A COMPARISON OF METHODS USED IN ESTIMATING MISSING RAINFALL DATA

R. P. De Silva[1], N.D.K. Dayawansa[1] and M. D. Ratnasiri[1]

ABSTRACT

Precipitation or rainfall (in tropics) is an important climatic parameter and the studies on rainfall are commonly hampered due to lack of continuous data. To fill the gaps (missing observations) in data, several interpolation techniques are currently used. However, the lack of knowledge on the suitability of these methods for Sri Lanka is a practical problem. In view of this problem, this study is aimed at comparing a few selected methods used for the estimation of missing rainfall data with a new method introduced by the authors to determine their suitability in Sri Lankan context. The methods studied were Arithmetic Mean (Local Mean) method, Normal Ratio method and Inverse Distance method. The new method introduced by the authors is named as Aerial Precipitation Ratio method.

In this approach, rain gauging stations where complete monthly rainfall data sets are available were selected in such a way that the selected stations represent each of the seven major Agro-ecological zones of Sri Lanka. This selection procedure of stations makes it possible to generalize the results to the entire country. The period of data ranged from 15 years in the case of mid country intermediate zone to 28 years Up country intermediate zone and Mid country wet zone. Subsequently, monthly rainfall data of each station were estimated using the data of surrounding stations based on the above selected methods so that actual data and the estimated data can be compared. Each estimated series was compared with the actual data series using different statistical comparison techniques. These comparisons include Descriptive Statistics of Error, Root Mean Square Error, Mean Absolute Percentage of Error and Correlation Coefficient. Results of the study show that the Inverse Distance method is the most suitable method for all three Low-country zones (wet, intermediate, and dry). However, for Mid-country and Upcountry Intermediate zones, Normal Ratio method is the most suitable method. Further, Arithmetic Mean method is more appropriate for Upcountry Wet zone while Aerial Precipitation Ratio method is more suitable for Mid-country Wet zone.

Keywords: *Rainfall, Missing data, Arithmetic Mean method, Normal Ratio method, Inverse Distance method, Aerial Precipitation Ratio method*

INTRODUCTION

Precipitation plays a significant role in agriculture and it is a major area in climatological studies (Ayoade, 1983). Studying about precipitation is important in (i) identifying precipitation characteristics; occurrence and temporal & spatial variability (ii) statistical modeling and forecasting of precipitation and (iii) resolving the problems such as floods, droughts, land slides, etc. In tropics, the term rainfall has acquired the place of precipitation, where snow is generally absent and the term precipitation is interchangeable with rainfall. The consistency and continuity of rainfall data are very important in statistical analyses such as

[1]*Department of Agricultural Engineering, Faculty of Agriculture, University of Peradeniya, Sri Lanka*

time series analysis. Both consistency and continuity may be disturbed due to change in observational procedure and incomplete records (missing observations) which may vary in length from one or two days to decades of years. However, inconsistency in a RF record can be identified by graphical or statistical methods such as Double mass curve analysis, the Von Neumann ratio test, cumulative deviations, likelihood ratio test, run test, etc. (Buishand, 1982). Nevertheless, filling of the gaps generated by inconsistent data is essential, and different procedures and approaches are available to accomplish this task. The most common method used to estimate missing rainfall data is Normal Ratio method (Chow *et al*, 1988). This method is based only on past observations of that rain gauge and surrounding gauges. However, there are other important factors such as distances among rain gauges, aerial coverage of each gauge etc., which are disregarded in this method but are proved to have significant influences on rainfall estimates. However, there are other techniques which use different other factors also to estimate missing rainfall data. This study focuses on few of them including Normal Ratio method, Inverse Distance method, and Arithmetic Mean method/ Local Mean method (Chow *et al*, 1988). The proposed Aerial Precipitation Ratio method by the authors looks at the area of influence of each surrounding gauge.

There are seven major climatic zones in Sri Lanka namely, Low-country wet zone, Low-country intermediate zone, Dry zone, Mid-country wet zone, Mid-country intermediate zone, Upcountry wet zone, and Upcountry intermediate zone (Agro-ecological map of Sri Lanka, 2003). The best method to estimate missing rainfall data can be different for each climatic zone depending on the rainfall pattern and spatial distribution.

OBJECTIVES

The main objective of the study is to identify the best method for specific climatic zone for the estimation of missing rainfall observations. The specific objectives of the research are to develop and introduce a new method for missing data estimation, compare and evaluate the estimates obtained from each method, and to study whether the suitability of each method varies with the factors like climatic zone, topography, distribution of rain gauges etc.

MATERIALS AND METHODS

The monthly rainfall data were used in this study. For each climatic zone, a cluster of four to five rain gauging stations was selected and altogether 31 stations were considered for the study (Table 01). Stratified random sampling method was used to select rain gauging stations for the study. The monthly rainfall data of selected stations were estimated using selected techniques based on the observations of surrounding stations. Details of data availability are given in Table 02. In some of these stations, data for one or more years were missing.

In the analysis, all those years were excluded for all the stations within that cluster. In the instances where none or only few months had missing values, the averages of those particular months were used in place of missing data. In order to test the accuracy of methods used in estimation of missing data, a rain gauge station (X) and neighbouring stations, for which data are available, are selected and assumed that observations from X station are missing. Then using each method, observations for X station are estimated and compared with the actual observations.

Table 01: Selected Rain gauging stations and their locations

Climatic Zone	Rain gauging stations	Location*	
		X (m)	Y (m)
Upcountry wet (WU)	Udaradella	192000	196000
	Abbergeldie Group	175800	191750
	Holmwood Estate	193500	183500
	Seeta Eliya	203250	192500
Upcountry Intermediate (IU)	Katukithula	188000	209750
	Kurundu-Oya	206500	208500
	Alma Estate	206000	209750
	Gonapitiya Estate	203000	205500
	Liddesdale	209000	202500
	Maha Uva Estate	209750	208750
Mid-country wet (WM)	Rassagala Estate	181500	166500
	Dethanagala Estate	192000	171000
	Pettigala Estate	188000	160750
	Alupola Estate	177500	167500
Mid-country Intermediate (IM)	Kundasale Farm	191500	232250
	Galphele (Wattegama)	192122	237680
	Kandy Kings Pavillion	185000	233500
	Delta Estate	190000	222500
Low-country wet (WL)	Agalawatta	132000	148000
	Pimbura Estate	133500	152750
	Bombuwela Agmet	116000	153000
	Sirikandura Estate	130500	144500
Low-country Intermediate (IL)	Mapalana	177000	95000
	Denagama	186500	98500
	Thihagoda	177848	89450
	Kamburupitiya	177000	98000
Dry Zone (DL)	Nachchaduwa	166868	336500
	Mihintale	172358	347480
	Anuradhapura	156956	345284
	Maha-Illuppallama	167000	322250

* Location information is given using National Grid (Sri Lanka) based on Transverse Mercator Projection.

Table 02: Availability of Rainfall data and Period of Data Availability

Principle Station	Used surrounding stations	Years of data used		
		From	To	Years
Udaradella (WU)	Abbergeldie Group Holmwood Estate Seeta Eliya Katukithula	1976	1999	24
Kurundu-Oya (IU)	Alma Estate Gonapitiya Estate Liddesdale Maha Uva Estate	1970	2000	28
Rassagala Estate (WM)	Dethanagala Estate Pettigala Estate Alupola Estate	1972	1999	28
Kundasale Farm (IM)	Galphele (Wattegama) Kandy Kings Pavillion Delta Estate	1979	1999	15
Agalawatta (WL)	Pimbura Estate Bombuwela Agmet Sirikandura Estate	1976	2000	25
Mapalana (IL)	Denagama Thihagoda Kamburupitiya	1971	1999	21
Nachchaduwa (DL)	Mihintale Anuradhapura Maha-Illuppallama	1970	1999	19

Arithmetic Mean method/ Local Mean method

If the normal annual precipitations at surrounding gauges are within the range of 10% of the normal annual precipitation at station X, then the Arithmetic procedure could be adopted to estimate the missing observation of station X (Chow *et al*, 1988). This assumes equal weights from all nearby rain gauge stations and uses the arithmetic mean of precipitation records of them as estimate (Tabios & Salas, 1985).

Normal Ratio method

This method is used if any surrounding gauges have the normal annual precipitation exceeding 10% of the considered gauge. This weighs the effect of each surrounding station (Singh, 1994). The missing data are estimated by,

$$P_X = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{N_X}{N_i}\right]P_i$$

where,
P_x = Estimate for the ungauged station
P_i = Rainfall values of rain gauges used for estimation
N_x = Normal annual precipitation of X station
N_i = Normal annual precipitation of surrounding stations
m = No. of surrounding stations

Inverse Distance method

In this method, weights for each sample are inversely proportionate to its distance from the point being estimated (Lam, 1983).

$$P_X = \frac{\sum_{i=1}^{N}\frac{1}{d^2}p_i}{\sum_{i=1}^{N}\frac{1}{d^2}}$$

where,
P_x = estimate of rainfall for the ungauged station
P_i = rainfall values of rain gauges used for estimation
d_i = distance from each location the point being estimated
N = No. of surrounding stations

Aerial Precipitation Ratio (APR) method

This method was developed based on spatial distribution of daily rainfall without accounting for the historical recurrence. The method leads the extension of point rainfall records to Thiessen Polygon areas. The APR method assumes the contribution of rainfall from surrounding stations is proportionate to the aerial contribution of each sub catchment (Thiessen polygon area claimed by each station without considering the missing gauge), when the station of missing values is excluded (De Silva, 1997). The formula of the method can be given as follows.

$$P_X = \frac{\sum_{i=1}^{N}\left[\left(A_j - A_i\right)P_i\right]}{\sum_{i=1}^{N}\left(A_j - A_i\right)}$$

$\sum_{i=1}^{N}\left(A_j - A_i\right)$ = Thiessen Polygon area for the station with missing values

A_j =Thiessen Polygon area when station with missing values is excluded

A_i = Thiessen Polygon area when station with missing values is included

P_i = annual precipitation of surrounding stations

P_x = estimate for monthly rainfall for the station with missing observations

Comparison of Estimates

The estimates obtained from each method are compared with actual records. The suitability of method is decided by how close the estimates and actual values are in a given time series. Several 'Descriptive statistics of error' can be used as criteria to estimate the closeness of estimated and actual values. These Descriptive statistics of error include Mean (μ), Standard Deviation (S), Correlation Coefficient (r), Root Mean Square Error (RMSE), Mean Absolute Percentage (MAPE).

RESULTS AND DISCUSSION

Error Means and Error Standard Deviations (SD)

Among Descriptive Statistics of Error or deviation between actual value and estimate, Error Mean is the representative value of the error. The SD of Error indicates the fluctuations of the deviations. The Error Means and Error SDs are presented in Table 03. The minimum error mean and minimum SD for all low country stations were recorded for Inverse Distance (ID) method. Both Intermediate zone stations (IU &IM) recorded minimum mean as well as minimum SD for Normal Ratio (NR) method. Records of WU and WM zones had no clear pattern like above and minimum mean and SD for WU were given by arithmetic mean (AM) method and minimum mean and minimum SD for WM were given by Aerial Precipitation Ratio (APR) method.

Root Mean Square Error (RMSE)

This also shows similar results to Mean and SD of error as shown in Table 04. Low country zones (WL, IL, and DL) gave least RMSE for ID method. Mid country and up country Intermediate zones gave minimum RMSE when estimated by NR method. Being similar to Mean and SD of error, minimum RMSE for WU was given by arithmetic mean (AM) method and minimum RMSE for WM was given by Aerial Precipitation Ratio (APR) method.

Correlation Coefficient

This is an indicator for the strength of the relationship between observations and estimates. Higher positive coefficients indicate that estimates will be high or low when actual is high or low respectively giving evidence about the suitability of estimation method. The correlation coefficients of each method studied are given in Table 05. The results of this parameter also agreed with the results of descriptive statistics of the error and RMSE. For all low country stations (WL, IL and Dl) the highest Correlation coefficient was resulted with ID method. Two intermediate zones (WU and IU) recorded maximum values for NR method. WU and WM showed highest Correlation Coefficients by arithmetic mean (AM) method and Aerial Precipitation Ratio (APR) method, respectively.

Mean Absolute Percentage Error (MAPE)

This indicates the deviation of the estimate value from the observed (actual) value with respect to the observed value. The calculated MAPE values are given in Table 06. WU, IM, and DL gave minimum values for Normal Ratio method while WL and IU gave minimum value for Inverse Distance method. The WM and IL zones gave minimum values for Aerial Precipitation Ratio method and Arithmetic Mean method, respectively. According to the results of MAPE, it does not give any clear pattern in suitability of methods for different zones.

Table 03: Error means and Error Standard Deviations for each method for seven climatic zones

	AM method		NR method		ID method		APR method	
	Error Mean	Error SD	Error Mean	Error SD	Error Mean	Error SD	Error Mean	Error SD
WU	66.01	65.18	69.25	73.06	68.52	68.83	67.66	69.04
WM	80.42	82.13	80.43	75.56	82.65	78.30	75.94	75.46
WL	65.12	54.25	53.64	50.49	52.41	48.03	53.99	48.51
IU	61.69	80.74	50.03	79.20	65.27	87.40	63.08	81.69
IM	69.38	56.23	34.62	31.63	64.64	52.12	61.10	50.88
IL	55.42	65.39	53.53	64.14	51.26	57.57	57.44	69.63
DL	41.76	56.56	37.86	52.78	36.82	51.30	41.84	56.50

Table 04: Root Mean Square Error for each method for seven climatic zones

Climatic Zone	AM method	NR method	ID method	APR method
WU	92.6919	100.574	97.0357	96.586
WM	114.853	110.278	113.771	106.973
WL	84.7020	73.6101	71.0394	72.5278
IU	100.299	94.8881	108.982	103.11
IM	89.2114	46.8298	82.9418	79.4189
IL	85.6193	83.4512	76.9957	90.1547
DL	70.2076	64.8595	63.0503	70.2026

Table 05: Correlation Coefficients for each method for seven climatic zones

Climatic Zone	AM method	NR method	ID method	APR method
WU	0.83898	0.83213	0.83137	0.82786
WM	0.83042	0.82113	0.80257	0.834
WL	0.94461	0.9414	0.95016	0.94744
IU	0.89463	0.89596	0.87184	0.88539
IM	0.84837	0.8615	0.85863	0.85731
IL	0.79414	0.80579	0.83211	0.7758
DL	0.78208	0.78221	0.78258	0.78191

Table 06: Mean Absolute Percentage Error for each method for seven climatic zones

Climatic Zone	AM	NR	ID	APR
WU	86.7516	76.1584	85.2407	84.0986
WM	35.6984	42.3452	37.1841	34.3058
WL	21.9117	20.9618	18.6136	19.2974
IU	48.6153	50.6863	47.1127	49.5192
IM	118.534	64.5799	115.905	115.147
IL	46.9635	50.0270	48.8454	51.3612
DL	110.575	88.6641	96.1255	110.223

CONCLUSIONS

In estimating missing rainfall data, for Low country stations (WL, IL & DL) Inverse Distance method is the most suitable method among the methods studied. For Mid country and Upcountry Intermediate Zone stations (IM & IU), Normal Ratio method is the most suitable method compared to other three methods. Arithmetic mean method is more suitable for Upcountry Wet Zone and Aerial Precipitation Ratio method is more suitable for Mid country Wet Zone. The degree of suitability of these estimation methods for each zone needs to be determined and validated by further studies.

References

Ayoade, J O. (1983). *Introduction to Climatology for the Tropics*. John Wiley and Sons: New York.

Buishand T A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*. 58: 11-27.

Chow V.T., Maidment D.R. and L.W. Mays (1988). Applied Hydrology, Mc Graw Hill Book Company, ISBN 0-07-010810-2.

De Silva R P. (1997). *Spatiotemporal Hydrological Modeling with GIS for the Upper Mahaweli Catchment, Sri Lanka.* PhD Thesis (Unpublished), Silsoe Campus, Cranfield University, UK.

Lam N S. (1983). Spatial Interpolation Methods review. *The American Cartographer.* 10: 129-149.

Singh V P. (1994). *Elementary Hydrology*. Prentice Hall of India: New Delhi.

Tabios G. Q. and Salas J. D. (1985). *A comparative analysis of techniques for spatial interpolation of precipitation.* Water Resource Bull. 21(3), 365-380.