d States
onmental Protection
cy

Environmental Research
Laboratory
Athens GA 30613

EPA-600/9-83-015
September 1983

rch and Development

EPA

Proceedings of Stormwater and Water Quality Model Users Group Meeting

January 27-28, 1983

ESTIMATION OF MISSING VALUES IN MONTHLY RAINFALL SERIES

by

Efi Foufoula-Georgiou[1]

ABSTRACT

Infilling of missing values is often necessary prior to the practical use of hydrological time series. In this paper, three different types of infilling methods are considered reflecting the following basic ideas:

(1) the use of regional-statistical information in four simple techniques:
- mean value method (MV),
- reciprocal distance method (RD),
- normal ratio method (NR),
- modified weighted average method (MWA);

(2) the use of a univariate stochastic (ARMA) model which describes the time correlation of the series;

(3) the use of a multivariate stochastic (ARMA) model which describes the time and space correlation of the series.

An algorithm for the recursive estimation of the missing values by a parallel updating of the univariate or multivariate ARMA model is proposed and demonstrated. All methods are illustrated in a case study using 55 years of monthly rainfall data from four south Florida stations.

INTRODUCTION

Many different kinds of statistical analyses may be performed on a given data set, e.g., determination of elementary statistical parameters, auto- and cross-correlation analysis, spectral analysis, frequency analysis, fitting time series models. For routine statistics (e.g., calculation of mean, variance and skewness) missing values are seldom a problem. But for techniques as common as autocorrelation and spectral analysis missing values can cause difficulties. In multivariate analysis missing values result in "wasted information" when only the overlapping period of the time series is used in the analysis, and in numerical inconsistencies (Valencia and Schaake, 1973; Fiering, 1968; Slack, 1973) when the incomplete series are used. The evaluation of the estimation methods analyzed has utilized monthly rainfall records from the South Florida Water Management District (SFWMD), and has been based upon: a) the statistical comparison of the methods to each other at a fixed level of percent of missing values, and b) the performance of each individual method at different levels of percent of missing values. Gaps (missing values) have been artificially created in the complete record of the interpolation station (station whose missing values are to be estimated) with

[1] Graduate student, Department of Environmental Engineering Sciences, University of Florida, Gainesville.

the following procedure: First, the lengths of the gaps have been generated from a discrete exponential distribution with mean $\bar{k}$ months. Then, for a given percent of missing values, m, the mean interevent length (missing values between two successive gaps), $\bar{\tau}$, has been calculated as $\bar{\tau} = \bar{k}\ (100-m)/m$ and the interevent lengths have been generated randomly from an exponential distribution with mean $\bar{\tau}$. The values used for $\bar{k}$ and m are based on a frequency analysis of missing values in SFWMD monthly rainfall records (Foufoula-Georgiou, 1982) and are: $\bar{k}$ = 2.4 months and m = 2, 5, 10, 15 and 20%. Overlapping and concurrent periods of 55 years of monthly rainfall data of the four SFWMD stations shown in Fig. 1 have been used in the analysis.

TRADITIONAL ESTIMATION TECHNIQUES

Mathematical Representation

In all the following equations y will be used for the interpolation station and x_j for index station j, j = 1, 2, ..., n. An estimated value at time t is y'_t.

Mean Value Method (MV) --

The simplest method simply replaces the missing values with the sample mean:

$$y'_t = \bar{y} \tag{1}$$

This method results in a reduced variance and a spurious correlation coefficient especially at a high percent of missing values.

Reciprocal Distance Method (RD)

A missing value y_t is estimated as:

$$y'_t = \sum_{j=1}^{n} a_j\ x_{j,t} \tag{2}$$

The weighting coefficients a_j are calculated from:

$$a_j = (1/d_j)^p \Big/ \sum_{j=1}^{n} (1/d_j)^p \tag{3}$$

where d_j is the distance between index station j and the interpolation station, and n is the number of index stations used for the estimation. It has been concluded (Shearman and Salter, 1975; Wei and McGuinness, 1975; Dean and Snyder, 1977) that p=2 better approximates the isohyetal map drawn by conventional methods.

Normal Ratio Method (NR) --

A missing value y_t is estimated as:

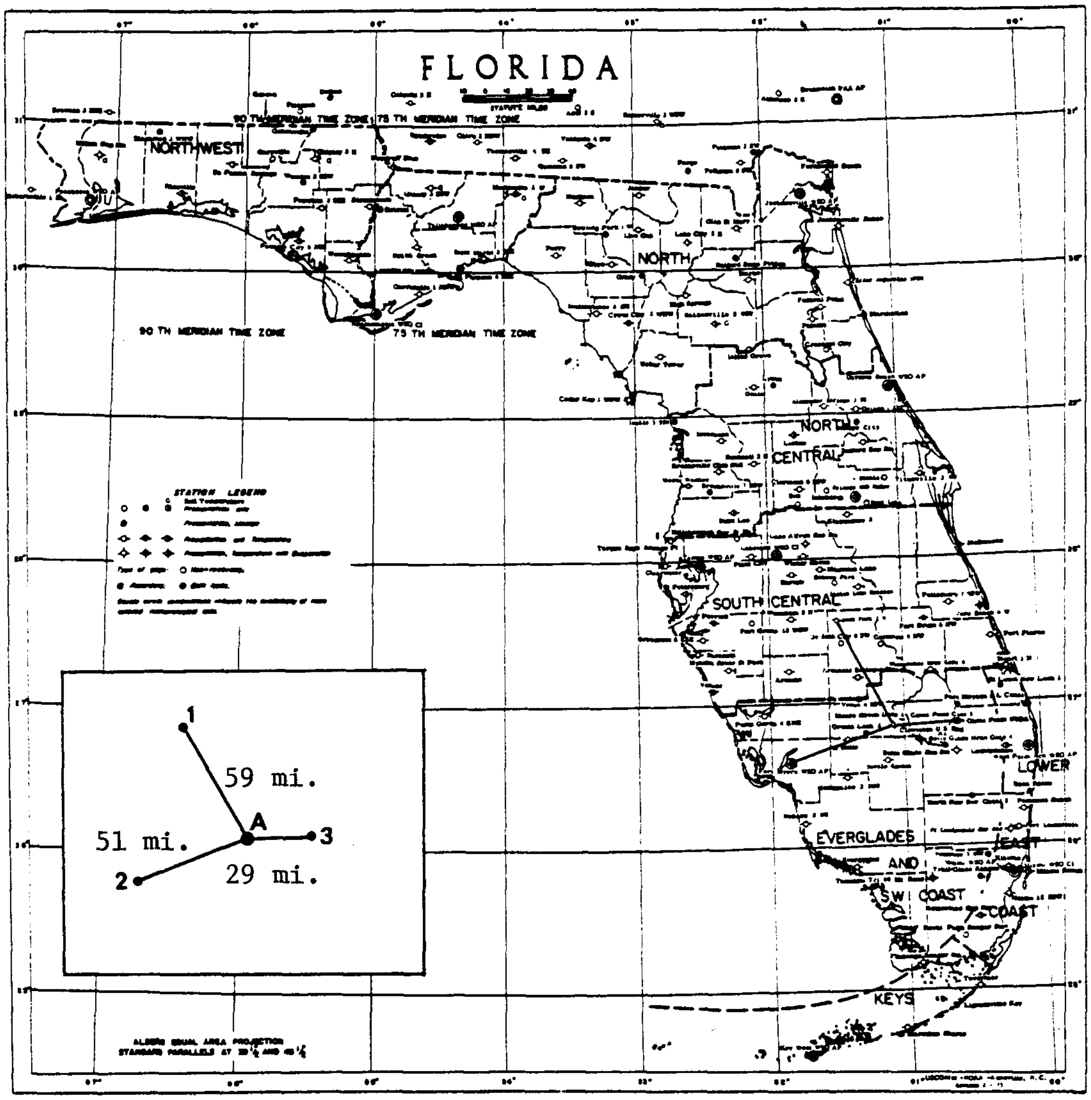

Fig. 1. The four south Florida rainfall stations used in the analysis.

A: 6038, Moore Haven Lock 1
1: 6013, Avon Park
2: 6093, Fort Myers WSO AP.
3: 6042, Canal Point USDA

$$y'_t = \frac{1}{n} \sum_{j=1}^{n} \left(\frac{R}{R_j} x_{j,t}\right) \qquad (4)$$

where R and R_j are the normal annual or monthly rainfall at the interpolation and index station j respectively. This method is recommended (Paulhus and Kohler, 1952) when the normal rainfall at any of the index stations differs from that of the interpolation station by more than 10 percent.

Modified Weighted Average Method (MWA) --

The RD and NR methods may be both written in the general form of a weighted average scheme:

$$y'_t = A\ X_t \qquad (5)$$

where A is a row vector (1 x n) and X_t is a column vector (n x 1). To preserve the mean, $\overline{y}$, and variance, s_y^2, estimated from the available data, a modified scheme may be used:

$$y'_t = B\ X_t + (\overline{y} - \overline{B}\ X) \qquad (6)$$

where

$$B = A \frac{s_y}{s_{y'}} \qquad (7)$$

and $s_{y'}$ is obtained from:

$$s_{y'}^2 = A\ \mathrm{cov}[X]\ A^T = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i\ a_j\ c_{ij} \qquad (8)$$

where c_{ij} is the covariance between elements x_i and x_j of the rainfall series of the index stations i and j (Kottegoda and Elgy, 1977; Foufoula-Georgiou, 1982).

Comparison of the Methods

Evaluation of the methods is based on the statistical comparison of the estimated series (mixture of existing and estimated values) to the incomplete series (what is really available in practice) and to the actual series (unknown in practice but known in this artificial case).

The following notation is introduced:

$\overline{y}_e$, s_e, r_e = mean, standard deviation and serial correlation coefficient of the estimated series;

$\overline{y}_i$, s_i, r_i = same as above but for the incomplete series, where i = 1, 2, 3, 4 and 5 for the five different percentages of missing values;

$\overline{y}_a$, s_a, r_a = parameters of the actual series;

$\overline{y}_r$ = mean of the residuals (estimated - actual values);

s_r^2, $s_{r,e}^2$ = variance of the residuals over the whole series and over only the estimated values respectively.

The criteria used for the comparison of the methods are:

(1) the bias in the mean as measured by (a) $\bar{y}_e - \bar{y}_i$ and (b) $\bar{y}_e - \bar{y}_a$;
(2) the bias in the standard deviation as measured by (a) s_e/s_i and (b) s_e/s_a;
(3) the bias in the serial correlation coefficient as measured by $r_e - r_a$;
(4) the bias of the estimation model as given by the mean of the residuals, $\bar{y}_r$ (this is also a way to detect a consistent over- or under-estimation by a method);
(5) the accuracy as determined by the variance of the residuals s_r^2 and $s_{r,e}^2$;
(6) the significance of the biases in the mean, standard deviation and correlation coefficient as determined by the appropriate test statistic for each.

Regarding comparison of the means the following can be concluded from Table 1:

(1) the bias in the mean in all cases is not significant at the 5% significance level as shown by the appropriate t-test;
(2) the bias in the mean of the incomplete series is relatively small but becomes larger the higher the percent of missing values;
(3) at high percents of missing values the NR method gives the less biased mean;
(4) except for the RD method which consistently overestimates the mean (the bias being larger the higher the percent of missing values), the other methods do not show a consistent over or underestimation.

Regarding comparison of the variances the following can be concluded from Table 2:

(1) although slight, the bias in the standard deviation is always significant, but this is so because the ratio of variances would have to equal 1.0 exactly to satisfy the F-test (i.e., be unbiased) with as large a number of degrees of freedom as in this study;
(2) the MV method always gives a reduced variance as compared to the variance of the incomplete series and of the actual series, the bias being larger the higher the percent of missing values;
(3) the bias in the standard deviation of the incomplete series is small;
(4) there is no consistent over or under-estimation of the variance by any of the methods (except the MV method);
(5) the MWA method does not give less biased variance even at the higher percent of missing values tested, as compared to the RD and NR methods.

Regarding comparison of the correlation coefficient the following can be concluded from Table 3:

(1) the bias in the correlation coefficient is in all cases not significant at the 5% significance level as shown by the appropriate z-test;
(2) the MV method gives the largest bias in the correlation coefficients, the bias increasing the higher the percent of missing values;
(3) all methods (except the MWA method) consistently overestimate the serial correlation coefficient of the incomplete series but not the serial correlation of the actual series, and therefore this is not considered

Table 1. Bias in the Mean.

	INC	MV	RD	NR	MWA	
			$(\bar{y}_e - \bar{y}_i)$			$\bar{y}_i$
2%	0.	0.009	0.008	0.002	0.003	4.116
5%	0.	-0.012	0.014	-0.008	0.003	4.113
10%	0.	-0.010	0.006	-0.024	-0.017	4.144
15%	0.	-0.089	0.042	0.000	-0.001	4.135
20%	0.	0.042	0.149	0.043	0.086	4.082
			$(\bar{y}_e - \bar{y}_a)$			$\bar{y}_a$
2%	-0.010	-0.001	-0.002	-0.012	-0.013	4.126
5%	-0.013	-0.025	0.001	-0.021	-0.010	
10%	0.018	0.008	0.024	-0.006	0.001	
15%	0.009	-0.020	0.051	0.009	0.008	
20%	-0.044	-0.002	0.105	-0.001	0.042	

Table 2. Bias in the Standard Deviation.

	INC	MV	RD	NR	MWA	
			s_e/s_i			s_i
2%	1.	0.995	0.998	0.996	0.998	3.680
5%	1.	0.983	1.007	1.001	1.013	3.671
10%	1.	0.972	0.996	0.986	1.005	3.705
15%	1.	0.957	0.988	0.978	0.994	3.671
20%	1.	0.944	1.006	0.973	1.011	3.701
			s_e/s_a			s_a
2%	1.002	0.997	1.000	0.998	1.000	3.673
5%	0.999	0.983	1.006	1.000	1.013	
10%	1.009	0.981	1.004	0.994	1.014	
15%	0.999	0.956	0.988	0.978	0.994	
20%	1.008	0.952	1.014	0.980	1.019	

Table 3. Bias in the serial correlation coefficient.

	INC	MV	RD	NR	MWA	
			$(r_{1,e} - r_{1,a})$			$r_{1,a}$
2%	--	0.005	0.001	0.002	-0.003	0.366
5%	--	0.006	0.003	0.001	-0.002	
10%	--	0.013	0.014	0.011	0.010	
15%	--	0.033	0.006	0.013	-0.009	
20%	--	0.042	0.004	0.011	-0.012	

Table 4. Accuracy--Mean and Variance of the Residuals.
N = total number of values = 660
N_o = number of missing values

	INC	MV	RD	NR	MWA	
			$r = (y_e - y_a)/N_o$			N_o
2%	--	-0.043	-0.061	-0.570	-0.589	13
5%	--	-0.440	0.034	-0.380	-0.176	33
10%	--	0.007	0.156	-0.113	-0.046	62
15%	--	-0.175	0.338	0.074	0.105	98
20%	--	0.037	0.502	0.038	0.200	130
			$s^2_{r,e} = (y_e - y_a)^2/(N_o-2)$			
2%	--	5.037	2.874	3.149	4.585	
5%	--	8.610	3.656	3.411	5.340	
10%	--	7.892	4.239	3.484	5.187	
15%	--	7.620	4.630	3.958	5.816	
20%	--	5.224	4.891	3.681	4.898	
			$s^2_r = (y_e - y_a)^2/(N-2)$			
2%	--	0.084	0.048	0.053	0.077	
5%	--	0.406	0.172	0.161	0.252	
10%	--	0.720	0.387	0.318	0.473	
15%	--	1.112	0.675	0.577	0.849	
20%	--	1.016	0.951	0.716	0.953	

a problem;
(4) the RD method seems to give the less biased correlation coefficient even at the higher percentage of missing values.

Regarding accuracy of the methods the following can be concluded from Table 4:
(1) no method seems to consistently over or underestimate the missing values at all percent levels, but at high percent levels the missing values are overestimated by all methods;
(2) the NR method is the most accurate method especially at high percents of missing values (i.e., it gives the smallest mean and variance of the residuals).

ESTIMATION BY A UNIVARIATE STOCHASTIC MODEL

Introduction

The observed monthly rainfall series, y_t, is normalized using the square root transformation (Roesner and Yevjevich, 1966; Stidd, 1970; Delleur and Kavvas, 1978) and the periodicity is removed by subtracting the monthly means and dividing by the standard deviations (Kavvas and Delleur, 1975). The reduced series, z_t, approximately normal and stationary is then modeled by an ARMA(1,1) model:

$$z_t = \phi\, z_{t-1} - \theta\, \alpha_{t-1} + \alpha_t \tag{9}$$

where ϕ, θ are the autoregressive and moving average parameters respectively, and α_t is a sequence of independent random variables from a normal distribution with zero mean and unit variance (white noise).

For an ARMA(1,1) model the minimum mean square error forecasts $z_t'(\ell)$ of $Z_{t+\ell}$, where ℓ is the lead time are:

$$\begin{aligned} z_t'(\ell) &= \phi\, z_t - \theta\, \ell_t \qquad &, \ell=1 \\ z_t'(\ell) &= \phi\, z_t'(\ell-1) \qquad &, \ell=2, \ldots, k \end{aligned} \tag{10}$$

as developed by Box and Jenkins (1976).

Proposed Estimation Algorithm

The estimation of the missing values in the series is performed recursively by the following procedure:
Step 1: The incomplete series $\underline{S_o}$ is filled-in with any initial estimates of the missing values giving the complete series, $\underline{S_1}$.
Step 2: An ARMA(1,1) model is fitted to the series $\underline{S_1}$ and the maximum likelihood estimates (MLE) of the parameters ϕ and θ are found.
Step 3: New estimates of the missing values are calculated as forecasts of the model $\underline{M_1} = (\phi,\theta)_1$ (made at the appropriate origin for each gap and using equation 10). The new estimates replace the old ones and the series becomes $\underline{S_2}$.
Steps 2 and 3 are repeated as many times as needed until convergence is obtained in the sense that no significant change occurs in the estimates of the

missing values as well as in the parameters of the model.

The above algorithm will be addressed as RAEMV-U (Recursive Algorithm for the Estimation of Missing Values - Univariate model) and is schematically shown in Fig. 2. A FORTRAN program has been developed for the above algorithm (Foufoula-Georgiou, 1982). Input is the incomplete rainfall series and the positions of the gaps. Output is the final estimated complete series as well as the final parameters of the fitted ARMA model.

Results of the Method

Little influence of the method used to determine initial estimates of missing values was found on the final values of parameters ϕ and θ and on the final estimates of missing values computed by the recursive scheme. All methods that were tried yielded identical estimates of missing values and model parameters after five iterations at 10 percent missing values ($\phi = 0.5095$, $\theta = 0.4333$) and eight iterations at 20 percent missing values ($\phi = 0.0776$, $\theta = -0.0293$). Moreover, by using zeroes as initial estimates the same results were obtained, suggesting the latter as a convenient choice.

The RAEMV-U method was assessed using the same statistical measures as used for the four traditional techniques described previously. Table 5 shows the bias in the mean, standard deviation and serial correlation coefficient for the final series (at 10% and 20% missing values). The bias in the mean and correlation coefficient is not significant at the 5% significance level; however, the bias in the standard deviation does not pass the stringent F-test (requiring exact equality of standard deviations) and thus is significant.

Table 5. Bias in the Mean, Standard Deviation and Serial Correlation Coefficient-Univariate Model.

	$\bar{y}_e - \bar{y}_a$	s_e/s_a	$r_{1,e} - r_{1,a}$
10%	-0.021	0.983	0.018
20%	-0.083	0.951	0.044

The forward mean square error forecasting procedure that was used worked satisfactorily in the sense that rapid convergence to a statistically acceptable series occurred. Damsleth (1980) introduced the optimal between-forecasts as that linear combination of forecasts and backforecasts which gives the minimum mean square error. For the case of monthly rainfall data the use of more sophisticated forecasts seems not to be justified. The parameters ϕ and θ of the fitted ARMA(1,1) model are very close to each other and the value of ϕ is small as compared to one, thus making the large white noise variance the predominant term in the calculation of the mean square forecast error (Box and Jenkins, 1976, p. 154).

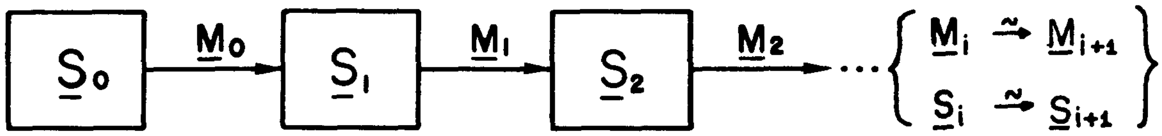

Fig. 2. Recursive Algorithm for the Estimation of Missing Values--Univariate model (RAEMV-U). $\underline{S}_i$ denotes the series, and $\underline{M}_i$ the model, $(\phi,\theta)_i$, at the ith iteration.

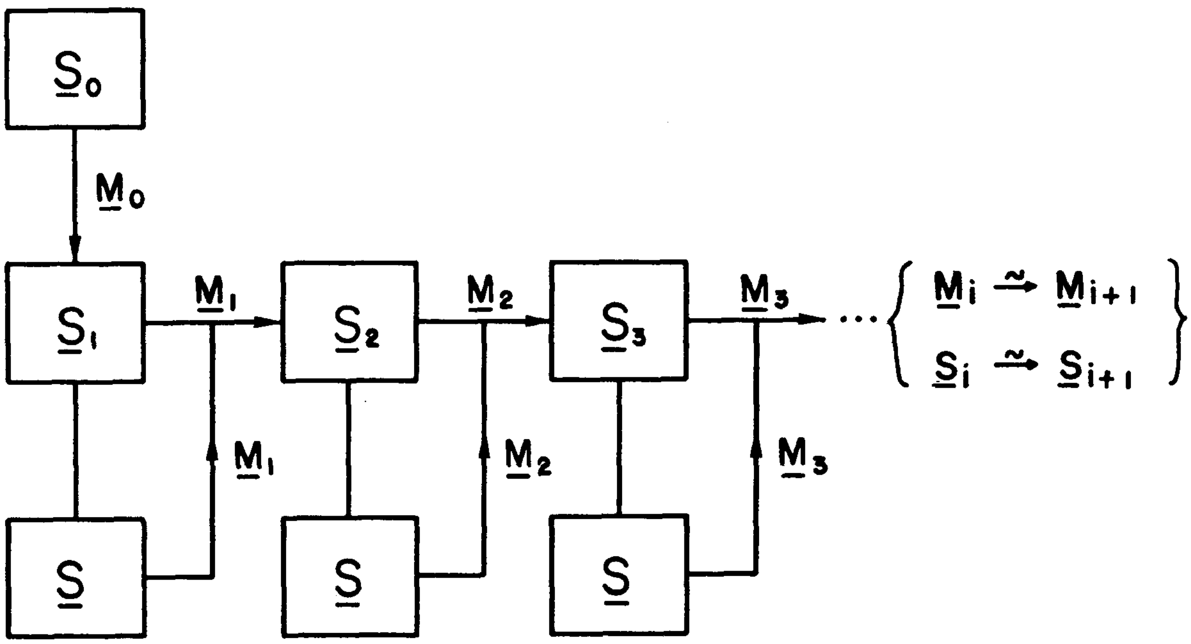

Fig. 3. Recursive Algorithm for the Estimation of Missing Values--Bivariate model (RAEMV-B). $\underline{S}_i$ denotes the series, and $\underline{M}_i$ the model, $(P,Q)_i$, at the ith iteration.

ESTIMATION BY A MULTIVARIATE STOCHASTIC MODEL

Introduction

When the concurrent rainfall series of nearby stations are available, their correlation with the series of interest may be incorporated in the model for an improved estimation of the missing values. The lag-one multivariate autoregressive model (Matalas, 1967) is expressed as:

$$Z_t = P\, Z_{t-1} + Q\, H_t \qquad (12)$$

where Z_t and Z_{t-1} are n-length vectors of the normalized and standardized variables at time t and t-1, H is an n-length vector of random components and n is the number of stations used. The above model preserves the lag-zero (M_o) and lag-one (M_1) correlation matrices when the coefficient matrices P and Q are estimated by:

$$P = M_1\, M_o^{-1} \qquad (13)$$

$$Q\, Q^T = M_o - M_1\, M_o^{-1}\, M_1^T \qquad (14)$$

Equation (14) may be solved for Q using a principal component analysis (Fiering, 1964) or much easier by an upper triangularization technique (Young, 1968; Young and Pisano, 1968). Missing values in any of the records may result in no solution at all or a solution that contains complex numbers since $Q\,Q^T$ may not be a positive semidefinite matrix as required for a real solution to occur (Valencia and Schaake, 1973; Slack, 1973).

The special case considered here is that of a bivariate AR(1) model between the interpolation station A and the index station 2 (Fig. 1). This model is written as:

$$\begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} q_{11} & 0 \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix} \qquad (15)$$

Following the Box-Jenkins forecasting procedure, the mean square error forecasts $z'_{1,t}(\ell)$ of $z_{1,t+\ell}$ are:

$$\begin{aligned} z'_{1,t}(\ell) &= p_{11} z_{1,t} + p_{12} z_{2,t} &&, \ \ell = 1 \\ z'_{1,t}(\ell) &= p_{11} z'_{1,t}(\ell-1) + p_{12} z_{2,t}(\ell-1) &&, \ \ell = 2, 3, \ldots, k \end{aligned} \qquad (16)$$

where k is the number of values missing in each gap.

Proposed Estimation Algorithm

An algorithm analogous to the one for the univariate case is also proposed for the bivariate case. The procedure is exactly the same, except that now the parameters of the model, $\underline{M}$ = (P, Q), are matrices calculated from equations (13) and (14), and the forecasts are calculated from eqn. (16)

The algorithm will be addressed as RAEMV-B (B stands for Bivariate model) and is shown schematically in Fig. 3. A FORTRAN program is also available (Foufoula-Georgiou, 1982). Input data are: the incomplete series of the interpolation station, the position of its gaps, and the complete series of the index station. Output results are: the final estimated complete series of the interpolation station, the parameters P and Q of the fitted bivariate model and the correlation matrices M_0 and M_1.

Results of the Method

Again, the scheme converges rapidly and independently of the method used to obtain initial estimates of missing values, thus suggesting their convenient replacement by zeroes to start the algorithm. Also, the convergence of the bivariate scheme seems to be less sensitive to the percentage of missing values as compared to the univariate one (three to four iterations were needed in both the 10% and 20% missing values).

Table 6 shows the bias in the mean, standard deviation and serial correlation coefficient for the final series (at 10% and 20% missing values). Again, the bias in the mean and correlation coefficient is not significant at the 5% significance level, but the bias in the standard deviation is.

Table 6. Bias in the Mean, Standard Deviation and Serial Correlation Coefficient--Bivariate Model.

	$\bar{y}_e - \bar{y}_a$	s_e/s_a	$r_{1,e} - r_{1,a}$
10%	-0.030	0.983	0.016
20%	-0.049	0.959	0.050

CONCLUSIONS

On the basis of the monthly rainfall data from the four south Florida stations used in the analysis, the following conclusions can be drawn:

(1) All the traditional estimation techniques give unbiased (overall and monthly) means and correlation coefficients at the 5% significance level even for as high as 20% missing values.
(2) At high percentages of missing values (greater than 10%) the MV method gives the more biased (although not significantly so) correlation coefficients.
(3) All methods give a slightly biased overall variance but unbiased monthly variance at the 5% significance level, and the MV method gives the most biased variances for all percentages of missing values.
(4) The NR method gives the most and the MV the least accurate estimates, at almost all levels of percent missing values.
(5) The proposed recursive algorithm works satisfactorily in both the univariate and bivariate case. It converges rapidly and independently of the initial estimates and gives unbiased means and correlation coefficients at the 5% significance level.

(6) The use of a bivariate model as compared to a univariate one did not improve the estimates except for a slight improvement at 20% missing values. However, the use of a multivariate model based on three or four nearby stations is expected to give much better estimates. The use of three adjacent stations is the main reason for the better performance of the NR method over the more sophisticated univariate and bivariate ARMA models which use only zero and one additional stations.

If the purpose of estimation is to calculate the historical statistics of the series (e.g., mean, standard deviation, and autocorrelations) the selection of the method matters little, and the simplest one may be chosen. However, if it is desired to fit an ARMA model to the incomplete series, to be used, say, to construct forecasts, the estimation of the missing values and the parameters of the model by the proposed recursive algorithm is recommended. In this case the equilibrium state (i.e., final series and parameters of the model) achieved upon convergence is unique, depending only on the existing information in the system (available data) and not on any external information added to the system (by the replacement of the missing values with some estimates derived by an arbitrary chosen method).

ACKNOWLEDGEMENTS

I would like to acknowledge with gratitude Professor Wayne C. Huber for the invaluable guidance and encouragement he provided along the course of this research.

The study was supported in part by the South Florida Water Management District.

REFERENCES

Box, G.E.P., and Jenkins, G.M., 1976, Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, Revised ed.

Damsleth, E., 1980, "Interpolating missing values in a time series," Scand. J. Stat., 7:33-39.

Dean, J.D., and Snyder, W.M., 1977, "Temporally and areally distributed rainfall," J. of the Irrigation and Drainage Div., ASCE, 103(IR2):221-229.

Delleur, J.W., and Kavvas, M.L., 1978, "Stochastic models for monthly rainfall forecasting and synthetic generation," J. Appl. Meteor., 17(10):1528-1536.

Fiering, M.B., 1964, "Multivariate technique for synthetic hydrology," J. Hydraul. Div., ASCE, 90(HY5):43-60.

Fiering, M.B., 1968, "Schemes for handling inconsistent matrices," Water Resour. Res., 4(2):291-297.

Foufoula-Georgiou, E., 1982, Estimation of Missing Values in Monthly Rainfall Series, Masters Thesis, University of Florida, Gainesville.

Kavvas, M., and Delleur, J., 1975, "Removal of Periodicities by differencing and monthly mean substraction," J. Hydrol., 26:335-353.

Kottegoda, N.T., and Elgy, J., 1977, "Infilling missing flow data," in Modeling Hydrologic Processes, Ed. by Morel-Seytoux, H., Salas, J.D., Sanders, T.G., and Smith, R.E., Water Resour. Res. Publications, Fort Collins, Colorado.

Matalas, N.C., 1967, "Mathematical assessment of synthetic hydrology," Water Resour. Res., 3(4):937-945.

Paulhus, J.L.H., and Kohler, M.A., 1952, "Interpolation of missing precipitation records," Mon. Weather Review, 80:129-133.

Roesner, L.A., and Yevjevich, V., 1966, "Mathematical models for time series of monthly precipitation and monthly runoff," Hydrology paper No. 15, Colorado State University, Fort Collins, Colorado.

Slack, J.R., 1973, "I would if I could (self-denial by conditional models)," Water Resour. Res., 9(1):247-249.

Shearman, R.J., and Salter, P.M., 1975, "An objective rainfall interpolation and mapping technique," Hydrological Sciences Bulletin, 20(3):353-363.

Stidd, C.K., 1970, "The nth root normal distribution of precipitation," Water Resour. Res., 6(4):1095-1103.

Valencia, D.R., and Schaake, J.C., Jr., 1973, "Disaggregation processes in stochastic hydrology," Water Resour. Res., 9(3):580-585.

Wei, T.C., and McGuiness, J.L., 1973, "Reciprocal distance squared method, a computer technique for estimating areal precipitation," ARS NC-8, U.S. Dept. of Agriculture, Washington, D.C.

Young, G.K., 1968, "Discussion of 'Mathematical assessment of synthetic hydrology' by N.G. Matalas," Water Resour. Res., 4(3):681-682.

Young, G.K., and Pisano, W.C., 1968, "Operational hydrology using residuals," J. Hydr. Div., ASCE, 94(HY4):909-923.

The work described in this paper was not funded by the U.S. Environmental Protection Agency. The contents do not necessarily reflect the views of the Agency and no official endorsement should be inferred.