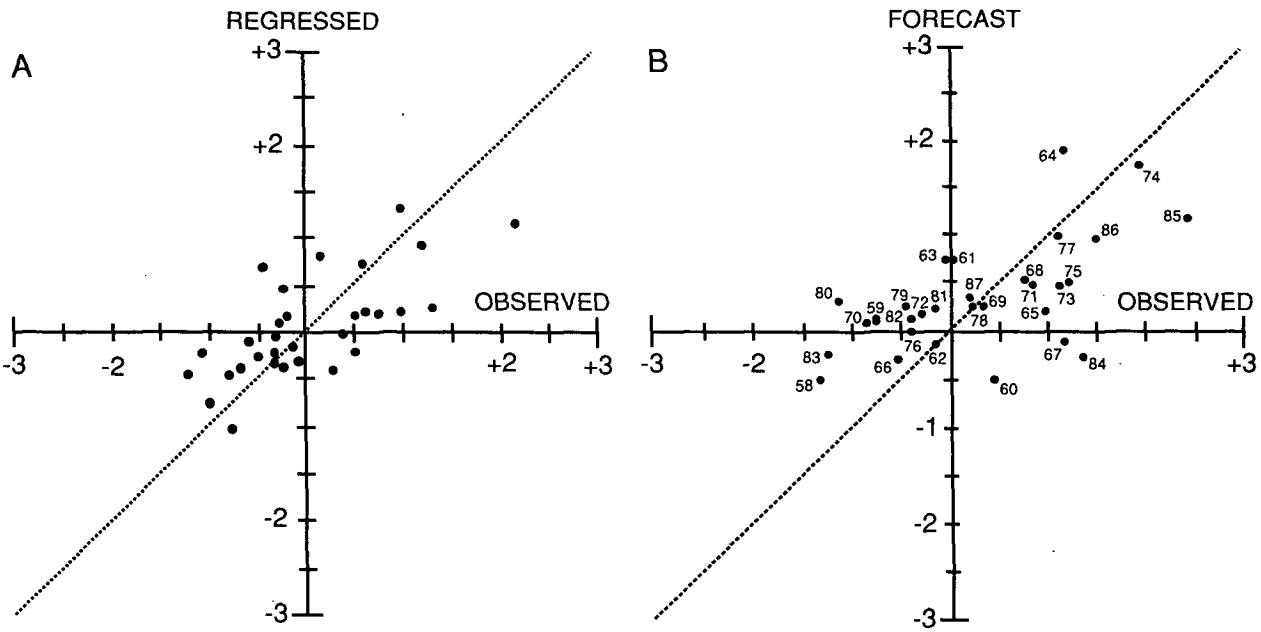

FIG. 4. Scatter diagram of March–September rainfall index (in 10^2) for model 1 (see Tables 2 and 4). Numbers indicate the years, and the dotted line is at a 45° angle. (a) Regression, period 1921–42 and 1948–57; (b) forecast, period 1958–87. For 1958–87 the CORR = +0.61**, RMSE = 76, ABSE = 63; BIAS = +4. For 1968–87 the CORR = +0.72**, RMSE = 74, ABSE = 58; BIAS = +4. Correlation coefficients are significant at the 1% level.

8. Optimal length of training period

The length of the regression base period most suitable for climate prediction is an issue addressed in earlier work (Hastenrath 1987, 1988b). While long series seem desirable to establish statistical relationships among variables, the most recent portion of the record may be most pertinent to the current circulation setting. These competing considerations need to be pursued from observations.

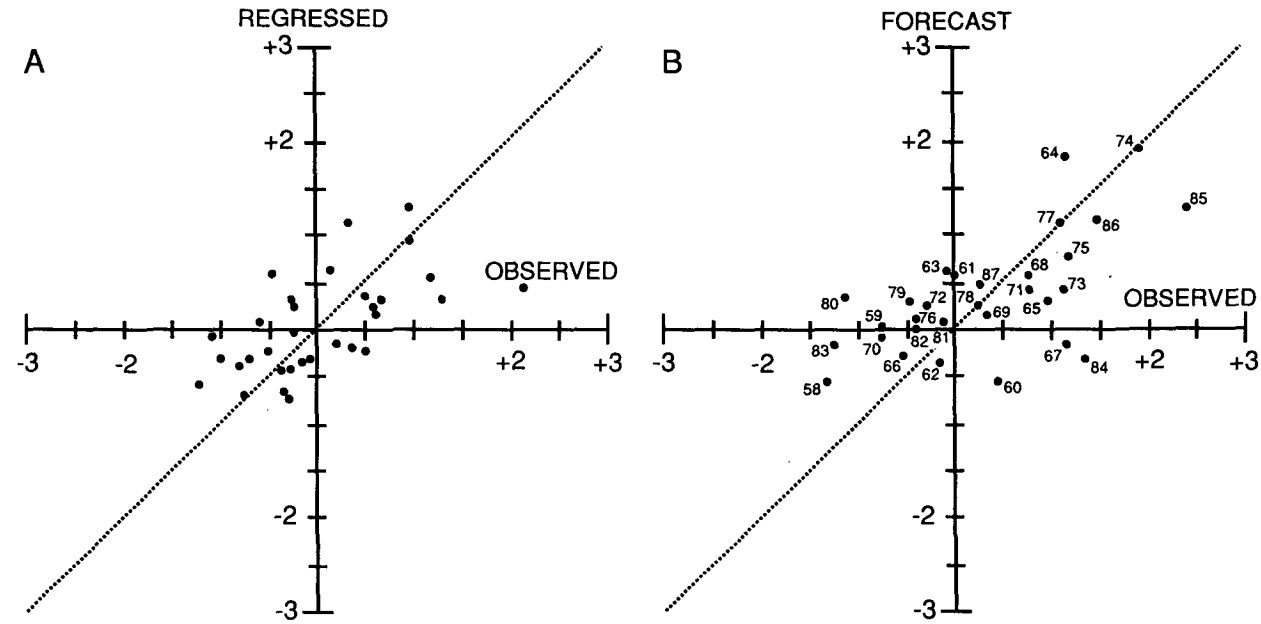

FIG. 5. Scatter diagram of March–September rainfall index (in 10^2) for model 3 (see Tables 2 and 4). Numbers indicate the years, and the dotted line is at a 45° angle. (a) regression, period 1921–42 and 1948–57; (b) forecast, period 1958–87. For 1958–87 the CORR = +0.64**, RMSE = 73, ABSE = 60; BIAS = +3. For 1968–87 the CORR = +0.71**, RMSE = 72, ABSE = 55; BIAS = +4. Correlation coefficients are significant at the 1% level.

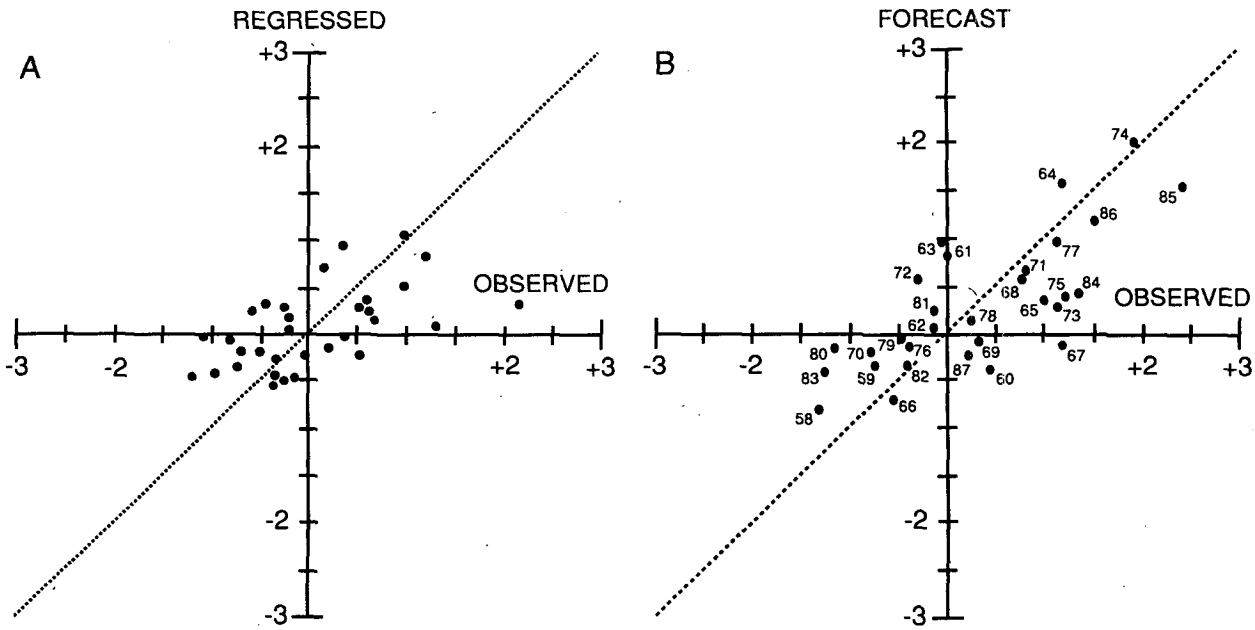

FIG. 6. Scatter diagram of March–September rainfall index (in 10^2) for model 6 (see Tables 2 and 4). Numbers indicate the years, and the dotted line is at a 45° angle. (a) Regression, period 1921–42 and 1948–57; (b) forecast, period 1958–87. For 1958–87 the CORR = +0.75**, RMSE = 63, ABSE = 54, BIAS = −1; For 1968–87 the CORR = +0.84**, RMSE = 58, ABSE = 49, BIAS = −4. Correlation coefficients are significant at the 1% level.

In accordance with the discussion in sections 6 and 7, models 1 and 6 are the types examined here. The versions corresponding to varying record length are denoted as models 8 and 9, respectively, in Tables 2 and 5. For both types, regression formulae were constructed from base periods of varying length but all ending in 1957. These models were then used to predict MS during 1958–87. Results are summarized in Table 5.

For models 8, using OJ as sole input (Table 5a), the forecast performance does not deteriorate substantially for base periods ranging from 30 to 20 years. Accordingly, a base period of 20 years appears appropriate. Consistent with this result, recall that OJ has maintained a remarkably stable relationship with MS since early this century (Fig. 3a).

Regarding model 9 (Table 5b), comparison of the various performance measures for the two forecast periods suggests little deterioration from a base period of 30 to 20 years. From Figs. 3c and 3d it is seen that the correlations of MS with AFV and PWT have remained relatively stable, although to a lesser extent than the MS–OJ correlations.

9. Updating of base period

Inasmuch as the general circulation setting may vary with time, the most recent portion of the record may be most appropriate for climate prediction. This conjecture has been considered in some earlier regional studies (Hastenrath 1987, 1986b) and is examined again here with reference to versions of models 1 and 6 denoted in Table 2 as models 10 and 11, respectively. In both cases, the training period is 32 years, ending immediately preceding the year to be predicted. Predictions are made for the 30 years 1958–87. Thus, the training period 1921–42 and 1948–57 serves to predict the MS of 1958; the training period 1922–42 and 1948–58 is used to predict 1959, and so on.

Regarding forecasts with OJ as the sole predictor, models 10 should be compared to model 1. Table 2 shows that updating yields no improvement in forecast

TABLE 5. Appraisal of predictions of March–September rainfall in Nordeste (index MS) with (a) models 8, and (b) models 9 during 1958–87. CORR = correlation coefficient between predicted (MS') and observed (MS) rainfall (one and two asterisks indicate significance at the 5% and 1% levels, respectively). RMSE = root-mean-square error; ABSE = absolute error; BIAS = bias.

No. of year	Training period	CORR	RMSE	ABSE	BIAS
a. Models 8					
30	1923–42, 1948–57	+61**	76	63	−6
25	1928–42, 1948–57	+61**	76	63	−11
20	1933–42, 1948–57	+61**	76	62	−3
15	1938–42, 1948–57	+61**	76	62	−2
b. Models 9					
30	1923–42, 1948–87	+76**	64	55	−15
25	1928–42, 1948–87	+68**	72	59	−20
20	1933–42, 1948–87	+72**	67	57	−12
15	1938–42, 1948–87	+71**	68	57	−9

performance. In fact, the percentage of MS variance explained by "updated" forecasts is distinctly less. Concerning forecasts with OJ, AFV, and PWT, models 11 should be compared to model 6 (Table 2). Again it is found that updating yields no improvement in forecast performance. Interesting in this context is the result pointed out in section 4 that the best forecast performance is achieved for the two most recent decades of the record.

10. Pseudoprognosis

Throughout the prediction experiments discussed in sections 6–9, a distinction was made between "dependent" and "independent" datasets. This discrimination was maintained in the EOF analysis of atmospheric and oceanic fields in the tropical Atlantic leading to the construction of the index series AFV, AFU, and AFT. By contrast, some recent climate prediction efforts using EOFs (Parker et al. 1988; Owen and Ward 1989) have blurred this distinction. Possible consequences of such approaches will be demonstrated in the present section.

Pertinent are the models 12 and 13 (Tables 2 and 4), both using as input only information on the SST field in the tropical Atlantic. For both the significance level of $\alpha = 5\%$ is used. In model 13, the sole regressor is AFT, and "dependent" and "independent" datasets are kept completely separate. Thus model 13 serves as a "control" experiment. By contrast, model 12 has as its sole regressor AFΘ. In the EOF analysis leading to the construction of this index, the portions of the record up to 1957 and from 1958 onward are not kept separate, as described in section 4. Consequently, the post-1957 portion of the AFΘ series is not independent of the pre-1958 portion. Therefore, the experiment with model 12 is called a "pseudoprognosis" here.

According to Tables 2 and 4, in the pseudoprognosis model 12 the calculated values explain 44% of the observed MS variance for 1958–87 and 42% for 1968–87, and RMSE, ABSE, and BIAS are relatively small. A comparison with the "control" model 13 is sobering, however. For the latter, the percentage of MS variance explained by predictions with the independent data is only 12% for both the 1958–87 and 1968–87 periods, and the other measures given in Table 4 are also less favorable than for model 12. These results serve to caution against unduly optimistic expectations when dependent and independent datasets are merged.

11. Ancillary indicators

Throughout this study, analyses concentrated on elements with records existing over 1921–87. The present section is concerned with observations also pertinent to Nordeste rainfall variations but with series as yet too short to be incorporated into a comprehensive empirical treatment.

A novel data bank of highly reflective clouds (HRC) beginning in 1971 and covering the global tropics (García 1985) documents the near-equatorial zone of maximum convection and thus intensity and latitudinal position of the ITCZ (Hastenrath 1990). Amount (AMT) and latitude position (LAT) of the HRC maximum at 30–60°W during January were evaluated to explore their potential for prediction purposes. Over the period 1971–87, MS possesses a correlation of −0.29 with LAT, and of +0.47 with AMT. (Consistent with the discussion of the meridional wind field and ITCZ position over the Atlantic, LAT and AFV are correlated at −0.72, significant at the 1% level.) These satellite-derived measures of ITCZ behavior could conceivably be produced in near real-time and may thus serve as qualitative indicators of the imminent rainy season, ancillary to the quantitative prognosis using models 1, 3, or 6 (Tables 2–4). HRC may also become a useful ingredient to a quantitative prediction scheme, as its period of record expands.

Regarding the wind field, the real-time monitoring at St. Peter and Paul's Rocks (0°55′N, 29°26′W) since the early 1980s (Garzoli and Katz 1984) is of obvious interest. For March–April, but not January, the short record indeed shows associations with Nordeste rainfall consistent with the discussion in section 2. Although the location in the western equatorial Atlantic is strategically favorable, measurements at a single site must be regarded as inferior to monitoring the basinwide wind field. Similarly, the predictive potential of the long-term wind measurements on the island of Fernando de Noronha remains to be ascertained, should these records become accessible.

Maps of 850 mb wind anomalies are being produced in near real-time by the Climate Analysis Center, NOAA. The 850 mb level is less informative than the surface and the record is short, however.

Fields of SST anomalies in the tropical Atlantic and Pacific are likewise being produced in near real-time by NOAA (Rao 1989). These may be of practical importance in the light of sections 6 and 7, which indicate that Atlantic SST information may improve on predictions based solely on the preseason rainfall OJ (models 3 versus 1; Tables 2 and 4) and that Pacific SST serves to improve performance compared to using only OJ and AFV as predictors (models 6 versus 5; Tables 2 and 4).

12. Conclusions

The Sêcas of Northeast Brazil have long been recognized as climatic disasters with extraordinary social and economic impact. Accordingly, the interannual rainfall variability in Brazil's Nordeste represents a prime "target of opportunity" for the development of climate prediction methods. With this motivation we ascertained in earlier work the intrinsic predictability of Nordeste rainfall anomalies from antecedent cir-

culation departures (Hastenrath et al. 1984). The present study expands the basis for operational application.

Central to this endeavor is the creation of "clean" and up-to-date rainfall index series, both for the predictand (MS) and a predictor (OJ). The real-time processing of these selected raingage measurements involves only relatively modest material resources; with this prerequisite alone it will be possible to forecast a substantial portion of the interannual rainfall variance on an operational basis.

Indeed, about half of the interannual variability of March–September (MS) precipitation can be predicted from the preseason rainfall (OJ) in the Nordeste itself (model 1; see section 6). Preseason rainfall (OJ) must, therefore, form the cornerstone of Nordeste climate prediction, even if the predictor basis is expanded to include information on the wind and SST fields in the tropical Atlantic and Pacific. Most important among such additional input is the meridional component of the surface wind field over the Atlantic. The 850 mb wind observations mentioned in section 11 are not so well suited as input to quantitative prediction models, although they may conceivably serve as ancillary qualitative indicators. Monitoring the Atlantic surface wind field is needed to produce a meridional wind index equivalent to AFV in near real time.

In a range of empirical and modeling studies, much attention has been given to SST. As shown in sections 6 and 7, however, the Atlantic SST field plays only a subordinate role as a predictor in the company of the preseason rainfall (OJ) and the Atlantic wind field. Nevertheless, the availability of Pacific and Atlantic SST fields in near real time makes it possible to update the SST indices PWT and AFT and to use them as predictors in the company of OJ, achieving some increase in forecast performance (models 3 and 4 versus 1; Tables 2 and 4).

From a series of experiments it is suggested that 20–30 years are adequate as a training period and that updating offers no advantage over a fixed training period. Predictability of Nordeste rainfall anomalies has increased over the past three decades. Finally, the portion of observed MS variance explained by the prediction can be raised to about 70% if the equatorial Pacific SST (PWT) and the meridional component of the Atlantic surface wind field are also provided as input (model 6). Barring this more costly undertaking, the present study underlines the importance of maintaining a near real-time processing of a selected conventional raingage network for the purposes of climate monitoring and prediction.

Acknowledgments. This study was supported by National Science Foundation Grant ATM-8722410. I thank the Fundação Cearense de Meteorologia, Fortaleza, and Aydil Gusmão, Superintendencia do Desenvolvimento do Nordeste, Recife, for cooperating in the creation of the new rainfall index series. Larry Greischar and Neil Simmons assisted me with the computations and Dan Sheehan with the graphics.

REFERENCES

Garcia, O., 1985: *Atlas of Highly Reflective Clouds for the Global Tropics: 1971–83.* NOAA-ERL, Boulder, 365 pp. [NTIS PB 87-129 169]

Garzoli, S. L., and E. J. Katz, 1984: Winds at St. Peter and Paul Rocks during the first SEQUAL year. *Geophys. Res. Lett.*, **11,** 715–718.

Hastenrath, S., 1986: On climate prediction in the tropics. *Bull. Amer. Meteor. Soc.*, **67,** 692–702.

——, 1987: Predictability of Java monsoon rainfall anomalies: A case study. *J. Climate Appl. Meteor.*, **26,** 133–141.

——, 1988a: *Climate and Circulation of the Tropics.* Reidel, 455 pp.

——, 1988b: Prediction of Indian monsoon rainfall: Further exploration. *J. Climate,* **1,** 298–304.

——, 1990: The relationship of highly reflective clouds to tropical climate anomalies. *J. Climate,* **3,** 353–365.

——, and L. Heller, 1977: Dynamics of climatic hazards in Northeast Brazil. *Quart. J. Roy. Meteor. Soc.*, **103,** 77–92.

——, and P. J. Lamb, 1977: *Climatic Atlas of the Tropical Atlantic and Eastern Pacific Oceans.* University of Wisconsin Press, 117 pp.

——, M-C. Wu and P.-S. Chu, 1984: Towards the monitoring and prediction of Northeast Brazil droughts. *Quart. J. Roy. Meteor. Soc.*, **110,** 411–425.

——, L.-C. Castro and P. Aceituno, 1987: The Southern Oscillation in the tropical Atlantic sector. *Contrib. Atmos. Phys.*, **60,** 447–463.

National Climate Program Office, 1988: National Climate Program Five Year Plan 1989–93, NOAA, Rockville, MD, 48 pp.

Norusis, M., 1988: SPSS-X Advanced statistics guide. SPSS, 2nd ed., 527 pp.

Oort, A. H., Y. H. Pan, R. W. Reynolds and C. F. Ropelewski, 1987: Historical trends in the surface temperature over the oceans based on the COADS. *Climate Dynamics,* **2,** 29–38.

Owen, J. A., and M. N. Ward, 1988: Forecasting Sahel rainfall. *Weather,* **44,** 57–64.

Parker, D. E., C. K. Folland and M. N. Ward, 1988: Sea surface temperature anomaly patterns and prediction of seasonal rainfall in the Sahel region of Africa in Recent climate change, S. Gregory, Ed., Belhaven Press, 166–178.

Quenouille, M. A., 1952: *Associated Measurements.* Butterworths, 242 pp.

Rao, D. B., 1989: A review of the program of the Ocean Products Center. *Weather and Forecasting,* **4,** 427–443.

Statware, Inc., 1986: STAT 80 TM for Apple Macintosh, professional version, release 2.10, version 1.0, 521 pp. [Available from Statware Inc., P.O. Box 510881, Salt Lake City, UT 54151.]

World Meteorological Organization, 1980: Outline plan and basis for the World Climate Programme, 1980–83. WMO-NO. 540, 64 pp.

Wright, P. B., 1984: Relationships between indices of the Southern Oscillation. *Mon. Wea. Rev.*, **112,** 1913–1919.