International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

Condensed Disaggregation Procedure of Rainfall Using Conservation Corrections

R. Subbaiah
College of Agricultural Engineering and Technology
Junagadh Agricultural University, Junagadh - 362 001
E-mail: ranga3711@gmail.com

Shashank Singh
IIM, Ahmedabad, INDIA

ABSTRACT: Meteorological models generate fields of precipitation and other climatological variables as spatial averages at the scale of the grid used for numerical solution. The grid-scale can be large, particularly for general circulation models and disaggregation is required, for example, to generate appropriate spatial-temporal properties of rainfall for coupling with surface-boundary conditions or more general hydrological applications. The lane's condensed disaggregation model is adopted to disaggregate the annual rainfall series to monthly values to avoid the estimation of excessive number of parameters. When transformed series are modeled rather than real space series in disaggregation, the generated series fails to preserve the additive property. To overcome this problem, two adjustment procedures, mainly, absolute difference method and proportional adjustment method were adopted. The proportional adjustment method yielded better results than the absolute difference method (ABS).

INTRODUCTION

Relative paucity of rainfall data at lower time-scales coupled with insufficient rain gauge network demands the generation of the hydrologic sequences the design of water resource system. Two basic approaches have been taken in the generation of monthly or seasonal flow series. Periodic autoregressive moving average extensions (Obeysekera and Salas, 1962) generate monthly or seasonal flows directly, which can be summed to obtain annual flows. Such models may not capture the distribution and persistence of annual totals. An alternative to generate annual flows is disaggregation approaches to obtain finer time scale rainfall series.

Disaggregation models were introduced in hydrology by the pioneering work of Valencia and Schaake (1972, 1973). Several disaggregation models have been developed for generating multi-season stream flow sequences for single and/or several sites. Different model structures and parameter estimation procedures intended to preserve the lagged covariance properties among lower-level variables belonging to consecutive periods have been suggested by Mejia and Rousselle (1976), Hoshi and Burges (1979) and Stedinger and Vogel (1984). These multisite and multi season disaggregation models have excessive number of parameters because of the many cross correlations that they attempt to reproduce. This led to the development of staged disaggregation models (Salas *et al.,* 1980; Stedinger and Vogel, 1984). The staged disaggregation models do not explicitly model the cross correlations among annual flows at the various key stations. The condensed disaggregation models (Lane, 1979, 1982; Grygier and Stedinger, 1988; Lane and Frevert, 1990; Stedinger and Vogel, 1984) reduce the number of required parameters by explicitly modelling fewer of the correlations among the lower-level variables. The marginal distributions are easy to model if transformed data is preferred over the real space flows (Stedinger and Vogel, 1984). The disaggregation technique should emphasize both on the issue of model size and the need for and impact of adjusting seasonal flows to maintain annual consistency. In the present effort Lane's condensed disaggregation model is used to disaggregate the seasonal rainfall series from annual to monthly series and the adjustment techniques were used to maintain the additivity of the generated series.

CONDENSED DISAGGREGATION MODEL

The approach uses the extended model form, but on a one-season-at-a-time basis and with only one lagged season, the model equation may be written as,

$$Y_\tau = A_\tau X + B_\tau \underline{\varepsilon} + C_\tau Y_{\tau-1} \quad \text{... (1)}$$

The τ mean denotes the current season being generated. Thus, if there are w seasons, there are w individual equations following the form of Eq. 1. Also, there are w sets of parameters $A\tau B\tau C\tau$. For the single site case the parameter matrices are all single element matrices. This model is designed to preserve covariances between the annual value and its seasonal values, and to preserve variances and lag-one covariances among the seasonal values. The main advantage of this model is the reduction in number of parameters. The parameters of this model are estimated by,

$$\hat{A}_\tau = \left[\frac{S_{YX}(\tau,\tau) - S_{YY}(\tau,\tau-1)S_{YY}^{-1}(\tau-1,\tau-1)S_{YX}(\tau-1,\tau)}{S_{XX}(\tau,\tau) - S_{XY}(\tau,\tau-1)S_{YY}^{-1}(\tau-1,\tau-1)S_{YX}(\tau-1,\tau)}\right] \quad \text{... (2)}$$

$$\hat{C}_\tau = \left[\frac{S_{YY}(\tau,\tau-1) - \hat{A}_\tau S_{XY}(\tau,\tau-1)}{S_{YY}(\tau-1,\tau-1)}\right] \quad \text{... (3)}$$

$$\hat{B}_\tau \hat{B}_\tau^T = S_{YY}(\tau,\tau) - \hat{A}_\tau S_{XY}(\tau,\tau) - \hat{C}_\tau S_{YY}(\tau-1,\tau) \quad \text{... (4)}$$

$$S_{YY}(\tau,\tau) = \frac{1}{N-1}\sum_{V=1}^{N}\left[y_{V,\tau} y_{V,\tau}^T\right] \quad \text{... (5)}$$

$$S_{YX}(\tau,\tau) = \frac{1}{N-1}\sum_{V=1}^{N}\left[y_{V,\tau} x_{V,\tau}^T\right] \quad \text{... (6)}$$

$$S_{XX}(\tau,\tau) = \frac{1}{N-1}\sum_{V=1}^{N}\left[x_{V,\tau} x_{V,\tau}^T\right] \quad \text{... (7)}$$

$$S_{YY}(\tau,\tau-1) = \frac{1}{N-1}\sum_{V=1}^{N}\left[y_{V,\tau} y_{V,\tau-1}^T\right] \quad \text{... (8)}$$

$$S_{YX}(\tau-1,\tau) = \frac{1}{N-1}\sum_{V=1}^{N}\left[y_{V,\tau-1} x_{V,\tau}^T\right] \quad \text{... (9)}$$

Where, $S_{YY}(\tau, \tau)$ = Autocorrelation between the sub series and series associated with current season; $S_{XY}(\tau, \tau)$ = Covariance between the sub series and key series associated with current season; $S_{XX}(\tau, \tau)$ = Auto-correlation between the key series associated with current season; $S_{YY}(\tau, \tau-1)$ = Cross-correlation between the sub-series series associated with current season and the previous season; $S_{YX}(\tau-1, \tau)$ = Correlation between the sub-series and key series associated with current season.

BOX-COX TRANSFORMATION

Box and Cox (1964) power transformation was used for data transformation. The transformation can be expressed as,

$$Y_i = \frac{\left(X_i^\lambda - 1\right)}{\lambda} \quad \text{for } \lambda \neq 0, X_i > 0 \quad \text{... (10)}$$

$$Y_i = \ln X_i \quad \text{for } \lambda = 0, X_i > 0 \quad \text{... (11)}$$

In which X_i = the variates of a given data series; Y_i = the transformed variates which has a normal distribution; and λ = a constant of transformation. Eqn. 10 X can be replaced by $(x + k)$ for $x > -k$ to overcome the difficulty arising from zero values of X in the historical series. The constant λ is non-linear and cannot be determined in the closed form. The proper value of λ is that value producing a transformed sample with skewness coefficient and excess coefficient equal to zero. The value of λ generally ranges from –1.0 to 1.0. It has been observed that an increase or decrease in λ follows an increase or decrease in the coefficient of skewness. This trend is helpful in the estimation of λ. Alternatively, the likelihood function can be used to estimate λ.

Since, normal distribution has coefficient of skewness (C_s) and excess coefficient (∈) equal to zero, the efficiency of transformation can be judged by checking whether these coefficients tend to zero in the transformed series. For a completely normalized data both C_s and ∈ should be equal to zero. However, Yevjevich (1972) proposed the value of tolerance +0.50 from 0 for C_s. However, following Bowman (1973) the normality of a data set of size n may be asserted at the 95% confidence level if C_s and ∈ fall within the range of $\pm 1.96\ (24/n)^{1/2}$ respectively (Phien *et al.,* 1982). Sometimes in certain data series the Kurtosis (C_k) will not reach three. Even though Kurtosis is not zero in the transformed series, its effect in hydrological studies is generally neglected. Moments higher than the third are not commonly used in the statistical analysis of hydrologic data because the records are too short to give reliable estimates of the higher order moments.

RAINFALL ADJUSTMENT PROCEDURES

The disaggregation method described above is attractive for dividing annual flows among seasons. Unfortunately, when the historical series is transformed rather than the actual real-space data series, the generated monthly series generally fail to sum to historical annual series or to the previously generated annual series. Two different procedures mainly *the* proportional method, ABS-absolute difference method (Lane, 1979) for adjusting generated monthly series to get the seasonal sum. A brief description is presented here.

PROPORTIONAL ADJUSTMENT METHOD

Proportional adjustment is most popular and simple procedure to allocate the corrections proportionally to the originally generated annual and monthly series. Proportional adjustment procedure is appropriate for lower-level variables with gamma distributions that satisfy certain constraints. Here one selects in each year y a factor d_y, difference between the values of historical and generated series, such that,

$$Q_y = d_y \sum_t I_{yt} d_y = \sum_t \left(d_y I_{yt}\right) = \sum_t I_{yt} \quad \dots (12)$$

$$d_y = \frac{Q_y}{\sum_t I_{yt}} \quad \dots (13)$$

$$\Delta_y = Q_{y-} \sum_t I_{yt} \quad \dots (14)$$

The proportional adjusting procedure gives exact, in a strict mathematical sense, results, only if the variables Q_y are two-parameter gamma distributed, have common scale parameter and are mutually independent (Koutsoyiannis 1994). The proportional adjustment procedure is simple. On the other hand, series in the stable months are adjusted as much as series in unstable months so that the distortions of the marginal distributions are in some sense unbalanced. This prompted to use other adjustment procedures.

ABS-ABSOLUTE DIFFERENCE METHOD

The absolute difference procedure is more general and can be applied to any distribution and it can preserve the first and second moments regardless of the type of distribution,

$$Q_y = \sum I_{yt}^* = \sum \left[I_{yt} + d_y \left| I_{yt} - m_t \right| \right] \quad \dots (15)$$

$$d_y = \frac{\Delta}{\sum \left| I_{yt} - m_t \right|} \quad \dots (16)$$

$$\Delta_y = Q_{y-} \sum_t I_{yt} \quad \dots (17)$$

Among the generated monthly and weekly flows I_{yt} producing adjusted monthly and weekly rainfall I_{yt}. This procedure makes larger adjustments in more variable months, but makes no adjustment at all to a generated flow, which happens to equal its mean, even in the months with high variable series. The procedure shifts the burden of adjusting monthly series to the observation that differs substantially from their mean; thus it is likely to distort the tails of distribution more than alternative adjustments.

ERROR FUNCTIONS

Before being able to generate the synthetic sequences and/or forecast future values, models have to be found which describe the past data adequately. Ideally, these models should preserve all the properties of the observed series. In practice, however, this cannot be achieved and criteria for evaluating the statistical resemblance between historic and generated hydrologic data have to be chosen. In general, the criteria should depend upon the purpose of the model and the cost of reaching a wrong decision. The error functions employed for evaluation include (i) Akaike Information Criterion, (ii) Bayesian Information Criteria, (iii) Root Mean Square Error, and (iv) Coefficient of skewness. A brief description of these criteria is presented below.

Akaike Information Criterion (*AIC*),

$$AIC = \ln(RMSE) + \frac{2n}{N} \quad \dots (18)$$

Where, *RMSE* = Root mean square error; n = number of parameter estimated; N = sample size.

Bayesian Information Criteria (*BIC*),

$$BIC = \ln(RMSE) + \frac{n \ln(N)}{N} \quad \dots (19)$$

Root Mean Square Error (*RMSE*),

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} (M_t - P_t)^2}{N}} \quad \dots (20)$$

Where, all the terms have been defined earlier.

Coefficient of Skewness

The coefficient of skewness is determined by using the following equation,

$$\gamma(X) = \left[\sum_{i=1}^{M} X_i^3 \big/ M - 3\mu(X)\sigma^2(X) - \mu^3(X) \right] \Big/ \sigma^3(X) \quad \dots (21)$$

Where, X_i = Historical rainfall series; $\mu(x)$ = Mean of historical series and $\sigma(X)$ = Standard deviation.

RESULTS AND DISCUSSION

Historical rainfall data of 47 (1956–2003) years was collected from Agro-meteorological observatory, Junagadh Agricultural University, Junagadh. The computed values of serial correlation coefficients between two successive weekly rainfall sequences were very small. For most cases, they turned out to be non-significantly different from zero thereby suggesting that the weekly rainfall sequences are independently distributed (Clarke, 1973) (i.e. ± $2/n$).

The performance of power transformation in modifying the skewness and kurtosis is presented in Table 1. It is clear that in all weeks skewness has been brought to near the recommended limit as compared to the other transformations. The mean and the standard deviation of C_s were 0.088 and 0.267 and for kurtosis coefficient 2.6347 and 1.0899 respectively. These values were found to be lower than the historical series. However, when zero values were more than 60% of the sample size, this transformation was observed to be unable to normalize the data. This can be reflected from 23 and 40 standard weeks. Even if zeros were less than 60%, the normalized data series were not statistically reliable enough to give satisfactory probable values of transformed variables. This was because when this transformed series was plotted on to normal probability paper, it did not follow a true straight line, a desirable feature of any probability distribution; hence, the normality assumption was no more realistic for such cases. Power transformation was also unable to normalize data series having a very high standard deviation even though there were no zeros present. This leads to the conclusion that power transformation is effective only for data series that do not have many zeros and are moderately dispersed, which is again difficult to meet in case of weekly rainfall.

The annual rainfall series was disaggregated to finer scale (monthly). The estimated values of *A, B* and *C* for various months are shown in Table 2. Minimum *A* value of –1.9905 for the month of September and maximum of 1.0143 for the month of July was observed. The value of *A* for June and August was 1.00236 and 1.03530 respectively. Minimum *B* value of 1.484377 for July and maximum value of 4.4679372 for June was observed. The value of *B* for August and September was 2.73033 and 3.77941 respectively. Minimum *C* value of –0.03712 for June and maximum value of 0.178225 for August was observed. The value of *C* for July and September was 0.13423 and –0.09097 respectively. No trend of results in *A, B* and *C* with months was observed.

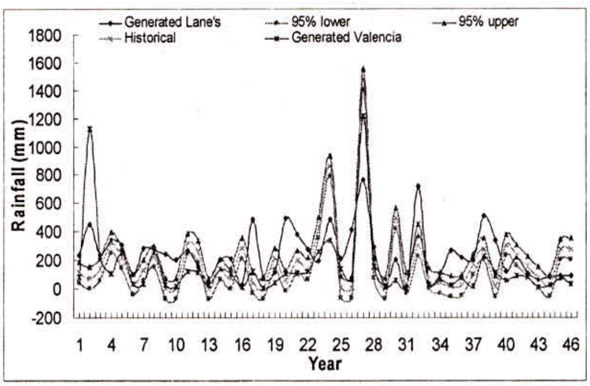
Table 1: Statistical Parameters of Historical and Transformed Rainfall Series

Week	*Historical Series*				*Transformed Series*		
	Mean (mm)	*SD (mm)*	C_s	C_k	λ	C_s	C_k
23	21.25	51.26	3.874	20.01	0.01	0.442	1.365
24	23.84	41.17	2.145	7.104	0.13	0	1.419
25	75.21	225.99	5.59	35.3	0.2	–0.05	3.205
26	70.68	115.31	3.104	15.16	0.26	0.013	2.131
27	77.5	91.36	1.419	4.981	0.35	0.002	2.04
28	87.7	110.81	1.126	5.615	0.32	0	2.546
29	103.03	116.48	1.265	4.073	0.34	0	2.187
30	69.55	101.87	1.545	9.61	0.28	0.043	3.289
31	57.82	76.2	2.365	10.1	0.31	0.035	2.938
32	60.22	95.82	2.477	8.81	0.21	0.078	4.301
33	82.73	235.49	5.41	33.09	0.22	0.034	5.532
34	23.16	34.51	2.805	11.62	0.31	0.005	2.743
35	51.73	106.72	3.906	19.21	0.24	0	3.286
36	30.47	40.24	1.946	5.724	0.32	0.024	2.04
37	24.68	42.93	3.147	14.74	0.25	0.036	2.195
38	25.63	49.87	3.1	12.86	0.01	0	1.168
39	12.01	17.56	1.4	3.82	0.16	0.009	1.415
40	10.6	27.02	3.2	12.83	0.01	1.08	2.35

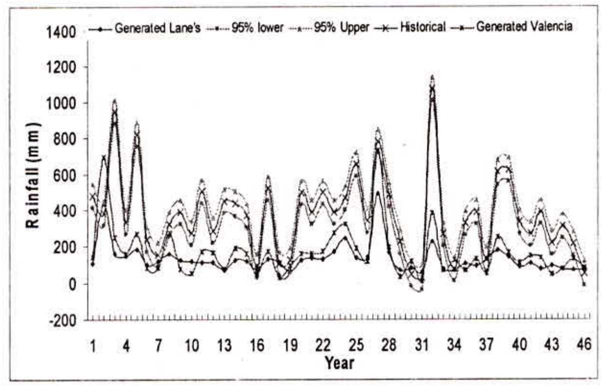
Table 2: Coefficients of the Lane's Condensed Model for Monthly Rainfall Series

Season	*Lane's Model*		
	A	*B*	*C*
June	1.0023688	4.4679372	−0.037120
July	1.0143161	1.4843775	0.134232
August	1.0353039	2.7303309	0.178225
September	−1.990589	3.7794180	−0.09097

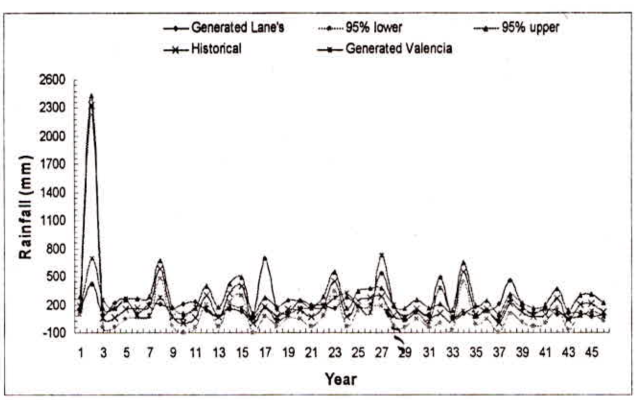
A comparison of historical and generated monthly series (Figure 1) showed that the generated values are slightly higher or slightly lower than historical series. The difference between historical and generated monthly rainfall are either positive or negative. The RMSE, skewness, AIC and the BIC values were determined using historical series and monthly-generated series for Lane's model (Table 3). The lowest value of these parameters indicates the better performance of the model.



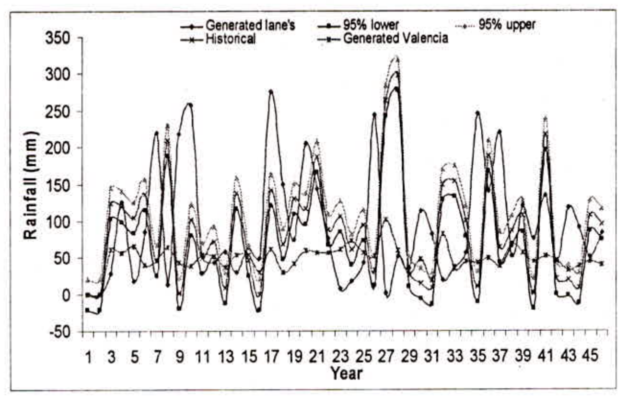
June



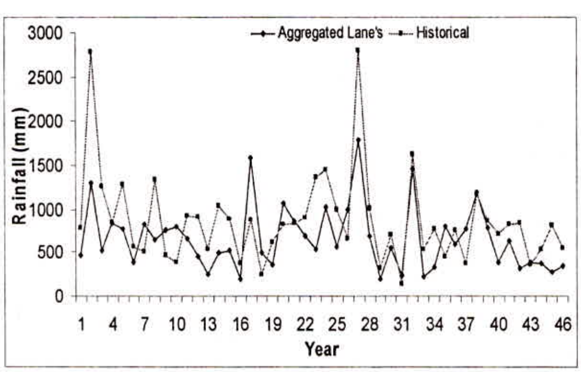
July



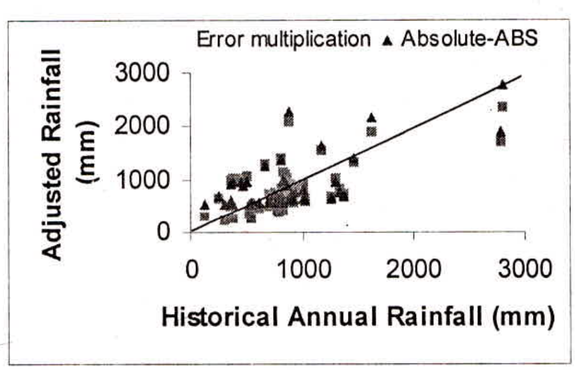
August



September



Annual Series without Adjustment



Adjusted and Historical Annual Rainfall Series

Fig. 1: Generated and Historical Rainfall Series Using Disaggregation Models for Monthly and Annual Series

Table 3: Error Functions for Lane's Monthly Disaggregation Model

Month	*RMSE*	*SKEWNESS*	*AIC*	*BIC*
June	228.51	3.446	13.022	1.564
July	311.68	2.885	13.766	1.339
August	332.23	6.674	13.919	2.399
September	111.18	1.887	11.295	0.803
Annual	368.24	3.023	7.040	7.417

Lowest value of RMSE, skewness, AIC and BIC were observed for the month of September. Lower RMSE was observed in June followed by July and August. Lower skewness and BIC was observed for July followed by June and August. No significant difference was found in AIC values. Although the seasonal variations are minimized by taking zero-mean normalized series and also various correlations, cross-correlations and covariance are preserved in parameter estimation of the model but still there is lack of consistency in the generated series due to highly variable nature of the data taken for the study.

COMPARISON BETWEEN HISTORICAL ANNUAL AND AGGREGATED ANNUAL RAINFALL SERIES

The aggregated annual series were obtained by adding the generated monthly series for the four months viz. June, July, August and September. The aggregated annual series were calculated for both the disaggregation models and are presented in Figure 1. The disaggregation models fail to preserve the additivity in generating the aggregated series, due to the fact that that rainfall data were transformed before using in the model. The aggregated annual series was compared with the annual historical series for Lane's disaggregation model. The visual inspection of the graph shows that the aggregated series had either less or more values than the historical series i.e. the residuals are either positive or negative (Figure 1). The error functions for aggregated series (Table 3) showed a poor performance of Lane's model in aggregated rainfall series from monthly values.

RAINFALL ADJUSTMENT PROCEDURES

The generated monthly rainfall fails to generate annual value upon addition. Two different procedures, proportional method, and ABS absolute difference method were employed for Lane's model in adjusting the generated annual and monthly series. The criteria for choosing an appropriate adjustment procedure are simplicity and the ability to preserve the fitted marginal distributions of the disaggregated monthly rainfall series. The second criterion may be tailored to individual situations; in particular, the high and low end of the marginal distribution might be given more weight depending on whether high or low data series were more critical at the time scale of interest.

Table 4: Error Functions for Adjusted and Historical Annual Rainfall Series Using Proportional Adjustment Method, ABS Difference Method

S. No.	*Error Functions*	*Proportional Method*	*ABS Method*
1.	RMSE	311.70	361.39
2.	Skewness	4.817	5.330
3.	AIC	6.890	7.067
4.	BIC	6.170	6.345

The absolute model yielded high RMSE, Skewness, AIC and BIC. The proportional adjustment method was found to give least value for the skewness and performed better than absolute method. The scatter plot diagrams between adjusted and historical rainfall series (Figure 1) indicate the visual distortions in the marginal distributions with the two adjustment procedures. The absolute model distorted more than the proportional adjustment procedure at the low and high ends of the distribution. The advantage of the proportional adjustment scheme may be explained by the hydrological characteristics of the rainfall series and the structure of the disaggregation model. Stedinger *et al.* (1985) observe that reproducing the sample estimate of the log-space mean and variance tends to produce slightly upwardly biased estimates of the real space moments for each month. They found that the proportional adjustment procedure counterbalanced this bias.

SUMMARY

Rainfall data is often required for engineering or hydrological purposes, but is also often severely lacking, both in terms of spatial coverage as well as length of recorded time. Condensed and staged disaggregation model are used to avoid proliferation of parameters. In the present study Lane's condensed disaggregation model is adopted. The problem of preserving annuity in disaggregation models is tackled through adopting two empirical correction procedures mainly proportional adjustment method and absolute difference method. The proportional adjustment method performed better than the absolute difference model.

REFERENCES

Grygier, J.C. and Stedinger, J.R. (1988). "Condensed disaggregation procedures and conservation corrections for stochastic hydrology". *Water Resou. Res.*, Vol. 24(10), pp. 1574–1584.

Hoshi, K. and Burges, S.J. (1979). "Disaggregation of stream volumes". *J. Hyd. Engrg.*, ASCE, pp. 105:27–41.

Lane, W.L. (1979). "Applied Stochastic Techniques", *User's Manual. Bureau of Reclamation, Engineering and Reach centre*, Denver Colorado.

Lane, W.L. (1982). "Corrected parameter estimates for disaggregation schemes, in Statistical Analysis of Rainfall and Runoff', edited by V.P. Singh, *Water Resources Publications*, Littleton, Colo.

Lane, W.L. and Frevert, D.K. (1990). "Applied Stochastic Techniques", *User's Manual. Bureau of Reclamation, Engineering and Research Center*, Denver, Co., Personal Computer Version.

Mejia, J.M. and Rousselle, J. (1976). "Disaggregation models in hydrology revisited", *Water Resou. Res.,* Vol. 12(2), pp. 185–186.

Rumelhart, D.E. and McClelland, J.L. (1986). "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", Foundations, Volume 1, MIT Press.

Stedinger, J.R. and Vogel, R.M. (1984). "Disaggregation Procedures for generating Serially Correlated Flow Vectors". *Water Resou. Res.*, Vol. 20(1), pp. 47–56.

Tao, P.C. and Delleur, J.W (1976). "Multistation, multiyear synthesis of hydrologic time series by disaggregation", *Water Resou. Res.*, Vol. 12(6), pp. 1303–1312.

Tarboton, D.G., Sharma, A. and Lall, U. (1998). "Disaggregation procedures for stochastic hydrology based on nonparametric density estimation". *Water Resou. Res* Vol. 34 (1), pp. 7–119.

Todini, E. (1980). "The preservation of skewness in linear disaggregation schemes", *J Hydro*, Vol. 47, pp. 199–214

Stedinger, J.R. and Vogel, R.M. (1984). "Disaggregation Procedures for generating Serially Correlated Flow Vectors.". *Water Resou. Res*. Vol. 20(1), pp. 47–56.

Valencia, D.R. and Schaake, J.C. (1973). "Disaggregation processes in stochastic hydrology", *Water Resou. Res*, Vol. 9(3), pp. 580–585.