EMPIRICAL PROBABILITY MODELS TO PREDICT PUERTO RICO MONTHLY RAINFALL PROCESS

JP3.22

Nazario D. Ramirez-Beltran*
University of Puerto Rico, Mayaguez, PR

K.-M. Lau
NASA Goddard Space Flight Center, Greenbelt, MD

Amos Winter
University of Puerto Rico, Mayaguez, PR

and

Joan M. Castro
University of Puerto Rico, Mayaguez, PR

Nazario Ramirez Escalante
University of Puerto Rico, Mayaguez, PR

1. INTRODUCTION.[1]

A methodology to predict monthly rainfall is proposed. The prediction strategy depends on dynamic probability models and empirical functions. The parameters of the dynamic probability model are changing with time while the structure of the probability model will remain unchanged. The parameters of the dynamic probability model are estimated at every point in time by using empirical functions. The empirical function is a time difference equation that establishes the relationship between a random vector that belongs to the probability model and a set of time series, which are observations of climatological phenomena. In time series literature the empirical functions are known as multivariate transfer function models (Box and Jenkins, 1976; Brockwell, and Davis, 1996). A mathematical relationship between the dynamic probability model and the empirical functions was derived after taking the first moment of both the probability and the empirical models. Thus, the parameters of the dynamic probability model become a set of empirical functions.

The maximum likelihood method was used to estimate the parameters of the dynamic probability model because this method is invariant under linear transformations and because most of the time their estimators are consistent (Bickel, and Doksum, 1977). Typically, the resulting likelihood function is a highly nonlinear function with some constraints. The suggested optimization method consists of two steps: The first one is dedicated to estimate the initial point, which is obtained after estimating the parameters of the empirical functions. The second step consists of obtaining the final estimates of the parameters of the dynamic model. The sequential quadratic programming (SQP) algorithm was selected to solve the constrained nonlinear optimization problem. Thus, if an initial point is carefully selected, then the nonlinear algorithm will converge to a satisfactory local maximum, and consequently, the optimal parameters of the dynamic probability model will be available. The dynamic probability model and the empirical functions will be used to compute the probability that in a particular station and time the rainfall level will exceed the normal behavior, or the rainfall level would be less than the normal level, or the rainfall level would be in the normal range. Once, the probability is known for one of the previous three stages, the conditional expected rainfall will be predicted.

The proposed algorithm was successfully applied to predict the monthly rainfall process of six rainfall stations located in Puerto Rico (PR) with the longest rainfall records. PR is a small Caribbean island. Nevertheless, as part of the NWS COOP program 91 rainfall gauges have been installed starting in 1899 and 65 of them are currently collecting rainfall data on a daily basis.

2. DATA

In this study, six rainfall stations were selected since they have the most complete and the longest records in Puerto Rico. The stations are: Coloso, Isabela, Manati, Maunabo, Mayaguez, and San Juan. These stations have one hundred and one years (1901-2001) of daily rainfall records. The rainfall records were compared to well know SST Oscillation Indexes. The studied time series are the following: the SST in the North Atlantic (5-20°N, 60-30°W), SST in the South Atlantic (0-20°S, 30°W-10°E), SST in Tropical Equatorial (10°S-10°N, 0-360°). The data set also includes the SST in the equatorial Pacific: el Niño

**Corresponding author address:* Nazario D. Ramirez, University of Puerto Rico, Dept. of Industrial Engineering, Mayaguez, PR 00680, e-mail: nazario@ece.uprm.edu.

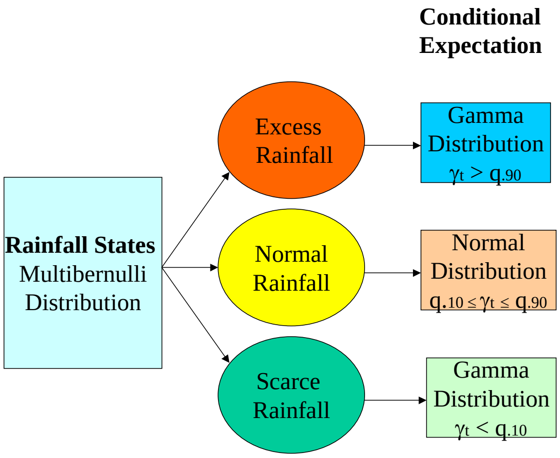

Figure 1. General Model Scheme

1-2 (0-10°S, 90-80°W), el Niño 3 (5°N-5°S, 150-90°W), el Niño 4, (5°N-5°S, 160°E-150°W), and el Niño 3-4 (5°N-5°S, 170-120°W), as well as the North Atlantic Oscillation index. These data sets was provided by the Climate Diagnostic Center located at Boulder, Colorado and by the Puerto Rico Climate Office .

The SST time series are complete ; however, rainfall observations have missing data. Therefore, missing rainfall observations were estimated in order to perform time series analysis and to develop the appropriate empirical functions. A method to perform time and spatial interpolation was implemented to estimate the missing values of the rainfall process (Ramirez-Beltran et al., 2002). The interpolation algorithm is a convex combination of spatial and time interpolation methods. The convex combination can be written as follow:

$$y_{i,t} = \alpha_i K_{i,t} + (1-\alpha_i) A_{i,t}\,, \qquad 0 \le \alpha_i \le 1 \tag{1}$$

where $y_{i,t}$ is an estimate of the missing value in the i^{th} station at time t, $K_{i,t}$ is the spatial interpolation in the i^{th} station at time t obtained by using the Kriging algorithm and $A_{i,t}$ is the time interpolation in the i^{th} station at time t obtained by the seasonal autoregressive integrated and moving average (ARIMA) model (Storch and Zwiers,2001; Brockwell, and Davis, 1996). The α_i is the convex coefficient at the i^{th} station obtained by using a cross-validation technique. A large subset of rainfall observations with no missing values were selected to estimated the α_i coefficients. The selected data include 187 observations, which start on May 1985 and finish on December 2001. The selected series were divided into two equal parts, the first part was used to fit the ARIMA model and the second one was used to perform time and spatial interpolation. The second part of the data was used for validation. Thus, 30% of the second part were randomly selected and eliminated for each series. The eliminated values were declared as missing values. The Kriging algorithm and the ARIMA model were used to estimate the missing values for each station. Since the actual values are known the α_i coefficients were estimated using linear regression techniques.

3. METHODOLOGY

The proposed methodology consists of probabilistic models in which the parameters of the models are changing because of the effects of meteorological changes. Thus, the parameters of the probability model can be written in terms of meteorological indexes and consequently developing empirical functions can capture the dynamics of the probabilistic model. Careful attention must be devoted to satisfy the constraints involved in the parameters of the probabilistic model. Thus, if empirical functions do not satisfy the parameter constrains the potential capability of the probabilistic model is destroyed and prediction will be meaningless.

3.1 *Probability of Stage*

We defined rainfall processes for any month or year as follows: A month is in excess stage when the amount of rainfall exceeded the 90 percentile. A month is in scarce stage when rainfall is less than the 10 percentile. A month is in a normal stage when the amount of rainfall occurs in the range between 10 to 90 percentiles. Once the model establishes that a particular stage has occurred then it proceeds to identify the conditional probability density function and to estimate the expected rainfall for that specific month.

It should be noted that each one of the stages is a mutually exclusive event. Thus, the rainfall process can be represented by a stochastic sequence of three multi exclusive events. Therefore, the stage of a single month for a particular year can be modeled by the generalization of the Bernulli distribution, which will is named here the Mutibernulli distribution. A random vector, $\mathbf{Y}_t$, has the Multibernulli distribution if its probability mass function can be written as follows:

$$f(y_{1,t}, y_{2,t}, y_{3,t}) = \binom{1}{y_{1,t} \quad y_{2,t} \quad y_{3,t}} p_{1,t}^{y_{1,t}} \quad p_{2,t}^{y_{2,t}} \quad p_{3,t}^{y_{3,t}} \qquad (2)$$

$$1 = y_{1,t} + y_{2,t} + y_{3,t} \quad \text{and} \quad 1 = p_{1,t} + p_{2,t} + p_{3,t} \qquad (3)$$

where

$$y_{1,t} = \begin{cases} 1, \text{ if } r_t > q_{i,0.90} \\ 0, \text{ otherwise} \end{cases} \qquad y_{2,t} = \begin{cases} 1, \text{ if } r_t < q_{i,0.10} \\ 0, \text{ otherwise} \end{cases}$$

$$y_{3,t} = \begin{cases} 1, \text{ if } q_{i,0.10} \le r_t \le q_{i,0.90} \\ 0, \text{ otherwise} \end{cases} \qquad (4)$$

r_t is the amount of rainfall for a given month on a year *t*, $q_{i,0.10}$ is the rainfall 10 percentile for the i^{th} month, $q_{i,0.90}$ is the rainfall 90 percentile for the i^{th} month. $p_{i,t}$ are the parameters of the Multibernulli distribution and each one represents the probability of success for the i^{th} stage at the year *t.* For example, $p_{i,t}$ represents the probability that a particular month is at the i^{th} stage on year *t.*

It should be noted that the constraints of the model (2) can be incorporated into the probability model and consequently the Multibernulli distribution can be written as follows:

$$f(y_{1,t}, y_{2,t}) = p_{1,t}^{y_{1,t}} \quad p_{2,t}^{y_{2,t}} (1 - p_{1,t} - p_{2,t})^{1-y_{1,t}-y_{2,t}} \qquad (5)$$

The parameters of the dynamic Multibernulli distribution will change through time and consequently an empirical model is required to express the parameters of the distribution by climatological variables, which also change throughout time. Therefore, the dynamic of the probabilistic model are captured by observing climatological variables. Thus, empirical models are needed to express the relationship between the parameters of the theoretical model and the observed climatological variables. The following empirical models were postulated to estimate the parameters of the dynamic probabilistic model.

$$y_{1,t} = a_0 + \sum_{(i,j,k)\in A} a_{ij} x_{ijk} + \varepsilon_{1ijk}, \qquad k = t - L \qquad (6)$$

$$y_{2,t} = b_0 + \sum_{(i,j,k)\in B} b_{ijk} z_{ijk} + \varepsilon_{2ijk} \qquad k = t - L \qquad (7)$$

where x_{ijk} is a meteorological index with a subscript that belong to set "*A*". In this study a meteorological index is one of the variables defined in the data section. "*A*" is a set of subscripts defined as follows: *i* represents the variable name, *j* represents the month, *t* indicates the year and *L* is the time lag expressed in years. The elements of "*A*" correspond to the variables that significantly contribute to explain the variability of $y_{1,t}$ which is a sequence of Bernulli process defined by equation (4). Similarly, the z_{ijk} is a meteorological index whose subscripts belong to set "*B*". The coefficients a_{ijk}. and b_{ijk} are the regression coefficients and ε_{1ijk} and ε_{2ijk} are sequences of independent random variables. Model (6) represents the occurrence of excess of rainfall while model (7) expresses scarce rainfall for a particular month.

It is necessary to express a relationship between the empirical models and the probabilistic model. The approach to derive this relationship is by determining the first moment of both empirical and probabilistic models. The first moment of the probabilistic model is obtained as follows:

$$E(\mathbf{Y_t}) = \sum \mathbf{Y_t} f(\mathbf{Y_t}) = [1 \quad 0 \quad 0] f(1,0,0) + [0 \quad 1 \quad 0] f(0,1,0) + [0 \quad 0 \quad 1] f(0,0,1)$$
$$= [1 \quad 0 \quad 0] p_{1,t} + [0 \quad 1 \quad 0] p_{2,t} + [0 \quad 0 \quad 1] p_{3,t}$$
$$= [p_{1,t} \quad p_{2,t} \quad p_{3,t}]$$

where "*E*" is the mathematical expectation operator. Thus, it follows:

$$E(\mathbf{Y_t}) = E[y_{1,t} \quad y_{2,t} \quad y_{3,t}] = [E(y_{1,t}) \quad E(y_{2,t}) \quad E(y_{3,t})]$$
$$= [p_{1,t} \quad p_{2,t} \quad p_{3,t}]$$

$$E(y_{1,t}) = p_{1,t} \quad E(y_{2,t}) = p_{2,t} \quad E(y_{3,t}) = p_{3,t} \qquad (8)$$

The expected values of the empirical models given that meteorological indexes are known (*x's* and *z's*) can be written as follows:

$$E(y_{1,t}|\mathbf{x}) = a_0 + \sum_{(i,j,k)\in A} a_{ijk} x_{ijk}, \qquad k = t - L \qquad (9)$$

$$E(y_{2,t}|\mathbf{z}) = b_0 + \sum_{(i,j,k)\in B} b_{ijk} z_{ijk} \qquad k = t - L \qquad (10)$$

Therefore, assuming that historical information of meteorological indexes are known up to time *jk* (*j*=month and *k*= year) the relationship between the empirical models and the theoretical model can be expresses as follows:

$$p_{1,t} = a_0 + \sum_{(i,j,k)\in A} a_{ijk} x_{ijk}, \qquad k = t - L \tag{11}$$

$$p_{2,t} = b_0 + \sum_{(i,j,k)\in B} b_{ijk} z_{ijk}, \qquad k = t - L \tag{12}$$

The parameters of the dynamic probabilistic model will be estimated by using the maximum likelihood estimation procedure. The maximum likelihood function for the Mutibernulli distribution is as follows:

$$L(\mathbf{p}) = \prod_{i=1}^{n} p_{1,i}^{y_{1,i}} \quad p_{2,i}^{y_{2,i}} \quad \left(1 - p_{1,i} - p_{2,i}\right)^{1-y_{1,i}-y_{2,i}} \tag{13}$$

where n is the total number of available data. Since the probability model is a positive function, maximizing (13) is equivalent to maximizing its natural log. It follows:

$$J(\mathbf{p}) = lnL(\mathbf{p}) = \sum_{i=1}^{n}\left\{y_{1,i}\ln(p_{1,i}) + y_{2,i}\ln(p_{2,i})\right\} + \sum_{i=1}^{n}\left\{(1 - y_{1,i} - y_{2,i})\ln(1 - p_{1,i} - p_{2,i})\right\} \tag{14}$$

where $0 \le p_{1,i} \le 1$ and $0 \le p_{2,i} \le 1$

Replacing equations (11) and (12) into equation (14) the maximum likelihood estimators for *a's* and *b's* can be obtained after maximizing the following expression:

$$J(\mathbf{a},\mathbf{b}) = \sum_{k=L+1}^{n}\left\{y_{1,k}\ln\left(a_0 + \sum_{(i,j,k)\in A} a_{ijk}x_{ijk}\right) + y_{2,k}\ln\left(b_0 + \sum_{(i,j,k)\in B} b_{ijk}z_{ijk}\right)\right\} + \sum_{k=L+1}^{n}\left\{(1 - y_{1,k} - y_{2,k})\ln\left(1 - a_0 - \sum_{(i,j,k)\in A} a_{ijk}x_{ijk} - b_0 - \sum_{(i,j,k)\in B} b_{ijk}z_{ijk}\right)\right\} \tag{15}$$

subject to

$$0 \le a_0 + \sum_{(i,j,k)\in A} a_{ijk}x_{ijk} \le 1 \text{ and } 0 \le b_0 + \sum_{(i,j,k)\in B} b_{ijk}z_{ijk} \le 1 \tag{16}$$

The constrained nonlinear optimization problem is solved by using the following strategy. First, regression values obtained after fitting equations (6) and (7) were used as the initial point for the nonlinear optimization routine. The Sequential Quadratic Programming (SQP) algorithm was selected to solve this problem. SQP algorithm approximates the objective function by a quadratic function and constraints by linear functions. Once, the approximation is complete the quadratic problem is solved by using the Wolf simplex method (Bazaraa et al, 1993). The next point is considered as a new initial point and a new approximation is designed and solved. This process is repeated until eventually a local maximum is found.

3.2 *Conditional Probability*.

Once the stage of a month is predicted the next task is to determine the expected amount of rainfall that will occur in the underlying month. It is required to derive the conditional probability of the rainfall given that the month is in a specific stage. Thus, three different populations are studied for every month and their corresponding distributions are developed. In order to illustrate the methodology the month of May was arbitrarily selected. Because of the limited space the suggested methodology will illustrate only the stage of excess rainfall. The identified model for the scarce rainfall was Gamma distribution whereas the normal stage follows a normal distribution.

The histogram during the last 101 years shows evidence that excess of rainfall for the month of May in the Mayaguez station follows a gamma distribution (histogram was omitted). Therefore, the conditional probability distribution for excess of rainfall can be written as follows:

$$f(z_{smt} | r_{smt} > q_{0.90}) = \frac{\lambda^{\alpha_t} z_t^{\alpha_t - 1} e^{-\lambda_t z_{smt}}}{\Gamma(\alpha)}, \tag{17}$$

$$z_{smt} > 0,\ \lambda_t > 0,\ \alpha_t > 0$$

where

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad \text{for} \quad \alpha > 0 \tag{18}$$

and z_{smt} is a random variable that represents the amount of rainfall that will occur in the s^{th} station given that the amount of rainfall will exceed the normal level in the m^{th} month of the t^{th} year; r_{smt} is the unconditional rainfall value for the s^{th} station on the month m^{th} and at the k^{th} year. α_t and λ_t are the parameters of the dynamic probability model. These parameters are changing according to the climatic conditions and the changes may not be linear. The dynamics of the model will be captured by means of time series of meteorological indexes. The postulated empirical model for the excess of rainfall includes several processes generated by meteorological indexes.

$$z_{smt} = c_0 + \sum_{(i,j,k)\in C} c_{ijk} w_{ijk} + \varepsilon_{ijk}, \qquad k = t - L \qquad (19)$$

where C is a set of subscripts associate with i^{th} variable, on the j^{th} month and the k^{th} year; w_{ijk} is the i^{th} meteorological index that is highly correlated with excess of rainfall for j^{th} month and for the year k^{th}.

The relationship between the probabilistic model (17) and the empirical model (19) can be obtained by computing the expected value of both models. Thus, the derived relationship can be written as follows:

$$\lambda_t = \frac{\alpha_t}{c_0 + \sum_{(i,j,k)\in C} c_{ijk} w_{ijk}} \qquad (20)$$

Parameter estimation can be derived by the maximum likelihood function, which can be written as follows:

$$L(\alpha,\lambda) = \prod_{i=1}^{n} \frac{\lambda_i^{\alpha_i} z_{smi}^{\alpha_i - 1} e^{-\lambda_i z_{smi}}}{\Gamma(\alpha_i)} \qquad (21)$$

It should be noted that the strategy introduced in this study consists of determining the parameters of the probabilistic model using empirical equations. Computing the natural log of the likelihood function and substituting equation (20) into equation (21) the problem reduces to find the values of α_t and *c's* so that the following function is maximized.

$$J(\alpha,c) = \sum_{k=1}^{n}\left\{\alpha_k \ln\left(\frac{\alpha_k}{c_0 + \sum_{(i,j,k)\in C} c_{ijk} w_{ijk}}\right) + (\alpha_k - 1)\ln(z_{smk})\right\} - \sum_{k=1}^{n}\left\{\ln(\Gamma(\alpha_k)) + \frac{\alpha_k z_{smk}}{c_0 + \sum_{(i,j,k)\in C} c_{ijk} w_{ijk}}\right\} \qquad (22)$$

subject to

$$\frac{\alpha_k}{c_0 + \sum_{(i,j,k)\in C} c_{ijk} w_{ijk}} > 0, \qquad \alpha_k > 0 \qquad k = t - L \qquad (23)$$

where z_{smk} is excess of rainfall and was defined on equation (17) and w_{ijk} was introduced in equation (19), and t represents the time unit in years and L is the number lagged years.

Equation (22) was maximized subject to constraint (23) by using the SQP algorithm. This nonlinear optimization algorithm will converge to an incorrect local optimum if the initial point is not selected properly. To accomplish convergence the initial point was carefully selected. Thus, the initial values of *c's* were obtained after solving the regression model represented by equation (19). Since the variance of the gamma distribution is given by the following expression: $Var(z) = \frac{\alpha}{\lambda^2}$ the initial value for the lambdas were selected as follows:

$$\alpha_t = \frac{\left(c_0 + \sum_{(i,j,k)\in C} c_{ijk} w_{ijk}\right)^2}{S_z^2} \qquad (24)$$

where S_z^2 is sample variance of the random variable z_{smk}. It should be noted that alpha's are initialized with a single value since the sample variance is a single value.

4. RESULTS

A model for each rainfall station and for each month during the last 101 years (1901 to 2001) was designed. Because of the paper space results are limited to Mayaguez station. The model identification scheme is given for a single month and the cross-validation is provided for 40 years (1961 – 2000).

Rainfall data for the Mayaguez station and for the month of May during 1901 to 2001 are exhibited in Figure 2. The pink line exhibits the 90 percentile and the yellow line shows the 10 percentile. Thus, if the amount of rainfall occurs in the range between 3.96 to 12.47 inches the considered month will be in the normal stage. On the other hand if the rainfall amount is greater than 12.47 inches the month will be in the excess stage, and if the rainfall amount is less that the 3.96 inches the month will be on scares stage. The 10 points that exhibit either excess or scarce rainfall reveals a gamma distribution, and the remaining 81 points follows a normal distribution (the histograms of these data were omitted).

It should be noted that at every point in time a different model is identified. A heuristic algorithm was developed to identify the best meteorological indexes that contribute to express the behavior of the rainfall for a specific month. The designed algorithm includes three major steps.

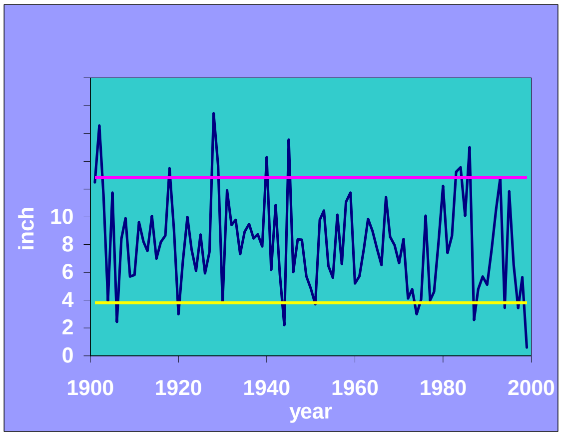

Figure 2. Rainfall stages.

(1) The algorithm generates 1,080 variables based on 9 meteorological indexes, twenty-four lagged values for each variable and five mathematical transformations are applied to each variable (24x5x9=1,080). (2) The number of variables that will be allowed in the model are fixed, depending on the available observations. For instance, 14 variables were allowed to identify the model that determines the stage of the month. Two variables were the maximum number of allowed variables for either excess or scarce model. Fourteen variables were allowed to identify a normal stage model. (3) The algorithm select the best variables based on model fitting properties, significance of variables and the explained variability by the model. Thus, each rainfall model will select the best fourteen variables out of the 1,080 to explain the stage of the predicted rainfall.

Table 1 shows the fitted model up to time 1989 to predict year 1990. The identified model has 14 variables that explain 83.22 % of the variability of $y_{1,t}$ variable, which was defined by equations (4).
The code 2 in mathematical transformations means the values are squared and 1 means that no transformation was required.
The first value of the fifth column of Table 1 is the estimated of a_0 of equation (6) and from row 2 to row 15 are the a-estimates of equation (6). Row 16 of the fifth column is the estimated of b_0 in equation (7) and the remaining values of the fifth column are the b-estimated of equation (7). The last column of Table 1 shows the result from the SQP and correspond to the optimal estimates of a's and b's. expressed in equations (15) and (16)

Table 1. Fitting Model to identify stage of the month.

Index	month	lag	transformation	Initial coefficients	Optimal coefficients
a_0				0.4115	0.2587
7	10	1	1	-0.2999	-0.0928
7	8	1	1	0.1697	0.0394
8	9	1	1	0.2741	0.1398
1	6	1	1	-0.0332	-0.0189
6	8	1	2	-0.0941	-0.0448
6	3	1	2	0.1612	0.1730
6	8	2	1	-0.0500	-0.0245
3	3	1	1	0.2569	0.0415
3	2	1	1	0.1873	0.0435
5	12	1	1	0.1557	0.0433
5	2	1	1	-0.1418	-0.0618
5	11	2	1	0.3906	0.2041
5	10	2	1	-0.3449	0.1705
5	7	1	1	-0.1772	-0.0863
b_0				-0.0110	-0.0020
4	4	1	1	0.2465	0.1365
4	2	1	1	-0.1521	-0.0718
5	8	1	1	0.0065	0.0006
5	6	1	1	-0.0123	-0.0011
5	2	1	1	0.0287	0.0035
5	9	2	2	0.0526	0.0045
5	7	2	2	-0.0115	-0.0017
6	6	1	1	-0.0218	-0.0036

The index numbers exhibited in Table 1 are defined in Table 2. It should be noted that the lagged values of Mayaguez rainfall are also used as variables to explain future values of the same variable.

Table 2. Meteorological indexes

Index number	Index Name
1	Rainfall at Mayaguez
2	North Atlantic Anomaly SST (5-20N, 60-30W)
3	South Atlantic Anomaly SST (0-20S, 30W-10E)
4	Tropical Equatorial Anomaly SST (10S-10N, 0-360)
5	El Niño 1,2 Anomaly SST (0-10S, 90-80W)
6	El Niño 3 Anomaly SST (5N-5S, 150-90W)
7	El Niño 4 Anomaly SST (5N-5S, 160E-150W)
8	El Niño 3,4 Anomaly SST (5N-5S, 170-120W)
9	North Atlantic Oscillation

Figure 3 shows the model fitting given that the month of May is in the normal stage. The blue squares show the observed values and the pink squares represent the estimated values

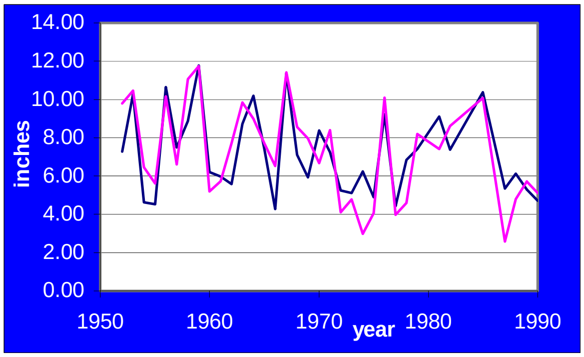

Figure 3. Normal stage model fitting (May: 1950-1990)

Figure 3 shows the model fitting given that the month of May is in the normal stage. The blue squares show the observed values and the pink squares represent the estimated values.

Figure 4 shows the results of 40 years of cross-validation, assuming that the model knows the correct stage of the month. The algorithm develops a model up to previous December and predict five month ahead what will be the expected amount of rainfall given that the stage of next May is known without error.

Calculations of the month stages require results from a nonlinear optimization routine. Once the model determines the month stage it proceeds to estimate the expected rainfall given that the stage of the month is known.

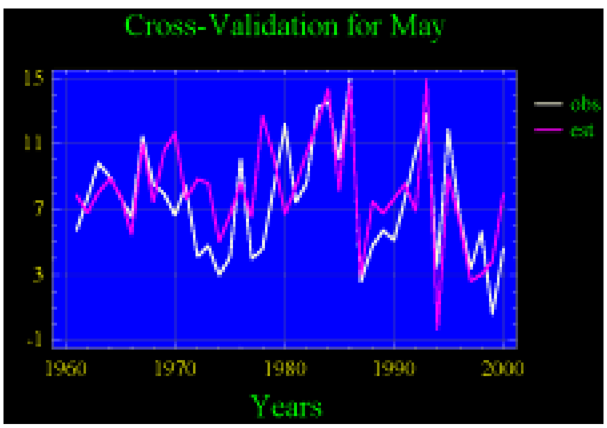

Figure 4. Cross-validation.

5. CONCLUSIONS

The proposed prediction algorithm has the capability of predicting rainfall 5 months ahead given that the stage of the month is known. The complete algorithm has been tested for the month of May and the preliminary results indicate that the algorithm is able to predict with a reasonable accuracy if the stage of the month is precisely known. If the model performs mistake on the stage of the month the prediction may not be accurate. Thus, some effort is required to improve the prediction of the stages.

The proposed algorithm is still in the process of development and exploration. It was observed that in a sample of ten years the algorithm predicted without error 80% of the times the month stage.

One of the major limitations of the proposed algorithm is the large computational effort that is required to derive predictions. 5 hours of computational time is required for a Pentium 4 to obtain twelve months predictions. The proposed algorithm processes the complete set of information up to December and it is able to predict the complete next year. Thus, to predict January the lead time will be one month, to predict February the lead time is two months and so on.

ACKNOWLEDDGEMENTS

This research is been supported by NASA-EPSCoR grant NCC5-595 and also by the University of Puerto Rico. Authors want to recognize the technical support provided by the National Diagnostic Center. Especially to Dr. Robert Webb and Dr. Jon Eischeid who provided the 101years of monthly meteorological indexes.

REFERENCES

Bazaraa, M.S. Sherali, H.D., and Shetty, C.M. 1993. *Nonlinear Programming: Theory and Algorithms.* Second Ed., John Wiley & Sons, Inc., New York.

Bickel, P.J., and Doksum,K.A. 1977. *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, Inc. Oakland, CA.

Box, G.E.P. and Jenkins, G.M. 1976. *Times Series Analysis Forecasting and Control,* Hodlen-Day Oackland, California.

Brockwell, P.J., and Davis, R.A., 1996. *Introduction to Time Series and Forecasting,,* Spring-Verlang New York Inc.

Ramirez-Beltran, N.D., 2002. Time and Spatial Interpolation algorithm. A working paper to be submitted to *Apply Meteorology*.

Storch, H.V. and Zwiers, W.Z. 2001. *Statistical Analysis in Climate Research*, Cambridge University Press., Cambridge, UK.