APPLIED COMPUTING, MATHEMATICS AND STATISTICS GROUP
Division of Applied Management and Computing

Spatial Interpolation of Rainfall in the Dry Zone of Sri Lanka

B.V.R. Punyawardena and Don Kulasiri

Research Report No: 99/05
March 1999

ISSN 1174-6696

RESEARCH REPORT

LINCOLN UNIVERSITY
Te Whare Wānaka O Aoraki

Applied Computing, Mathematics and Statistics

The Applied Computing, Mathematics and Statistics Group (ACMS) comprises staff of the Applied Management and Computing Division at Lincoln University whose research and teaching interests are in computing and quantitative disciplines. Previously this group was the academic section of the Centre for Computing and Biometrics at Lincoln University.

The group teaches subjects leading to a Bachelor of Applied Computing degree and a computing major in the Bachelor of Commerce and Management. In addition, it contributes computing, statistics and mathematics subjects to a wide range of other Lincoln University degrees. In particular students can take a computing and mathematics major in the BSc.

The ACMS group is strongly involved in postgraduate teaching leading to honours, masters and PhD degrees. Research interests are in modelling and simulation, applied statistics, end user computing, computer assisted learning, aspects of computer networking, geometric modelling and visualisation.

Research Reports

Every paper appearing in this series has undergone editorial review within the ACMS group. The editorial panel is selected by an editor who is appointed by the Chair of the Applied Management and Computing Division Research Committee.

The views expressed in this paper are not necessarily the same as those held by members of the editorial panel. The accuracy of the information presented in this paper is the sole responsibility of the authors.

This series is a continuation of the series "Centre for Computing and Biometrics Research Report" ISSN 1173-8405.

Copyright

Correspondence

This paper represents work to date and may not necessarily form the basis for the authors' final conclusions relating to this topic. It is likely, however, that the paper will appear in some form in a journal or in conference proceedings in the near future. The authors would be pleased to receive correspondence in connection with any of the issues raised in this paper. Please contact the authors either by email or by writing to the address below.

Any correspondence concerning the series should be sent to:

The Editor
Applied Computing, Mathematics and Statistics Group
Applied Management and Computing Division
PO Box 84
Lincoln University
Canterbury
NEW ZEALAND

Email: computing@lincoln.ac.nz

SPATIAL INTERPOLATION OF RAINFALL IN THE DRY ZONE OF SRI LANKA

B. V. R.. Punyawardena, Natural Resources Management Centre, Department of Agriculture, Sri Lanka

Don Kulasiri, Applied Management and Computing Division, Lincoln University, Canterbury, New Zealand

Abstract: One of the problems which often arises in climatology is either data at a given site is missing or the site is ungauged. In this study, a spatial interpolation model was developed to estimate the weekly rainfall of the Dry zone of Sri Lanka at ungauged sites assuming that the spatial continuity of rainfall at two neighbouring locations are exponentially correlated. Twenty years of weekly rainfall data from six stations located in the Dry zone was used in the study. To support the methodology, the results of the exponential model were compared with the other two methods of spatial interpolation techniques, namely, the local mean and the inverse distance methods. The results of the study indicate that the exponential correlation model is a promising candidate for estimating mean weekly rainfall of the Dry zone. However, the local mean and the inverse distance methods compare quite well along with the exponential model, indicating that more complex models have no particular advantage over simple models for estimating rainfall in the Dry zone of Sri Lanka. Nevertheless, the results point towards the relative importance of the exponential model as opposed to the other two models when the neighbouring locations do not have long series of historical records.

Key words: Dry zone, Rainfall, Spatial interpolation, Exponential correlation

INTRODUCTION

The complex, interacting atmospheric processes that give rise to rainfall make it a variable phenomenon across the landscape. Therefore, recorded rainfall from a rain gauge usually represents only an extremely small area of the catchment. Most of the time rain gauges in the Dry zone are usually separated by several kilometres. Therefore, existing network of gauges may not be sufficient enough to estimate the parameters that are needed for

hydrological and climatological applications. This problem is further aggravated by the frequent appearance of missing data in the observed rainfall sequences. Thus, there is a need for a methodology of spatial interpolation of rainfall that uses only minimum available data.

Spatial interpolations of data available at other sites are being used in the field of hydrology and climatology to generate the data for ungauged locations. In most cases, simple methods of point estimation are applied. The availability of computing facilities has encouraged the development of advanced methods of interpolation. As a result, a number of spatial interpolation techniques are available today with varying degrees of complexity such as local mean, Thiessen polygon, inverse distance, inverse square distance, isohyetal and krigging.[1, 2] Some of them are very simple with limited applicability while others involve complex mathematical frameworks and needs a large number of data points to obtain a reasonable level of accuracy.

METHODS AND MATERIALS

Spatial continuity exists in most earth science data sets and two data sets close to each other are expected to have closer values than those that are far apart.[3] A function can be developed to describe the continuity of the relationship between the value of one variable at a point and the value of the same variable at another point, a given distance away.[1] Correlation, covariance and variogram functions have been used to express the spatial continuity of a random variable. Similar assumptions have been made about rainfall phenomena over an area, and estimation methods used in earth science have been applied to rainfall data to estimate the values at ungauged sites.

The spatial correlation models for rainfall have been presented in inverse power and exponential forms[4]:

$$\gamma_{ab} = (1+\alpha d)^{-c} \qquad (1)$$

$$\gamma_{ab} = e^{-\alpha d} \qquad (2)$$

where,

γ_{ab} = spatial correlation coefficient between stations A and B

α = a coefficient

c = a power coefficient

d = distance between the pair of stations

Assuming homogeneity and isotropicity,[5] a study was undertaken to determine the appropriateness of point estimation of weekly rainfall by an exponential spatial correlation model for the Dry zone's climatic environment. In this study, two distinctive regions of the Dry zone were considered, the north-central part and the southern part of the Dry zone (Figure 1). Both regions exhibit fairly similar physiography of gently-undulating to rolling, with 3 to 4 per cent slopes. However, some geographical features are not alike. The north-central part of the Dry zone, abbreviated NCDZ, is generally an inland region. The southern part of the Dry zone, abbreviated SDZ, resembles an area that is more closer to the ocean. Therefore, the amount of water vapour in the atmosphere, what is available to become cloud with the chance of subsequently becoming rain, may not be comparable in the two regions. Thus, the correlation structure of the rainfall process could be different in the two regions. This necessitates the evaluation of the spatial correlation model for the two regions separately to meet the assumptions made on the isotropicity and homogeneity.

The selected rainfall recording stations from the NCDZ region are located at Maha-Illuppallama, Pelwehera and Maradankadawala. Out of these three stations, Maradankadawala which lies in between the other two stations was considered as the location where rainfall values to be estimated. The areal distances from Maradankadawala to Maha-Illuppallama and from Maradankadawala to Pelwehera are 17 km and 25 km respectively, while the areal distance between Maha-Illuppallama and Pelwehera is 38 km. From the SDZ region which represents a coastal area, Angunakolapellessa, Ambalantota and Weerawila were selected for the study. In this region, Ambalantota which lies in between the other two stations was considered as the location where the rainfall values to be estimated. The areal distances from Ambalantota to Angunakolapellessa and from Ambalantota to Weerawila are

15 km and 27 km respectively, while the areal distance between Angunakolapellessa and Weerawila is 38 km. In the selection of the rainfall recording stations, care was given to select the locations with reliable data with a maximum number of record lengths to be on par with the guidelines stipulated by the Hydrology and Water Resources Program, Department of Civil Engineering, Colorado State University.[5] The said guidelines prescribe that data records with more than 30 years should be used. But most of the time, the available length of the records from the selected locations were 20 years. Although there are some other locations in the Dry zone which have the minimum of 30 years of records, a large number of missing data and unreliability of the measurements forced us not to select them for the study.

In addition, the models were also evaluated for situations with short series of historical data and when the stations are located relatively far away. Weekly rainfall values for Angunakolapellessa were interpolated using 10 years of historical rainfall records from two neighbouring locations, Embilipitiya and Tangalle. As the reliable rainfall data from the immediate vicinity of a rainfall station in the Dry zone was unavailable, these two locations were selected from the neighbouring Intermediate zone. The areal distances from Embilipitiya to Angunakolapellessa and from Angunakolapellessa to Tangalle are 18 km and 20 km respectively, while the areal distance between Embilipitiya and Tangalle is 35 km. An evaluation of the interpolation models for the stations located farther apart was carried out by interpolation of weekly rainfall at Weerawila using Tangalle and Lahugala (eastern part of the Dry zone). The areal distance from Lahugala to Tangalle is 143 km whereas distances from Lahugala to Weerawila and from Weerawila to Tangalle are 93 and 58 km respectively.

The spatial correlation coefficient γ_{ab} for weekly rainfall values can be determined using contemporaneous observation pairs from stations A and B. Using the calculated γ_{ab} and the distance between stations A and B, the coefficient (α) of equation (2) can be found.

$$\gamma_{ab} = e^{-\alpha d}$$

$$\ln \gamma_{ab} = -\alpha d$$

$$\alpha = -\frac{\ln \gamma_{ab}}{d} \qquad (3)$$

The observed spatial correlation coefficients between the two stations, A and B, and the value for α from equation (3) can be used in equation (2) to estimate the correlation coefficients between stations A and B with the station C (γ_{ac} and γ_{bc}). Let the unbiased linear estimator for the normalised rainfall at station C be:

$$R_c^* = w_a . R_a^* + w_b . R_b^* \qquad (4)$$

where

R_c^* = estimated normalised rainfall at station C

R_a^* = observed normalised rainfall at station A

R_b^* = observed normalised rainfall at station B

w_a = weight assigned to the station A

w_b = weight assigned to the station B

The least squares regression for equation (4) can be written in matrix notation:

$$\tilde{C} \quad . \quad w \quad = \quad \gamma$$

$$\begin{bmatrix} 1 & \gamma_{ab} & 1 \\ \gamma_{ab} & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_a \\ w_b \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{ac} \\ \gamma_{bc} \\ 1 \end{bmatrix} \qquad (5)$$

The matrix $\tilde{C}$ consists of the covariance value of rainfall between the two sample locations. The vector γ consists of the covariance values of rainfall between two sample locations and the location where we need the estimation. The vector w consists of the weight given to each location and the Lagrange parameter μ[3]. To solve for the weights, multiply both sides of equation (5) by $\tilde{C}^{-1}$.

$$w = \tilde{C}^{-1} . \gamma \qquad (6)$$

Based on the homogeneity and isotropicity assumptions, the estimated mean rainfall and the standard deviation of the rainfall at station C can be calculated using the following linear estimation:

$$\hat{R}_c = AC\left[\frac{\overline{R}_a - \overline{R}_b}{AB}\right] \tag{7}$$

$$\hat{\sigma}_c = AC\left[\frac{\sigma_a - \sigma_b}{AB}\right] \tag{8}$$

where

$\hat{R}_c$ = estimated mean rainfall at station C

$\overline{R}_a$ = observed mean rainfall at station A

$\overline{R}_b$ = observed mean rainfall at station B

$\hat{\sigma}_c$ = estimated standard deviation of rainfall at station C

σ_a = observed standard deviation of rainfall at station A

σ_b = observed standard deviation of rainfall at station B

AB = distance between stations A and B in km

AC = distance between stations A and C in km

Once the above parameters are determined from the observed data, equation (4) can be used to estimate the rainfall in each week of the year.

RESULTS

The validity and applicability of the foregoing interpolation model was examined by comparing the model output with the observed data from Maradankadawala and Ambalantota. In addition, a further comparison of the model output was made with the other two interpolation techniques, namely, the local mean method and the inverse distance method. Use of local mean or the arithmetic mean in spatial interpolation is the most simplistic approach. It assumes equal weight from all nearby sample locations, using the sample mean as the estimate. Inverse distance method is a technique which gives more weight

to the closest samples and less to those that are farthest away. Thus, weight for each sample is inversely proportional to its distance from the point being estimated:

$$\hat{R} = \frac{\sum_{i=1}^{n} \frac{1}{d_i} v_i}{\sum_{i=1}^{n} \frac{1}{d_i}} \qquad (9)$$

where

$\hat{R}$ = estimate of rainfall for ungauged location

v_i = observed value at the i^{th} location

d_i = distance from each location to the point being estimated

Comparison of estimated and observed rainfall

Figures 1 and 2 show the mean estimated and observed rainfall in each week for Maradankadawala and Ambalantota respectively. Typically, we want a set of estimates that comes as close as possible to the true values. Thus, we would prefer the results shown in Figures 1 and 2. There was no significant difference between the observed values and the estimated values at both Ambalantota and Maradankadawala. The standard deviations of the observed sequences of rainfall were comparable with the estimated sequences of rainfall from the exponential model (Tables 1 and 2). However, most of the time the variability of the estimated values from the exponential model was less than that of the observed variability. This trend was more apparent at Ambalantota in the SDZ region. Reduced variability of estimated values is often referred to as "smoothing" and is a consequence of combining two or more sample values to form an estimate.[3] As more sample values are incorporated in a weighted linear combination, the resulting estimates generally become less variable.

Figure 3 shows the performance of the exponential model in interpolating weekly rainfall at Angunakolapellessa with a short series of historical data from neighbouring two locations. It is clear that the estimated data do not represent the observed data well compared to the estimations at Maradankadawala and Ambalantota. However, the differences between the estimated mean values and the observed mean values were not significant at the 5% probability level. The estimated mean weekly rainfall at Weerawila from the historical data of Tangalle and Lahugala are shown in Figure 4. It is clear that there is a distinct deviation of the estimated mean values from the observed mean values. In general, these mean deviations were significant at the 5% probability level, especially during the second half of the year.

Comparison between different interpolation methods

The results of the other two interpolation methods described in a preceding section of this paper were compared with the outcome of exponential correlation model. As the first criterion for comparing the different methods, the means in each week were computed. Figures 5 and 6 show the means of weekly interpolated rainfall values from the three methods for Maradankadawala and Ambalantota, respectively. It may be seen that practically all of the interpolation techniques reproduce the means well. None of these means were significantly different from each other and also from the observed values. The estimated values from all three models at Maradankadawala are almost identical (Figure 5). At Ambalantota, though it is not significant, a small discrepancy between estimated values from the three models is noticeable during the two dry periods and during the Yala season, mid-March to mid-May, (Figure 6). The differences between estimated values from the three

models at both Angunakolapellessa and Weerawila were also not significant. However, most of the time the closset value to the observed value was found with the exponential model.

Another way of checking the appropriateness of the model is to calculate the correlation coefficient between the observed and the estimated values. It is a good index for summarising how close the points on a scatter plot come to falling on a straight line, and therefore can make use to compare different estimation models. The correlation coefficient between the observed and the estimated values from each model was calculated for every week of the year. These values were averaged over a four different time periods of the year, namely, first dry season (early February to mid March), second dry season (late March to late September), Yala season (mid March to mid May) and Maha season (early October to late January). At both Maradankadawala and Ambalantota, the seasonal correlation coefficients were always above 0.65 except during the Yala season at Ambalantota (Table 3). It is interesting to note that when the correlation between the estimated and the observed values is low, it is consistent with all the three models.

The correlation coefficients between the estimated and the observed values at Angunakolapellessa were always above 0.5 with all the three models where the estimation was based only on 10 years of historical data (Table 4). However, at Weerawila, where the estimation was based on two neighbouring locations separated by over 100 km distance, only the first dry season exceeded the 0.5 boundary. During the Maha season, correlation coefficient was closer to 0.5 with the exponential model. But, during both the Yala and the second dry seasons, correlation coefficients were below 0.40 with all the three models (Table 4).

DISCUSSION

Overall, the results show that mean weekly rainfall of both Ambalantota and Maradankadawala are well preserved. However, the discrepancy between the observed and the estimated values at Maradankadawala is less than the same at Ambalantota. The correlation of rainfall between any two locations is highest for places which are close to each other, in flat country and away from the coast.[6] The areal distances between the two sample locations at both regions are almost equal. The topography of the two regions is also comparable to each other. Thus, closeness to the ocean could be the main determining factor for the small discrepancy between the observed and the estimated values at Ambalantota in SDZ region. However, the performance of the exponential model based on a short series of historical data (ie, 10 years) is not comparable to its performance with a long historical data. This confirms that the more historical data available, the better the estimation will be. The situation becomes worse when the neighbouring locations are further apart (ie, 100 km or in excess). This could be attributed to the fact that the correlation structure is being weakened when the stations are farthest away.

The exponential method and the inverse distance method should give a better estimation of weekly rainfall data compared to the local mean method. Because, they give a varying weight depending on the distance apart rather than a equal weight as in the case of local mean method. Therefore, it is reasonable to expect an improvement to the correlation with the exponential method and the inverse distance method over the local mean method. However, the correlation coefficient values of three models are almost similar at both Maradankadawala and Ambalantota (Tables 3) suggesting that performances of all the three models are similar under the given environments. Thus, if one is interested only in mean

rainfall, as is often the case in climatological applications, then there is no particular advantage in computing a complex exponential relationship; rather a simple inverse distance or local mean will suffice.

Although the difference between the correlation coefficient values among the three models is not large, both at Weerawila and Angunakolapellessa, the resulted correlation coefficient values from the exponential model were always higher than that from the other two models except during the Yala season at Weerawila (Table 4). Thus, when the neighbouring locations are farther away or especially with short series of historical data, the individual estimations of the exponential model would approximate the real values comparatively better than the other two models.

In conclusion, the results of this study suggest that the exponential correlation model is a promising candidate for estimating weekly rainfall of the Dry zone of Sri Lanka. However, the less sophisticated local mean and inverse distance methods rate quite well along with the exponential model. There is no particular basis to claim that the exponential model is significantly better than the other two methods tested under the given environments. Nevertheless correlation analysis shows an improvement to the estimates with the exponential model, especially when the neighbouring locations do not have long series of historical data.

Acknowledgment

This study is a part of graduate studies completed when the first author is on study leave at the Lincoln University, New Zealand. The first author is grateful to the Department of Agriculture, Sri Lanka and the Ministry of Foreign affairs and Trade, New Zealand for the study leave granted and the financial assistance during the study period, respectively. The

authors also indebted to Dr. V. Bidwell of the Lincoln Ventures, Lincoln University, New Zealand for stimulating discussions. Also, the authors wish to acknowledge two anonymous reviewers. Their comments improved and clarified several points of this paper.

References

1. Abtew, W., Obeysekara, J. and Shih, G. (1993). Spatial analysis of monthly rainfall in Florida. *Water Resources Bulletin* **29**(2): 179-188.

2. Singh, V.P. and Chowdhury, P.K. (1986). Comparing some methods of estimating mean areal rainfall. *Water Resources Bulletin* **22**(2): 275-282.

3. Issaks, E.H. and Srivastawa, R.M. (1989). *Applied Geostatistics.* pp 249-295. Oxford University Press, New York.

4. Yevjevich, V. and Karplus, A.K., (1973). Area-time structure of the monthly precipitation process. *Hydrologic papers*, Colorado State University, Fort Collins, Colorado, USA.

5. Tabios, G.Q. and Sallas, J.D. (1985). A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resources Bulletin* **21**(3): 365-380.

6. Linacre, E. (1992). *Climate and data resources. A reference and guide.* pp 250-311. Routledge, London and New York.

Table 1 Standard deviations of the observed and the estimated rainfall from the exponential model during major dry seasons at Maradankadawala and Ambalantota in the Dry zone of Sri Lanka.

	Maradankadawala		Ambalantota	
Standard Week No.	Observed (mm)	Estimated (mm)	Observed (mm)	Estimated (mm)
First dry season				
6	21.5	27.0	20.7	19.8
7	27.4	31.5	19.1	33.1
8	15.5	14.3	16.2	13.1
9	41.6	45.0	21.2	14.0
10	40.6	55.3	29.1	19.5
11	13.4	10.1	13.3	17.0
Second dry season				
22	30.9	14.6	26.9	24.6
23	4.2	5.5	12.6	11.1
24	6.4	5.1	16.2	11.6
25	0.9	2.5	27.8	13.7
26	1.8	4.2	10.6	9.0
27	10.9	11.7	12.6	7.1
28	28.6	29.0	15.2	10.0
29	25.5	23.5	18.2	17.0
30	21.6	10.1	10.5	8.1
31	4.6	11.3	11.9	7.8
32	28.2	18.8	26.4	16.2
33	1.4	2.3	16.1	9.5
34	15.4	13.6	11.9	5.4
35	9.5	4.2	26.6	11.1
36	5.7	7.9	8.5	6.3
37	38.2	38.7	19.5	17.9
38	35.3	23.5	21.2	16.0
39	33.7	44.3	22.2	17.5

Table 2 Standard deviations of the observed and the estimated rainfall from the exponential model during major rainy seasons at Maradankadawala and Ambalantota in the Dry zone of Sri Lanka.

	Maradankadawala		Ambalantota	
Standard Week No.	Observed (mm)	Estimated (mm)	Observed (mm)	Estimated (mm)
Rainy season - Yala				
12	16.0	14.5	14.6	12.6
13	31.6	17.2	19.8	14.9
14	31.3	33.8	14.6	21.0
15	30.8	37.1	26.9	24.7
16	57.8	54.8	22.0	29.5
17	43.8	50.1	14.8	19.9
18	48.4	43.6	18.3	14.7
19	25.7	18.7	23.2	25.8
20	29.7	33.0	27.1	21.0
21	27.0	29.7	20.2	28.7
Rainy season - Maha				
40	42.8	51.6	17.1	16.2
41	49.6	57.6	18.5	16.3
42	49.4	44.4	26.5	26.2
43	58.7	62.2	50.3	26.2
44	62.8	51.4	39.1	28.9
45	39.9	38.0	40.9	26.0
46	60.8	59.5	62.6	49.0
47	52.2	56.6	29.0	24.4
48	50.9	47.3	37.1	26.0
49	27.0	31.3	18.5	17.6
50	34.0	30.6	30.2	24.6
51	60.5	63.3	14.4	19.6
52	33.4	30.9	16.6	10.9
1	52.3	47.9	26.3	20.1
2	51.6	45.8	19.5	15.1
3	21.3	33.5	30.6	35.2
4	10.5	11.6	4.3	8.1
5	21.6	14.7	4.6	7.4

Table 3 Seasonal correlation coefficients between the observed and the estimated values from the three models at Maradankadawala and Ambalantota.

	Exponential model	Inverse distance model	Local mean model
	Maradankadawala		
First dry season	0.77	0.77	0.78
Second dry season	0.66	0.66	0.65
Yala season	0.72	0.72	0.73
Maha season	0.77	0.77	0.77
	Ambalantota		
First dry season	0.76	0.74	0.74
Second dry season	0.71	0.72	0.71
Yala season	0.41	0.40	0.38
Maha season	0.69	0.69	0.68

Table 4 Seasonal correlation coefficients between the observed and the estimated values from the three models at Angunakolapellessa and Weerawila.

	Exponential model	Inverse distance model	Local mean model
	Angunakolapellessa		
First dry season	0.79	0.76	0.76
Second dry season	0.67	0.63	0.64
Yala season	0.52	0.50	0.50
Maha season	0.63	0.57	0.57
	Weerawila		
First dry season	0.62	0.59	0.59
Second dry season	0.34	0.31	0.32
Yala season	0.36	0.38	0.39
Maha season	0.49	0.45	0.45

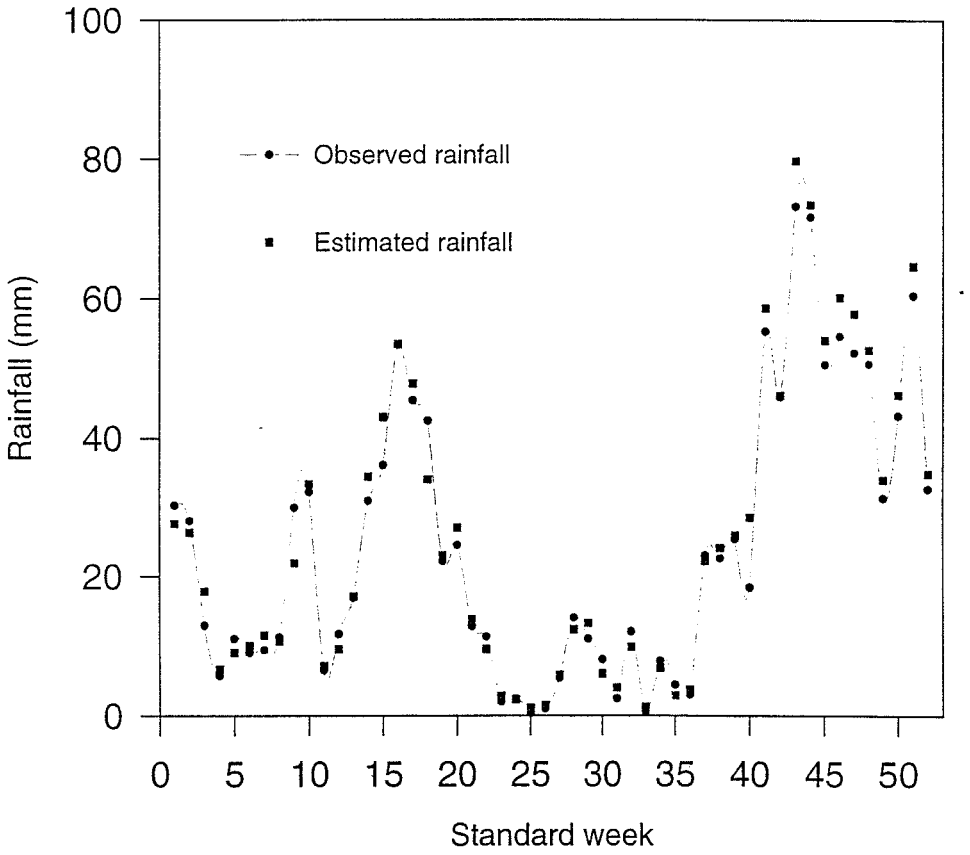

Figure 1 Observed and estimated rainfall at Maradankadawala in the Dry zone of Sri Lanka.

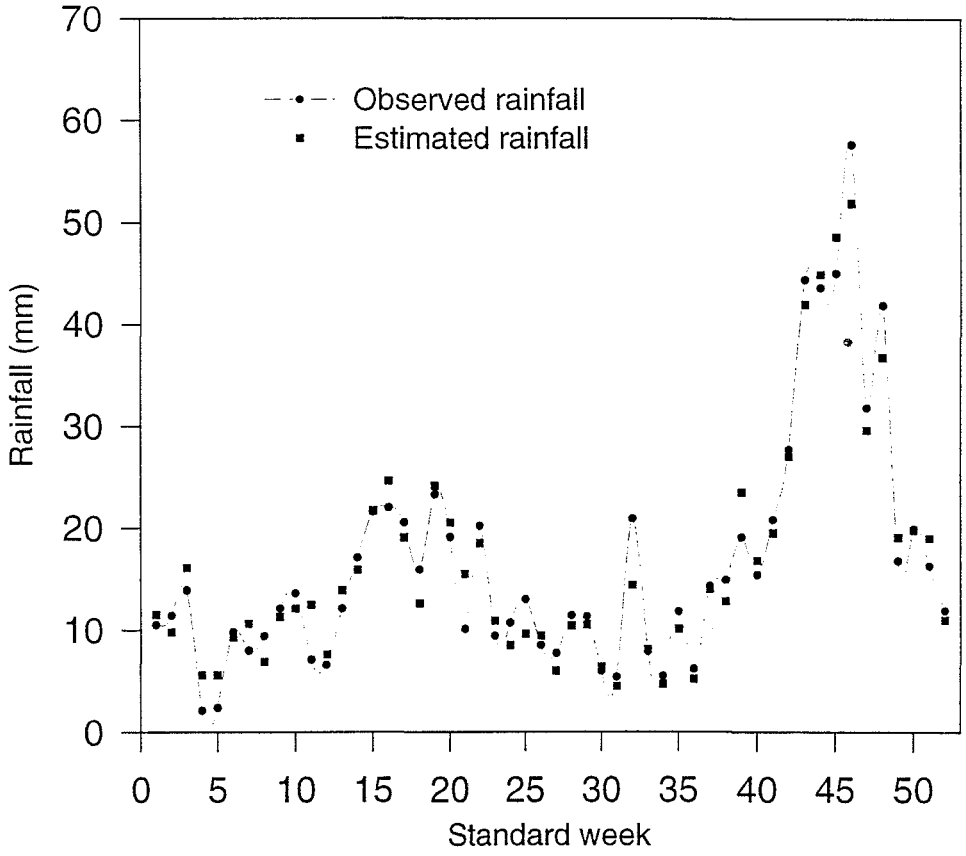

Figure 2 Observed and estimated rainfall at Ambalantota in the Dry zone of Sri Lanka.

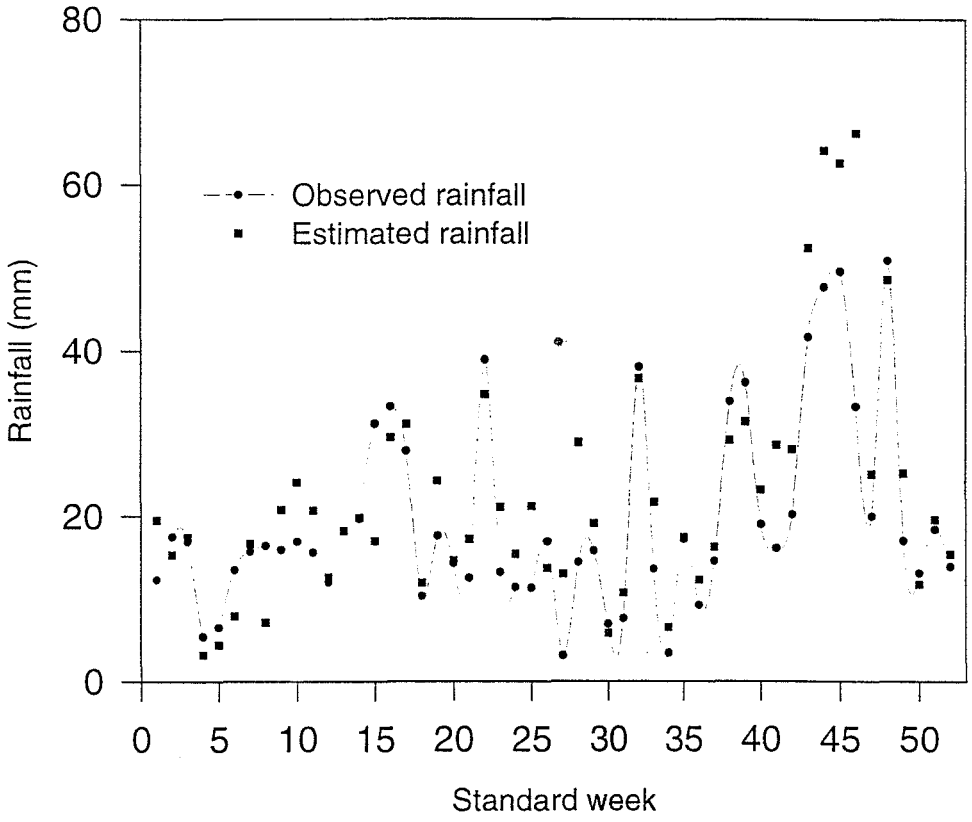

Figure 3 Observed and estimated rainfall at Angunakolapellessa in the Dry zone of Sri Lanka.

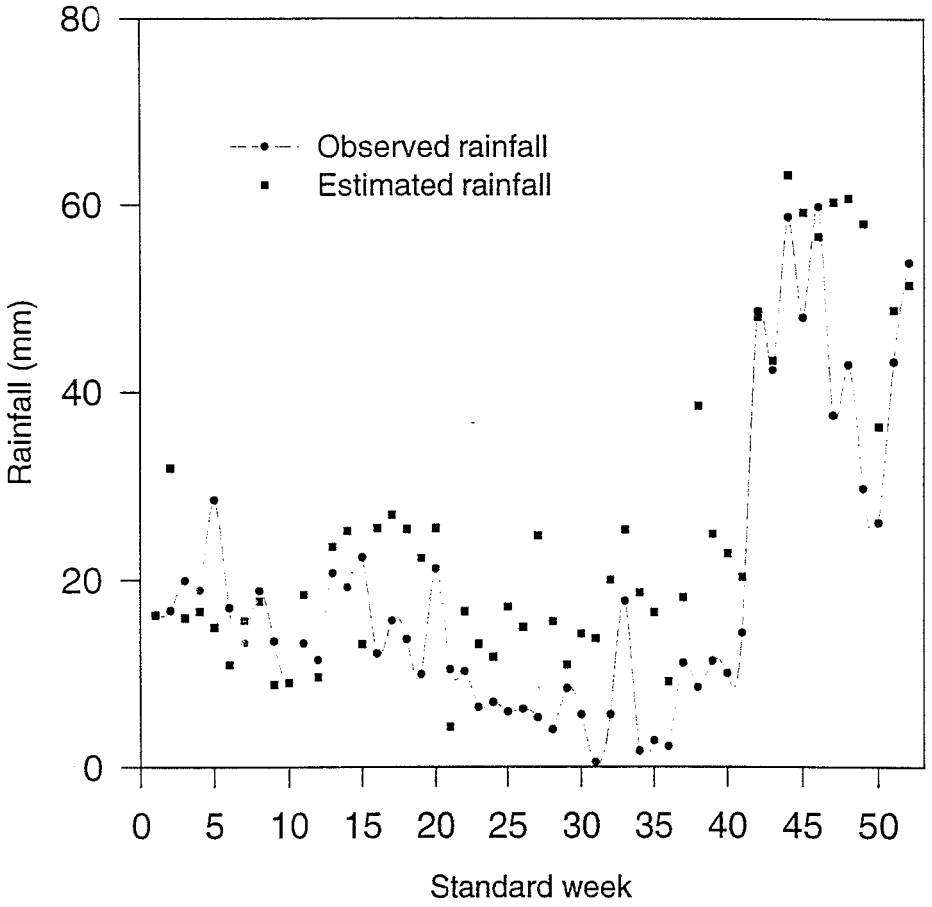

Figure 4 Observed and estimated rainfall at Weerawila in the Dry zone of Sri Lanka.

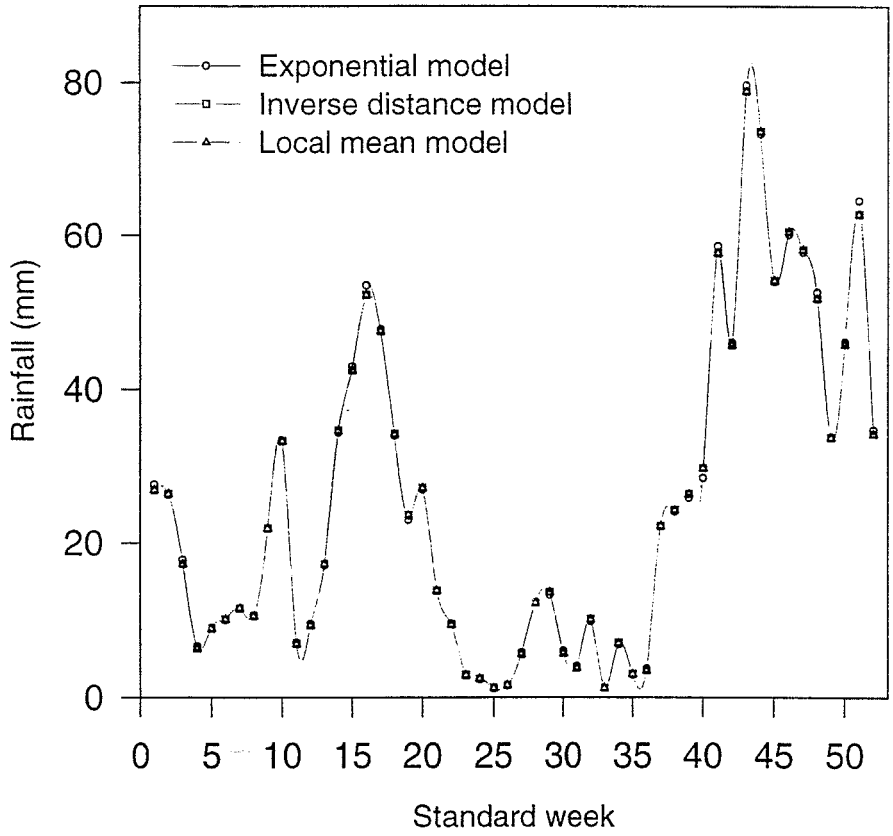

Figure 5 Estimated weekly rainfall from three models at Maradankadawala in the Dry zone of Sri Lanka.

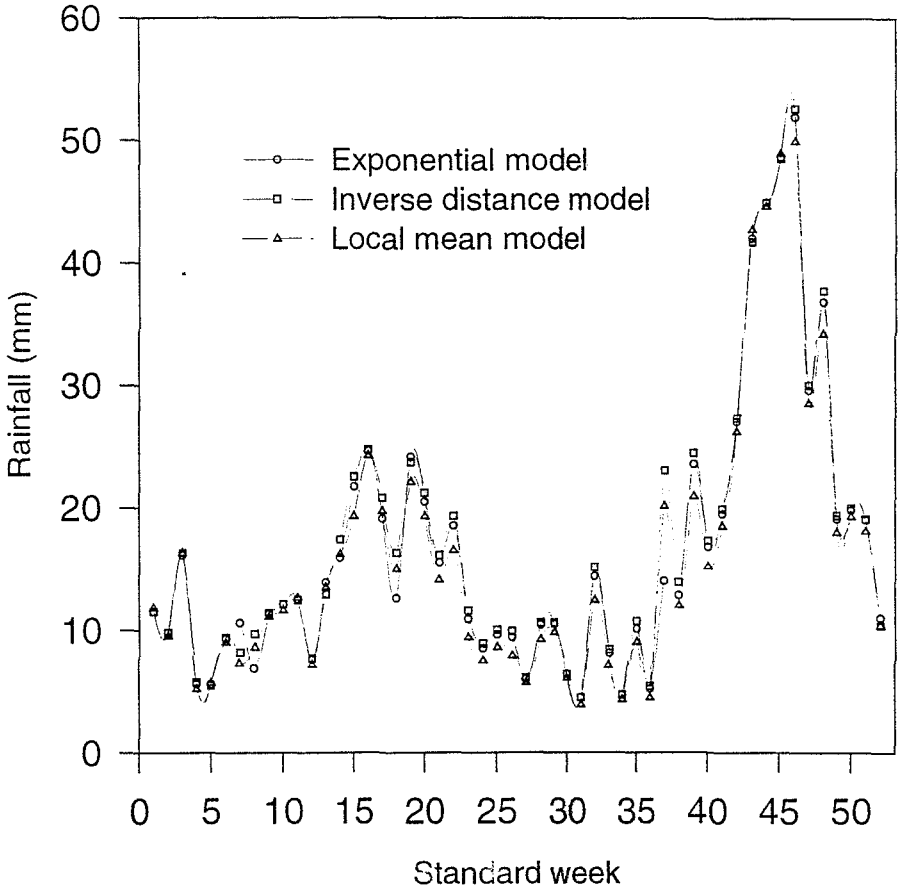

Figure 6 Estimated weekly rainfall from three models at Ambalantota in the Dry zone of Sri Lanka.