

## A probabilistic model for predicting seasonal rainfall in semi-arid lands of northeast Brazil

**BERNARDO B. DA SILVA<sup>1</sup>, CÉLIA C. BRAGA<sup>2</sup>,  
SUZANA M. G. L. MONTENEGRO<sup>1</sup>, VICENTE DE PAULO R. DA SILVA<sup>2</sup>,  
LEIDJANE M. M. DE OLIVEIRA<sup>1</sup> & FRANCISCO DE ASSIS S. DE SOUSA<sup>2</sup>**

*1 Federal University of Pernambuco – Acadêmico Hélio Ramos Street, CFCH / Department of Geographic Sciences, 6th Floor, 50670-901 Recife, PE, Brazil*

[bbdasilva.ufpe@gmail.com](mailto:bbdasilva.ufpe@gmail.com)

*2 Federal University of Campina Grande – Aprígio Veloso Avenue, 882. CTRN, Department of Atmospheric Sciences, Block CL – 58492-900 Campina Grande, PB, Brazil*

**Abstract** In most of the northeast region of Brazil rainfall is relatively low, presenting significant inter-annual fluctuations, especially when compared to rainfall in other areas of Brazil. Moreover, evaporative rates (like the ones found in the northeast semi-arid region) are too high, sometimes reaching over 2800 mm annually. Owing to such a climate character, very large areas in northeast Brazil are subjected to recurrent droughts. This paper presents a methodology for the prediction of seasonal rainfall in semi-arid lands of northeast Brazil. A total of 72 raingauge stations of Paraíba State, and 84 in Ceará State were employed, all of them distributed in three and seven homogeneous areas, respectively. A rainy season with different subdivisions was established for each homogeneous area. The  $z_i$  proportions – the ratio between the cumulative rainfall of the first rainy season period and the rain that falls during the whole rainy season were made to fit the Beta probabilistic model used for calculating the second and eighth deciles and the probability of rainfall above the average rainfall for the second period of the rainy season. The performance of the prognostic model for individual stations of Paraíba State in the period 1996–2000 was evaluated. In the period 1996 to 2000, with rainfall above average, the error was less than 20%. The methodology adopted proved very accurate for forecasting droughts in northeast Brazil.

**Key words** seasonal forecast; Beta probability model; drought; Brazil

### INTRODUCTION

The weather and climate in northeast Brazil are highly influenced by the phenomena associated with various scales: from the planetary scale to the small scale, represented by isolated convection. The atmospheric circulations associated with anomalies in the sea surface temperature (SST) – such as the ones that most characterize the phenomenon El Niño, the Southern Oscillation (ENSO), the Atlantic dipole, the Atlantic subtropical anticyclones, the inter-tropical converging zone (ITCZ) and the 30–60 days oscillation belong to the planetary scale. On a synoptic scale, the austral frontal systems or the converging zones originating from them, the high troposphere cyclonic vortexes and the eastern disturbances have been given considerable relevance. In northeast Brazil, mesoscale phenomena occur, the systems coming from the ITCZ (mesoscale convective complexes and instability lines that build up along the northern coast of northeast Brazil) and the circulations occasioned by thermo contrast between solid and liquid surfaces. In most parts of the region, rainfall is relatively low, presenting significant inter-annual fluctuations, especially when compared to rainfall in other areas of Brazil. Moreover, evaporative rates like the ones found in the Northeast semi-arid are too high, sometimes reaching over 2800 mm annually. Owing to such a climatic character, very large areas in northeast Brazil are subjected to recurrent droughts.

In spite of the great advances in the last 10 years, rendering the dynamic and statistical models far more efficient, too much divergence still persists among climate forecasters as regards their predictions, which serves to demonstrate the high degree of complexity that climate forecasting entails, at the same time showing that meteorology still has a long way to go in order to satisfy the scientific community where accuracy and anticipation of prognostics are concerned. While one tries to reach the accuracy and the anticipation needed, one should apply the existing models so as to provide the answers that would implement the planning and the decision-making procedures aimed at minimizing the impact of undesirable climatic conditions on the region, mainly those conditions affecting farming and water resources.

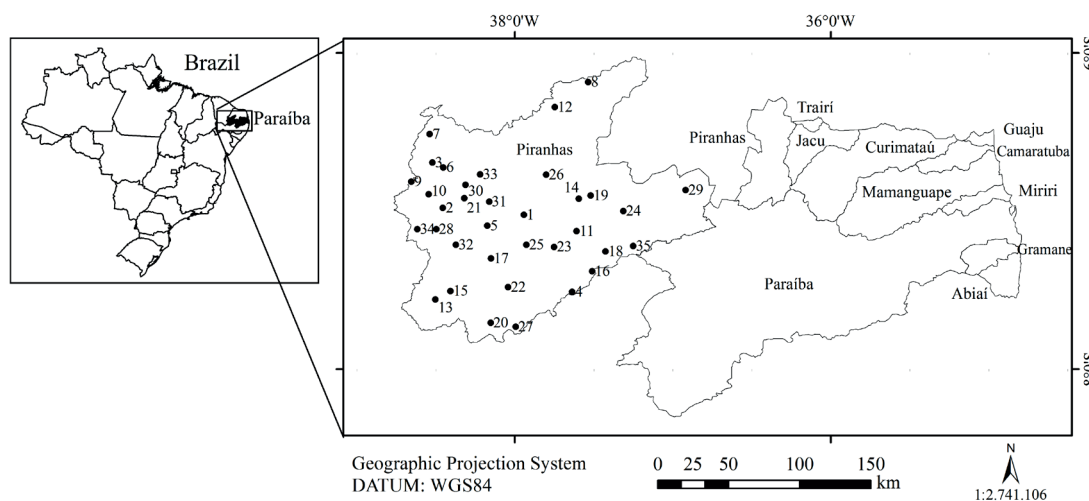
Silva (1985) suggests a new probabilistic technique that differs a lot from the ones that have been developed so far (Ward and Folland 1991). However, this new technique has been successfully

employed in a variety of pluvial regimes in northeast Brazil. The same model was employed by Azevedo *et al.* (1998) in the State of Ceará, aiming at predicting rainfall for the second half of the rainy season (from 20 March to 30 June) for each homogeneous area in Ceará State. The methodology proposed by Silva rendered itself rather effective in assessing the rainfall minimum and maximum values for the second half of the rainy season in various homogeneous micro-regions in Ceará, mainly the minimum values for the 1960–1969 period. Santos *et al.* (2002), employing the Silva (1985) method, used monthly totals from 34 raingauge stations on the eastern coast of northeast Brazil, which extends from Rio Grande do Norte State down to Sergipe State. This methodology proved rather effective at predicting the maximum and minimum rainfall values for the second period of each rainy season for both the northern and central region, whereas for the southern region it proved far less effective.

The present paper aims at producing a methodology based on the Silva (1985) climatic model for predicting rainfall in semi-arid lands of northeast Brazil. As for the specific aim, a technique is presented to determine the probability of occurrence of rainfall higher than the climatic value for the RS second period.

## DATA AND METHODS

**The study region** The region selected for illustration of the study (Fig. 1) were the states of Paraíba and Ceará, on account of the existing facilities in the States, such as Laboratório de Meteorologia, Recursos Hídricos e Sensoriamento Remoto – LMRS in Campina Grande city, and FUNCEME – Fundação Cearense de Meteorologia, located in Fortaleza city. The LMRS and FUNCEME conduct weather and climate monitoring for Paraíba and Ceará states, respectively, and benefit from Silva (1985) findings. Moreover, there exist in Paraíba a number of different pluvial regimes, which makes it easier to evaluate the effectiveness of this technique.



**Fig. 1** Study region with the raingauge stations and water basins of Paraíba State.

**The data used** The data, on which the present survey was based were all collected from 72 raingauge stations in Paraíba and 84 in Ceará, during the period 1910–1990. The data are available in digital archives at the Academic Unit of Atmospheric Science, Federal University of Campina Grande. The data used for validation of the model, that is, from 1996 to 2001, were provided by the LMRS. The stations consulted – with their geographic coordinates – and the study mesoregions are found in Silva *et al.* (2004) and Azevedo *et al.* (1998).

**Definition of study mesoregions** Three homogeneous areas for the State of Paraíba, here defined as Western (Sertões), Central (Cariri) and Eastern (Zona da Mata and Litoral), and seven for the State of Ceará, have been considered for the present work. These homogeneous areas were identified and have been used by the LMRS for a number of purposes, including weather and climate forecast.

In the case of the Silva (1985) model, it may be seen in the next section, the rainy season (RS) and their two periods for all localities must be identified. Based on the “Atlas Climatológico do Estado Paraíba” (Varejão-Silva *et al.* 1987) the RS for the Western, Central and Eastern homogeneous areas includes the periods from January to June, January to July and from January to August, respectively. For each mesoregion, the rainy season initial periods with differing duration were taken into account. This would determine a prognostic for two or more distinct RS occasions. The two different periods used for each particular RS at the Sertões mesoregion are found in Table 1. In order to show another way of exploring the technique, for the State of Ceará only the RS of the period from January to June was considered, comprising all the stations of the same homogeneous area.

**Table 1** Rainy Season – RS and the first and second periods for the Sertões (Western) mesoregion, and Ceara State.

Rainy Season	First Period X	Second Period Y	Mesoregion
RS1	JF	MAMJ	Western
RS2	JFM	AMJ	Western
EC3	JF19M	20MAMJ	Western and Ceará

**The Silva Model** The Silva (1985) model considers the corresponding proportions to the ratio between rainfalls that occurred during the first period (X) of the rainy season – RS. Given that the rainfall of the remaining RS period is termed Y, a Z proportion, as described by  $X/(X+Y)$ , would be provided for each year for the study period. As a result, the Z values vary from 0 (zero) to 1 (one). On different occasions (Silva 1985), it has been considered for the Paraíba hinterland that X would stand for the months of January, February and the first 19 days of March (JF, JF19M, JFM) and the RS would encompass the period from January to June (JFMAMJ). By means of this model it is possible to predict with different probabilities the maximum (and minimum) rainfall for the second period (Y) of the rainy season. An 80% probability level has been normally used (Azevedo *et al.* 1998, Santos *et al.* 2002), but other levels may be considered. Once the RS and its initial period are identified, the Z proportions will then be established on a year-to-year basis against a set of data that have been collected for not less than 30 years. The Beta probabilistic model is then applied to the set of  $z_i$  values because such a model is applicable to proportions, on which occasion its adjustment is checked against the already mentioned  $z_i$  set, in accordance with the Kolmogorov–Smirnov test. Once it is confirmed that the Beta model has successfully represented the  $z_i$  set of values, one proceeds to estimate the second and eighth deciles, which will eventually be used to establish the prognostic.

Should  $Q_2$  be the second decile of  $z_i$  proportions, one can surely count on an 80% value probability of equal or greater than  $Q_2$ . Thus, for a given year, selected at random, one would see that:

$$z_i = \frac{X_i}{(X_i + Y \max)} \geq Q_2 \quad (1)$$

according to which the possibility of  $z_i$  results larger than  $Q_2$  is likely to occur. However, on considering the eighth decile ( $Q_8$ ) of the proportions concerned, one can easily state that there is an 80% probability of occurrence of values less than or equal to  $Q_8$ , which would be the same for a given year, selected at random; it therefore follows that:

$$z_i = \frac{X_i}{(X_i + Y \min)} \leq Q_8 \quad (2)$$

where  $X_i$  is the rainfall for the first RS period of that year. Developing equations (1) and (2) further, one obtains the accepted maximum and minimum rainfall heights with 80% probability level.

It may be noticed that  $z_i$  is related to rainfall belonging to two different time-periods. Once some statistical and rainfall features of their  $X_i$  related periods are known, the Y rainy prognostic is made, and that constitutes the core of the Silva (1985) model.

**Beta probability density function** A continuous, independent, and random variable  $Z$ , with values set between zero and one, will distribute itself, according to the Beta probabilistic model, if its probabilistic density function is given by (Wilks 1995):

$$f(Z_i) = \frac{Z_i^{(a-1)}(1-Z_i)^{(b-1)}}{B(a,b)} \quad (3)$$

where  $a$  and  $b$  are the parameters of the model,  $B(a,b)$  is the Beta mathematical function and  $\Gamma$  represents the Gamma mathematical function.

**The Beta model parameters evaluation and the Goodness-fit test** The method of maximum likelihood was used for assessing the parameters  $a$  and  $b$  of the Beta probabilistic model, and that was accomplished according to solutions recommended by Mielke (1976). The efficiency of the Beta probability distribution model was considered by the Kolmogorov-Smirnov (K-S) test, a non-parametric test, as it can be applied to small samples unrestrictedly.

**The assessment of the probability of occurring rainfall above the average value of  $Y$**  In order to calculate the probability of rainfall above the RS second average value in a given locality over a given year, it is enough to determine the probability of  $Z$  being equal or less than  $Z_{ave}$ , where  $Z_{ave} = X_i / (X_i + Y_{ave})$ , and where  $X_i$  would correspond to the rainfall recorded during the RS first period for that particular year, and that  $Y_{ave}$  would correspond to the average value for the rainfall of the RS second period. This can be obtained using the Beta model.

## RESULTS AND DISCUSSION

**Adjusting the Beta Model** For the Western homogeneous area (Sertões) 37 stations and three different periods were assigned in order to establish the expected rainfall prognostic for the RS second period, which corresponds to the beginning of March, the middle of March and the beginning of April. This matches the Beta distribution application to 111 samples. To all of them the Kolmogorov-Smirnov test was applied at a significance level of  $\alpha = 0.10$ . The Beta model adjusted satisfactorily to all samples under investigation, which shows the high significance of the Beta model. When applying the Beta model to the seven mesoregions of the Ceará State, good fits were verified. Santos *et al.* (2002) applied the Beta model to  $z_i$  proportions from 34 stations located on the northeast Brazil eastern coast for different rainy seasons. The Beta best fittings were obtained from Natal-RN data, Itabaiana-PB, Mamaguape-PB, and Palmares-PE. The poorest fittings occurred at Propriá-SE and Aracaju-SE, notwithstanding the fact that the null hypothesis ( $H_0$ ) for  $\alpha = 0.10$  has been accepted.

**The Silva Model Performance** The performance of the Silva model was estimated in order to see the successes and mistakes related to the expected  $Y_{max}$  and  $Y_{min}$  predictions for the RS  $Y$  period in 2000, serving only to illustrate the method's many potentialities. Table 2 exhibits the results of the application of the forecasting model for the year 2000 and for the rainy season RS1. The same table shows the second decile ( $Q_1$ ) and eighth decile ( $Q_4$ ) of  $z_i$  proportions, the average rainfall for the RS first ( $X_m$ ) and second ( $Y_m$ ) periods, the rainfall values for the first ( $X_{2000}$ ) and second ( $Y_{2000}$ ) period of the 2000 RS, the maximum ( $P_{max}$ ) and the minimum ( $P_{min}$ ) rainfall values predicted by the Silva model for the second part of the 2000 rainy season, the  $z_i$  proportion values for the year 2000. Also in Table 2, one can see that the forecast of  $Y_{min}$  failed in 10 years (bold numbers), which represents 23%, when the expected one would be 20%. In the case of  $Y_{max}$ , the forecast presented 100% success.

All the results predicted for the whole period 1996–2001 in Sertões of the State of Paraíba for the RS1, are represented in Table 3. The first period of that RS is formed by January and February (JF), and the second one by March to June (MAMJ). On examining the results, one verifies a 89.19% success in predicting maximum values, and 97.29% success in predicting minimum values for 1996 year. The results obtained by the  $Y_{max}$  and  $Y_{min}$  predictions were rather satisfactory, resting within range of the probable (86.08% for  $Y_{max}$  and 78.97% for the  $Y_{min}$ ). In the State of

**Table 2** Raingauge stations of Sertões with their correspondents data for the RS EC1: second ( $Q_2$ ) and eighth ( $Q_8$ ) deciles of  $z_i$ , mean values of the first ( $X_m$ ) and second ( $Y_m$ ) period of EC1; forecasted maximum ( $P_{max}$ ) and minimum ( $P_{min}$ ) rainfall values for 2000 year; and the proportion variable for 2000 year –  $z_i$ .

Stations	$Q_1$	$Q_4$	$X_m$	$Y_m$	$X_{2000}$	$Y_{2000}$	$P_{max}$	$P_{min}$	Z
1, Açude Coremas	0.192	0.444	246.4	525.8	371.2	420.2	1562.1	<b>464.8</b>	0.41
2, Açude E. Ávidos	0.253	0.475	304.4	528.7	431.9	475.8	1275.2	<b>477.4</b>	0.45
3, Açude Pilões	0.166	0.420	233.0	515.9	447.2	683.2	2246.8	617.6	0.46
4, Água Branca	0.131	0.405	169.1	444.6	303.4	363.8	2012.6	<b>445.7</b>	0.41
5, Aguiar	0.182	0.449	236.3	530.5	298.2	577.2	1340.3	365.9	0.36
6, Antenor Navarro	0.191	0.432	276.7	595.8	467.1	717.0	1978.4	614.2	0.44
7, Barra do Juá	0.150	0.368	175.0	489.0	464.8	538.0	2633.9	<b>798.2</b>	0.49
8, Belém B. Cruz	0.114	0.429	178.8	452.0	253.3	284.2	1968.6	<b>337.1</b>	0.36
9, Bom Jesus	0.185	0.416	237.5	549.0	474.2	716.4	2089.0	665.7	0.46
10, Cajazeiras	0.212	0.465	269.8	537.1	419.7	694.6	1560.0	482.9	0.44
11, Catingueira	0.139	0.390	241.5	667.9	202.6	418.0	1255.0	316.9	0.23
12, Catolé do Rocha	0.130	0.371	195.5	595.6	619.7	431.0	4147.2	<b>1050.7</b>	0.51
13, Conceição	0.210	0.488	274.9	490.7	301.7	365.6	1135.0	316.5	0.38
14, Condado	0.142	0.398	206.1	514.0	252.8	409.5	1527.5	382.4	0.33
15, Ibiara	0.277	0.464	312.6	553.0	370.7	486.6	967.6	428.2	0.40
16, Imaculada	0.151	0.414	157.0	409.8	175.5	380.9	986.8	248.4	0.30
17, Itaporanga	0.181	0.446	239.0	517.2	385.2	602.4	1743.0	478.5	0.43
18, M. D'A. Dentro	0.138	0.477	208.6	532.7	226.9	211.9	1417.3	<b>248.8</b>	0.30
19, Malta	0.139	0.436	187.5	487.7	242.4	413.0	1501.5	313.6	0.33
20, Manaíra	0.211	0.469	190.1	372.4	290.3	445.4	1085.5	328.7	0.44
21, Nazarezinho	0.228	0.486	287.3	513.9	428.9	484.6	1452.2	453.6	0.45
22, Nova Olinda	0.198	0.471	270.4	540.5	309.2	453.7	1252.4	347.3	0.36
23, Olho D'Água	0.141	0.399	278.9	774.2	441.9	440.3	2692.1	<b>665.6</b>	0.36
24, Patos	0.166	0.508	218.5	457.3	160.4	371.1	805.9	155.3	0.26
25, Piancó	0.184	0.459	250.3	531.2	393.2	505.3	1743.8	463.4	0.43
26, Pombal	0.157	0.413	194.8	454.8	182.7	548.1	981.0	259.7	0.29
27, Princesa Isabel	0.178	0.482	234.4	450.5	195.9	327.9	904.7	210.5	0.30
28, S. J. Piranhas	0.254	0.484	316.0	543.1	424.8	499.0	1247.6	452.9	0.44
29, Santa Luzia	0.119	0.424	146.9	351.8	187.4	247.8	1387.4	254.6	0.35
30, São Gonçalo	0.208	0.446	273.7	526.4	517.8	458.9	1971.6	<b>643.2</b>	0.50
31, S. J. L. Tapada	0.244	0.452	320.0	585.4	318.6	464.5	987.1	386.3	0.35
32, Serra Grande	0.221	0.473	253.3	460.4	389.5	502.8	1372.9	434	0.46
33, Souza	0.178	0.482	237.8	456.3	329.9	509.9	1523.5	354.5	0.42
34, St Arapuá	0.252	0.482	267.9	464.2	356.6	464.2	1058.5	383.2	0.43
35, Teixeira	0.153	0.473	198.4	434.8	333.1	335.3	1844.0	<b>371.1</b>	0.43

**Table 3** General results (%) of forecasting success into the period 1996 to 2001 for RS1, RS2 and RS3.

Year	RS1		RS2		RS3	
	$P_{max}$	$P_{min}$	$P_{max}$	$P_{min}$	$P_{max}$	$P_{min}$
1996	89.19	97.29	83.78	100.00	97.29	100.00
1997	94.59	83.78	100.00	97.29	72.97	100.00
1998	100.00	30.56	100.00	58.33	97.22	66.67
1999	81.08	91.89	97.30	86.49	100.00	70.27
2000	100.00	70.27	100.00	97.29	100.00	97.14
2001	51.61	100.00	96.77	54.83	100.00	48.65
<b>Mean</b>	<b>86.08</b>	<b>78.97</b>	<b>96.31</b>	<b>82.37</b>	<b>94.58</b>	<b>80.46</b>

Ceará for the central area (mesoregion E6), according to Table 4, and in the period 1960–1969 both  $Y_{max}$  and  $Y_{min}$  presented just one failure, which shows the importance of the model. Santos *et al.* (2002) have obtained similar results after investigating raingauge stations on the eastern coast of northeast Brazil, and so have Azevedo *et al.* (1998) when they studied the homogeneous microregions of the State of Ceará. For all RS the results are extremely precise for rainfall forecasting, mainly for the  $Y_{max}$ . Considering all investigated years, 1998 and 2001 were the ones that presented the largest differences between predictions involving all rainy seasons. One can see that for all rainy seasons, the  $X_{1998}$  values exhibited low  $z_i$  values, and consequently reduced

$P(Y > Y_{\text{clim}})$ . This means that 1998 may be considered a non-typical year, given the rise in discrepancies between the probabilities  $P(Y > Y_{\text{clim}})$  and the rainfall above the  $Y_{\text{clim}}$  climatic average at every station and RS. It may be important considering that 1998 was the year of the last century with the strongest El Niño. As to the predicted maximum and minimum rainfall, the success indexes were very good, which therefore recommends the use of the Silva model in the LMRS and FUNCEME operational routine. Moreover, there is strong evidence towards the application of a filter in case of extreme events (rainy days), which would enhance the Silva model performance.

**Table 4** Results for  $X_i$ ,  $Y_i$ ,  $Y_{\text{max}}$  and  $Y_{\text{min}}$  associated with the raingauges of EC6 in the State of Ceará.

Year	$X_i$	$Y_i$	$Y_{\text{max}}$	$Y_{\text{min}}$
1960	251.6	393.3	510.7	223.9
1961	429.8	443.7	872.5	382.5
1962	283.7	367	576	252.5
1963	531.2	332.1	1178.3	<b>472.8</b>
1964	576.10	635.90	1169.50	512.70
1965	142.90	662.00	<b>290.10</b>	127.10
1966	220.50	287.70	447.60	196.20
1967	347.60	583.00	705.60	309.30
1968	317.40	420.10	644.30	282.50
1969	258.30	444.40	524.30	229.90

## CONCLUSIONS

Broadly speaking, the Silva model (1985) presented a success index equal or around to 80%, for general estimations of the  $Y_{\text{max}}$  and  $Y_{\text{min}}$ . The best results obtained with the application of  $P(Y > Y_m)$  were verified at the RS2 and RS3 stations, and were rather factual, being, as a result, recommended for both the northeast Brazil State laboratories of climatic analyses.

**Acknowledgements** The authors thank the National Council for Scientific and Technological Development (CNPq), the Financier of Studies and Projects (FINEP), the Foundation for Science and Technology of Pernambuco State (FACEPE), Coordination for the Improvement of Higher Education Personnel (CAPES) and Water and Climate Pernambuco Agency (APAC).

## REFERENCES

- Azevedo, P. V. de; Silva, B. B. da & Rodrigues, M. F. G. (1998). Previsão estatística das chuvas de outono no estado do Ceará. *Revista Brasileira de Meteorologia* 13, 19–30 (in Portuguese).
- Clarke, A. B. & Ralph, D. (1979) Probabilidade e Processos Estocásticos, Ed. Livros Técnicos e Científicos Editora S. A., Rio de Janeiro, 1979, 338 p (in portuguese).
- Mielke, P. W. Jr. (1976) Convenient beta likelihood techniques for describing and comparing meteorological data. *Journal of Applied Meteorology* 14, 985–990.
- Santos, F. A. S., Azevedo, P. V. & Silva, B. B. da. (2002). Previsão de chuvas de outono na costa leste do Nordeste do Brasil. *Revista Brasileira de Meteorologia*, São Paulo, 17(2), 195–206 (in Portuguese).
- Silva, B. B. (1985) Estudo da precipitação no estado da Paraíba: regimes pluviais caracterização de anos secos e chuvosos. M.Sc. Dissertation. UFPB. CCT, Campina Grande, 100p.
- Varejão-Silva, M. A. V., et al. (1987) Atlas Climatológico do Estado da Paraíba. Universidade Federal da Paraíba. Núcleo de Meteorologia Aplicada (in Portuguese).
- Walker, G. T. (1928) Ceará, Brazil, famines and general air movement. *Beitrag zur Physik der freien. Atmosphere* 4, 88–93.
- Ward, M. N. & Folland C. K. (1991) Prediction of seasonal rainfall in the North Nordeste of Brazil using eigenvectors of sea-surface temperature. *International Journal of Climatology* 11, 711–743.
- Wilks, D. S. (1995) *Statistical Methods in the Atmospheric Sciences – An Introduction*. International Geophysics Series. Academic Press. 467 p.