Vol. 29: 223–231, 2005 | CLIMATE RESEARCH Clim Res | Published October 17

Spatial interpolation of weekly rainfall depth in the dry zone of Sri Lanka

H. K. W. I. Jayawardene[1], D. U. J. Sonnadara[1,*], D. R. Jayewardene[2]

[1]Department of Physics, and [2]Department of Mathematics, University of Colombo, Colombo 3, Sri Lanka

ABSTRACT: Daily rainfall data recorded at 13 stations were analyzed to study the spatial patterns of rainfall in the dry zone of Sri Lanka. Principal component analysis was utilized to classify the dominant spatial regions. The first 2 eigenvectors accounted for 70.2% (the first eigenvector 54.8% and the second 15.4%) of the total variation, which clearly supports the commonly used major climatic division of Sri Lanka into wet and dry zones. Both the inverse distance weighting method and kriging successfully estimated weekly average rainfall in the North Central dry zone of Sri Lanka. For both methods, high correlation coefficients of 0.88 and 0.91 were observed for the southwest and northeast monsoon periods, respectively, with slightly lower values for intermonsoon periods. For inter-monsoon periods, the inverse distance weighting method produced better results than kriging. This work shows that the strength of the predictions depends on the rainfall seasons as well as the geometrical placement of the stations in the dry zone.

KEY WORDS: Spatial correlation · Kriging · Distance weighted · Principal components analysis

1. INTRODUCTION

Sri Lanka, an island situated between latitudes 5.92 and 9.83° N and between longitudes 79.68 and 81.88° E, is heavily dependent on the rainfall it receives for its agriculture. In most parts of the country rainfall follows a bi-annual pattern leading to a main cropping season called 'Maha' (September to March) and a minor cropping season called 'Yala' (April to August). The wet zone receives rainfall throughout the year which is adequate for year round cultivation. In the dry zone, the rainfall is adequate for crop growth only during the Maha season. Rice is the principal crop, cultivated mainly in the northern and eastern plains of the country in the dry zone. Rainfall variation is one of the principal factors that affect the interannual variability of rice production in Sri Lanka (Yoshino & Suppiah 1983). Also, most legumes (mung bean, cowpea, soybean) can be grown only in the northern and southeastern parts of the island due to the high rainfall (>2000 mm) in the other parts of the country. However high rates of evaporation and the pattern of rainfall pose a significant problem to the rain-fed agriculture in the dry zone, greatly limiting water availability at certain times of the year (especially in the Yala season). Also, the nature of tropical rainfall can lead to marked fluctuations over extremely small distances due to terrain, water bodies and the effect of meteorological conditions (Jackson 1992). Therefore, to make correct decisions regarding cultivating rain-fed rice and other shallow-rooted crops, it is important to know with reasonable accuracy the time of occurrence and depth of rainfall at any given location. Although the country is covered by over 350 rainfall stations, only a small fraction are in the catchment areas. In addition, the stations in the dry zone are usually separated by several kilometers (Punyawardena & Kulasiri 1998) and many suffer from missing values. Moreover, some of the main meteorological stations in the northern and eastern parts have not been functioning since 1990 due to the ethnic conflict. Thus, it is important to explore the spatial relationships that can be used to estimate rainfall at ungauged sites in the dry zone of Sri Lanka.

*Corresponding author. Email: upul@phys.cmb.ac.lk

A number of methods have been proposed for the interpolation of rainfall data. The most simple and frequently used technique uses the weighted average of surrounding values to estimate the rainfall at a given location. The weights are reciprocal to the square of the distances from the unknown location. The literature indicates that a regionalized variable such as precipitation, which is strongly correlated with elevation, would be well estimated by using geostatistical techniques such as kriging. One objective of this work is to compare the validity and strength of both methods in estimating rainfall in the dry zone of Sri Lanka.

Punyawardena & Kulasiri (1998) applied a spatial interpolation model to rainfall in the dry zone of Sri Lanka assuming an exponentially correlated spatial continuity. Their results indicate that a simple model (inverse distance method) rather than complex models based on exponential correlation is adequate for weekly rainfall estimation. They used standard weekly mean estimates (weekly values averaged for 10 yr) to test their model.

Regionalization of daily rainfall in Sri Lanka was studied by Domroes & Ranatunge (1993) using orthogonal factor analysis on rainfall data over a period of 15 yr. They identified 10 homogeneous daily rainfall regions. Suppiah & Yoshino (1984) used Empirical Orthogonal function (EOF) analysis based on monthly rainfall anomalies for the period 1881 to 1980. They found 4 major eigenvectors which described 63 % of the total variance.

In this work, the S-mode principal component method (multiple stations over time) with varimax rotation was applied to identify the dominant rainfall regions of Sri Lanka. The results were used to select a region (approximately 100 km^2) in the North Central dry zone of Sri Lanka where a set of rainfall measuring stations is available to study spatial interpolation. Ordinary kriging with a Gaussian semivariogram model was applied on weekly rainfall to model spatial autocorrelation among selected stations and to estimate weekly rainfall at a location at which the measurements are available. The same analysis was carried out using the inverse distance method and the estimates extracted by both methods were compared with the actual measurements. The accuracy of the estimates was tested using cross-validation techniques.

2. DATA

Rainfall data on 2 time scales were used in this work. First, monthly rainfall data at 15 stations over a 90 yr period (1901 to 1990) were used to classify the rainfall regions. The data were chosen according to different elevation levels as well as to cover most parts of Sri Lanka. An additional requirement was imposed on the selection of rainfall stations on the basis of missing values. The stations were considered as variables (S-15) and monthly rainfall was treated as cases (12×90). Missing values were excluded in pairwise order (i.e. if a case has a missing value for any variable, then the pair of affected values is ignored) within the computation.

Daily rainfall data at 13 stations during 3 time periods, namely 1970–1979, 1980–1987 and 1990–1999, were used for the spatial interpolation model. Spatial interpolation is not effective for the dry zone when the aerial distance is greater than 110 km (Punyawardena & Kulasiri 1998). Thus, the stations were selected by requiring that the aerial distance be less than 115 km from the selected location and within the region identified by the principal component analysis (PCA). To obtain accurate estimates from a geostatistical technique such as kriging the number of stations must be sufficiently large. However, due to the lack of a dense rain gauge network within the study area as well as the large number of missing values and the unreliability of the measurements, the number of stations available for this study was somewhat limited. The percentage of missing values in the weekly rainfall time series within the study periods used in the present work for each of the stations is below 5 % for all except 2 stations in the 1980–1987 period (Medawachchiya, 11.8 %, and Maradankadawala, 12.5 %) and 2 stations in the 1990–1999 period (Vavuniya, 26 %, and Mannar, 41 %). The details of the station locations and their altitudes together with the type of data used are shown in Table 1.

3. METHODOLOGY

3.1. Principal component analysis

S-mode PCA (multiple stations over time) with varimax rotation was applied to identify the rainfall regions of Sri Lanka.

To determine which eigenvectors should be retained in the analysis, the Kaiser (1959) criterion (to accept only those principal components with an eigenvalue greater than 1) and the 'scree test' proposed by Castell (1966) were used.

Factor loadings (i.e. the eigenvectors of each eigenvalue) reflect the correlations between the variables and the extracted factors or principal components. A simple structure can be achieved by rotating factors around the origin until each factor is maximally collinear with a distinct cluster of vectors. An orthogonal rotation scheme, 'varimax rotation' was used in this study. Varimax rotation uses the iterative maximization of the variances of the factor loadings.

Table 1. Data used in the principal component analysis and the spatial interpolation method

Station	Latitude (°N)	Longitude (°E)	Altitude (m)	Daily rainfall	Monthly rainfall
Anuradhapura	8.35	80.38	93	1970–1999	1901–1990
Badulla	6.98	81.03	670		1901–1990
Batticaloa	7.72	81.70	3		1901–1990
Colombo	6.90	79.87	7		1901–1990
Diyatalawa	6.82	80.97	1248		1901–1990
Elyapaththuwa	8.40	80.32	16	1980–1987	
Galle	6.03	80.22	13		1901–1990
Hambantota	6.12	81.12	16		1901–1990
Jaffna	9.65	80.02	4	1980–1987	1901–1990
Kandy	7.33	80.63	477		1901–1990
Kurunegala	7.47	80.37	116		1901–1990
Maha Illuppallama	8.12	80.47	138	1970–1999	
Mannar	8.97	79.90	4	1970–1999	1901–1990
Maradankadawala	8.13	80.57	137	1970–1999	
Medawachchiya	8.55	80.48	95	1980–1987	
Mihintale	8.35	80.52	16	1980–1987	
Mullaitivu	9.27	80.80	13	1980–1987	
Nachchaduwa	8.25	80.47	16	1970–1999	
Nochchiyagama	8.27	80.20	16	1980–1987	
Nuwara Eliya	6.97	80.77	1895		1901–1990
Puttalam	8.03	79.83	2	1970–1999	1901–1990
Rathnapura	6.68	80.40	34		1901–1990
Trincomalee	8.58	81.25	79	1970–1999	1901–1990
Vavuniya	8.75	80.50	106	1970–1999	

For PCA, the calculation of eigenvectors can be carried out by using either the covariance matrix or the correlation matrix of the data set. The correlation matrix method is commonly used when different variables have different variances. Even with the same seasonal period and with same atmospheric controls, monthly rainfall of selected stations could vary due to elevation effects. Therefore the correlation matrix method was used in this study to calculate eigenvalues (Comrie & Glenn 1998).

The ultimate goal in PCA is the identification of the underlying structure of the spatial distribution of rainfall. Factor loadings and principal components were plotted on a contour plot to identify the spatial pattern of each component. A commonly used rule specifies that only variables with loadings of 0.40 or higher on a factor or component should be considered (Raven 1994). We used a much higher level (0.60) contour line to identify rainfall regions according to the extracted principal components.

3.2. Spatial interpolation

The analysis was carried out for weekly averages of rainfall data in 3 separate periods within the years 1970 to 1999. The data contained a fair number of missing values. Therefore prior to the analysis, missing values were removed from the daily time series records. A week which contained at least one missing day was taken as a missing week. These missing weeks were removed when computing weekly average rainfall for stations.

Two spatial interpolation methods, namely kriging and the inverse distance weighting interpolation method, were performed to estimate rainfall values and the results were compared with the measured weekly average rainfall values.

3.2.1. Kriging

Kriging is a geostatistical estimation technique for spatial interpolation which uses a linear combination of surrounding observed values to make predictions. To make such predictions, one must estimate the weights for each surrounding observation. Kriging allows us to derive weights by minimizing the error variance and systematically setting the mean of the prediction errors to zero. Ordinary kriging was used assuming the absence of trend in observations over the considered area.

In kriging, the spatial correlation or the spatial dependence structure is quantified by a 'semivariogram', which provides a measure of variance as a function of distance between data points. The experimental semivariance $\gamma(h)$ is calculated from the observed data using the equation:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{x_i - x_{i+h}\}^2 \qquad (1)$$

Here $N(h)$ is the number of pairs of observed points separated by distance h or lag. x_i and x_{i+h} refer to observed rainfall values at locations around the station to be estimated.

A semivariogram is the graphical display of $\gamma(h)$ versus h. Usually, $\gamma(h)$ will increase with h and after a certain distance, called the 'range' (L), $\gamma(h)$ will level out at a value called the 'sill' (C_1). At a distance beyond L the deviations in observations are no longer correlated.

The experimental semivariogram was modelled by fitting a theoretical function to ensure that the solution is unbiased and has minimum variance. Since the rainfall data display high spatial correlation between nearby points (i.e. an S-shaped behavior at short lags in the semivariogram), a Gaussian function was chosen to model the experimental semivariogram:

$$\gamma(h) = C_0 \left\{ 1 - \exp\left[-\left(\frac{h}{L}\right)^2 \right] \right\} + \gamma(0) \qquad (2)$$

Here h is such that $0 \le h \le L$ and C_0 is the difference between the sill C_1 and 'nugget' $\gamma(0)$. The value x_0 at the estimated location (i.e. unobserved location) can be calculated by using a weighted average of the observed rainfall values x_i at locations around it. Optimal weights are those that produce minimum variance and they can be obtained by solving a set of simultaneous equations. A spatial interpolation program, Easy Kriging 2.1 (EasyKrig2.1 2000), was used to model the semivariance and kriging.

Each time step (a week) in rainfall data represents a separate set of data for kriging. In this work, semivariogram parameters were estimated only for 4 cases using mean seasonal values. Initially, the parameters were estimated by inspection of the shape of the semivariogram. Then, the cross-validation criteria were applied to ensure the correct range that produces a minimum residual error for each parameter. The best sets of parameters for L, $\gamma(0)$ and C_1 were obtained by performing a least square fit. Data records at the estimation point were not used in the estimation of parameters.

3.2.2. Inverse distance weighted interpolation (IDW)

In inverse distance weighted interpolation the values at ungauged locations are determined by the weighted average of the values at observed locations. In this method weights w_i are calculated depending on the radial distance between the observed data points and the estimated point:

$$x_0 = \sum_{i=1}^{N} x_i w_i \left(\sum_{i=1}^{N} w_i \right)^{-1} \qquad (3)$$

where $w_i = 1/d_i^{\,n}$; x_0 is the value at the location to be estimated; x_i is the observed values at locations i around it; N is the number of locations considered to estimate the value at the ungauged station; and d_i's are radial distances from the ungauged station to the chosen stations around it. Here N was chosen by comparing the correlation between the observed and the estimated values for standard weeks and on the basis of minimum mean absolute percentage error (MAPE).

To estimate the accuracy of the model, 3 forms of cross-validation were used, namely, root mean squared error (RMSE), mean relative error (MRE) and correlation between the estimated and the observed data values.

4. RESULTS AND DISCUSSION

4.1. Principal component analysis

Table 2 shows the eigenvalues and their explanatory capability for the monthly rainfall data observed at 15 rainfall stations. The eigenvalue for a given component measures the variance in all the variables which are accounted for by that component.

The first principal component explains the maximum variance from all eigenvalues, and it accounts for approximately 54.8% of the total rainfall variability. The second component accounts for about 15.4% of the variation in the data set and the percentages drop off gradually for the rest of the principal components. Cumulatively, components 1 and 2 together account for 70.2% of the variation in the data set. According to the Kaiser Criterion, the first 2 components alone are adequate, because the subsequent eigenvalues are all less than 1. The Scree test also recommends using only the first 2 factors, because the subsequent eigenvalues do not decrease rapidly. Thus, a 2-component PCA model was chosen for the study.

Factor loadings represent the correlation between input variables and the extracted principal components. To achieve a simple structure, factor loadings were rotated by the varimax rotation criteria. The rotated factor loadings are summarized in Table 3. Here, the first principal component is denoted by PC1 and the second by PC2.

According to Table 3, the first component has large positive loadings on Trincomalee, Batticaloa, Mannar, Badulla, Jaffna, Anuradhapura, Diyatalawa and Puttalam. The second component has large positive load-

Table 2. Total variance explained by eigenvalues

Component	Initial eigenvalues		
	Eigenvalue	% Variance	Cumulative %
1	8.217	54.780	54.780
2	2.311	15.404	70.185
3	0.916	6.104	76.288
4	0.640	4.265	80.554
5	0.569	3.793	84.347
6	0.438	2.918	87.265
7	0.295	1.969	89.234
8	0.290	1.933	91.168
9	0.269	1.796	92.963
10	0.228	1.519	94.482
11	0.215	1.432	95.914
12	0.174	1.163	97.077
13	0.158	1.054	98.131
14	0.145	0.970	99.100
15	0.135	0.900	100.00

Table 3. Rotated factor loadings

Station	Principal component	
	PC1	PC2
Trincomalee	0.869	0.097
Batticaloa	0.868	–0.045
Mannar	0.850	0.240
Badulla	0.821	0.222
Jaffna	0.814	0.229
Anuradhapura	0.810	0.350
Diyatalawa	0.741	0.429
Puttalam	0.684	0.465
Hambantota	0.572	0.461
Kandy	0.453	0.716
Kurunegala	0.442	0.722
Nuwara Eliya	0.233	0.700
Colombo	0.229	0.776
Galle	0.162	0.787
Ratnapura	–0.096	0.875

ings on Ratnapura, Galle, Colombo, Nuwara Eliya, Kurunegala and Kandy.

Fig. 1 shows the rotated loading patterns for selected principal components, with PC1 representing the largest variance in the rotated solution and PC2 the next largest. For PC1, the factor loadings are comparatively high (>0.60) in the northern and eastern regions while the southwest quadrant has lower factor loadings. On the other hand, for PC2 the southwestern region has high (>0.60) loadings and the rest have low loadings. Therefore we can clearly discern 2 regions according to loading patterns separated by the contour line at 0.60. The 2 components classify the commonly used climatic regions in Sri Lanka, the wet and dry zones, where PC1 corresponds to the dry zone and PC2 to the wet. If we compare the percentage of seasonal rainfall distribution against the annual totals in Sri Lanka (Perera 2003), the relationship between the 2 regions and the 2 monsoons can be clearly identified. On the same figure, the commonly used demarcation between the wet and dry zones is also shown (dashed line).

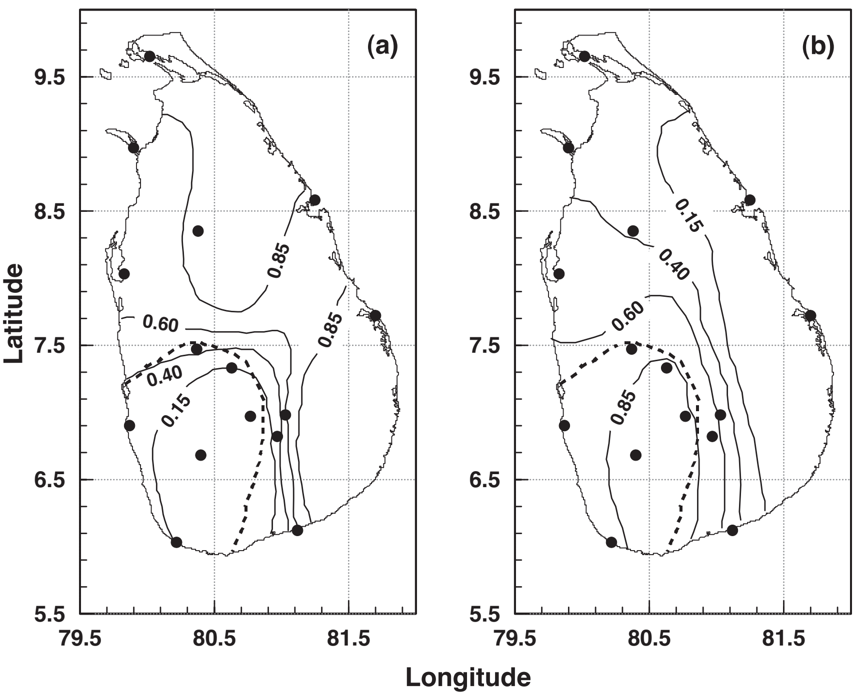

Fig. 1. (a) Rotated loadings for the first principal component, PC1. (b) Rotated loadings for the PC2. Dashed line indicates the commonly used demarcation between the wet and dry zones

To test the spatial correlation, Anuradhapura was selected as the test station. Using the results of the PCA, 12 stations were selected in the area bounded by the contour line 0.60 that are less than 115 km aerial distance from Anuradhapura. The aerial distance of the selected 12 stations together with the percentage of missing values estimated for each of the 3 study periods is shown in Table 4. The spatial distribution of the stations is shown in Fig. 2.

4.2. Spatial interpolation

4.2.1. Kriging

The procedure of kriging was applied in 2 steps. The first step was modeling the semivariogram or covariance and the second was spatial interpolation. To model the experimental

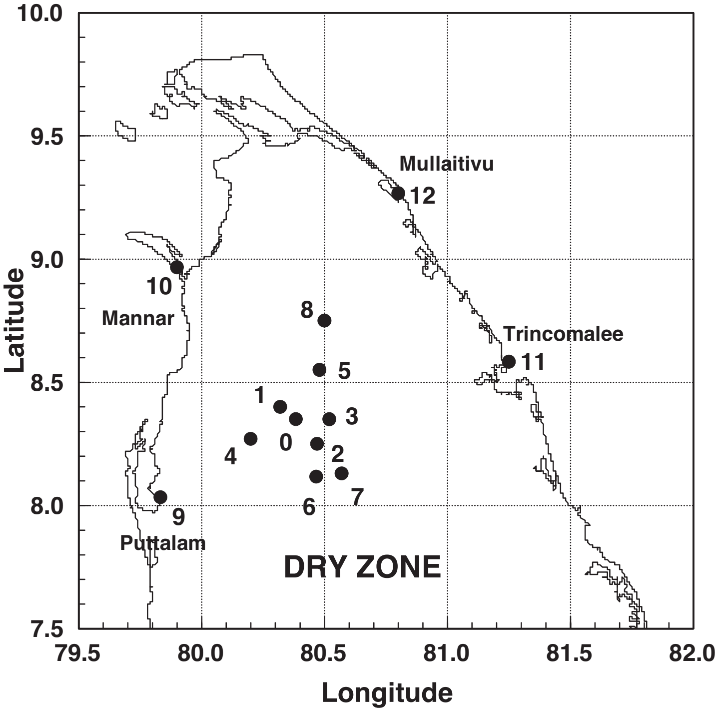

Fig. 2. Stations selected for the spatial interpolation analysis. Stn 0 indicates the test station; Anuradhapura. Station identification numbers together with the corresponding station names are given in Table 4

semivariogram one must estimate the parameters, nugget [γ(0)], sill (C_1), range (L) and resolution (lag or distance h) of the Gaussian function. Lag h was estimated relative to the full length scale. The parameters were estimated for each season separately using standard weeks. By using the commonly considered 4 seasons in Sri Lanka, we averaged Weeks 1 to 10 and Weeks 48 to 52 for the northeast monsoon (NEM), Weeks 11 to 20 for the first intermonsoon (IM1), Weeks 21 to 37 for the southwest monsoon (SWM) and Weeks 38 to 47 for the second intermonsoon (IM2). Therefore for the 4 seasons 4 Gaussian semivariogram models were fitted. The estimated parameters of the modelled Gaussian semivariogram for each season together with residual errors are tabulated in Table 5.

Table 4. Stations selected for the spatial interpolation analysis. Stn 0 indicates the test station; Anuradhapura for the 1980–1987 period. Estimated percentages of missing values are also shown for all 3 study periods

Station		Distance from Anuradhapura (km)	Missing values 1970–1979 %	1980–1987 %	1990–1999 %
0	Anuradhapura	–	0.4	0.0	0.2
1	Elyapaththuwa	9.36	–	0.2	0.0
2	Nachchaduwa	14.56	3.8	4.1	–
3	Mihintale	14.90	–	1.2	–
4	Nochchiyagama	22.55	–	1.2	–
5	Medawachchiya	25.04	–	11.8	–
6	Maha Illuppallama	27.81	0.6	0.0	0.6
7	Maradankadawala	32.34	1.9	12.5	3.7
8	Vavuniya	46.68	1.2	1.9	26.0
9	Puttalam	71.10	0.2	0.0	0.4
10	Mannar	88.10	0.2	2.4	41.0
11	Trincomalee	100.55	0.2	3.1	1.7
12	Mullaitivu	113.51	–	4.1	–

Table 5. Parameters of fitted Gaussian semivariogram models for each season

Parameters	Gaussian model IM1	SWM	IM2	NEM
Nugget [γ(0)]	0.0000	0.652	0.0352	0.073
Sill (C_1)	1.2000	1.500	1.0600	1.460
Range (L)	0.0834	0.362	0.8280	0.380
Resolution (lag h)	0.0650	0.073	0.0990	0.045
Mean residual error	–0.4694	0.6334	–0.1617	–0.553

Kriging interpolation was performed for each week between 1980 and 1987 using the modelled semivariogram parameters. The kriging results for Week 34 in 1984 are shown as a contour map in Fig. 3a and the corresponding kriging variance is shown in Fig. 3b. A grid size of 20 × 20 points was used in evaluating both maps.

The variance map indicates the relative reliability in the different parts of the contour map for kriging estimates. The magnitude of the kriging variance for a specific location depends on the semivariogram model and its parameters, the data locations used in the estimation, and the relative position of the location to be estimated.

4.2.2. Inverse distance weighted interpolation (IDW)

The first step of the inverse distance weighted method was selecting the optimum power parameter n in Eq. (3). We chose n on the basis of mean absolute percentage error (MAPE) and by comparing the correlation coefficient (r) between observed weekly average rainfall values at Anuradhapura and estimated values for the period 1980 to 1987. The results are shown in Table 6.

The minimum absolute percentage error was observed when the power parameter $n = 2$. When $n = 1$ the correlation coefficient r is slightly higher than the case of $n = 2$. Thus, both values were used in this work.

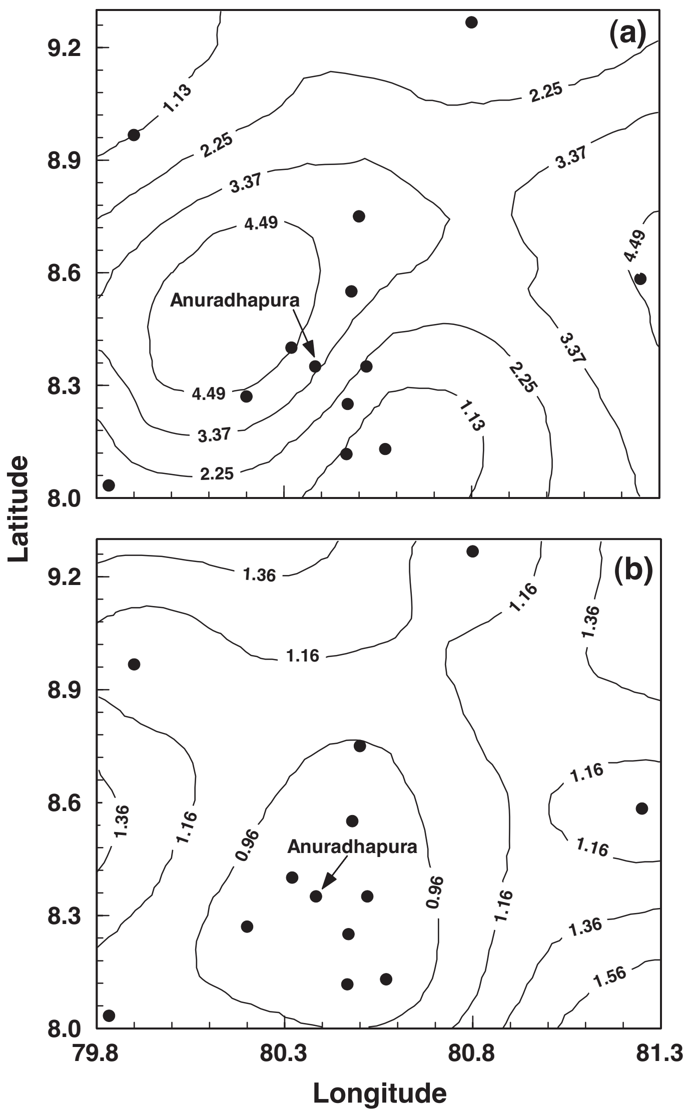

Fig. 3. (a) Kriging map for Week 34 in 1984. (•) Stations used in the analysis. Values shown are rainfall amounts in mm. (b) Kriging variance map for Week 34 in 1984

4.2.3. Accuracy of prediction

Table 7 presents 2 estimates of prediction accuracies. Best predictors have minimum root mean squared error (RMSE) values. The RMSE is slightly larger for kriging than for IDW with power parameter 2 (IDW-n2) while it is the minimum for IDW with power parameter 1 (IDW-n1). On the other hand the mean relative error (MRE) of IDW-n1 is greater than both kriging and IDW-n2. The MRE is the relative variant of bias measure which indicates the difference between the mean values of an interpolated set and a true dataset. As large rainfall values generally show larger errors than small values, the MRE measure partially compensates

Table 6. Mean absolute percentage (MAPE) error of inverse distance weighted interpolation (IDW) for power parameter n = 1 to 5

Power parameter (n)	MAPE for 1980–1987	Correlation coefficient (r)
1	45.6	0.91
2	34.2	0.89
3	37.4	0.87
4	45.0	0.85
5	48.8	0.84

Table 7. Prediction accuracy measures for kriging, IDW with power parameter 1 (IDW-n1) and IDW with power parameter 2 (IDW-n2)

Method	RMSE	MRE
Kriging	3.1	113.2
IDW-n1	2.7	123.1
IDW-n2	3.0	112.9

this dependency and downweights large values with respect to small values. On the other hand the RMSE measure is more strongly influenced by extreme differences than the MRE. Therefore according to Table 7, IDW-n2 predicts much better than IDW-n1.

A visual understanding of accuracy associated with the 2 spatial interpolation methods can be obtained through scatter plots of actual and estimated values (Fig. 4). The inverse distance method shows a slightly better relationship than kriging with less scattered points.

The correlation of observed and estimated values for the 1980–1987 period indicates IDW-n1 has slightly better correlation with observed values. Kriging and IDW-n2 estimates are slightly less correlated with observed values. However, all estimates correlate sufficiently with each other reflecting the equality of estimates in both methods.

The variation of the prediction accuracy with season is shown in Table 8. During the northeast monsoon the prediction performance is high. It is less in Intermonsoon 1 which receives convective rain. In the convection period there could be differences between rainfall amounts even at stations separated by only a few kilometers due to elevation effects. Even very small hills significantly enhance rainfall due to seeder-feeder effects (Schwarb 2000). As given in Table 1 the elevations of the stations vary from 2 to 138 m above sea level. However for Intermonsoon 2 the correlation coefficients were higher. The reason may be the influence of cyclones in that period. Cyclones bring heavy rains to a wide area of the lowlands of northern and

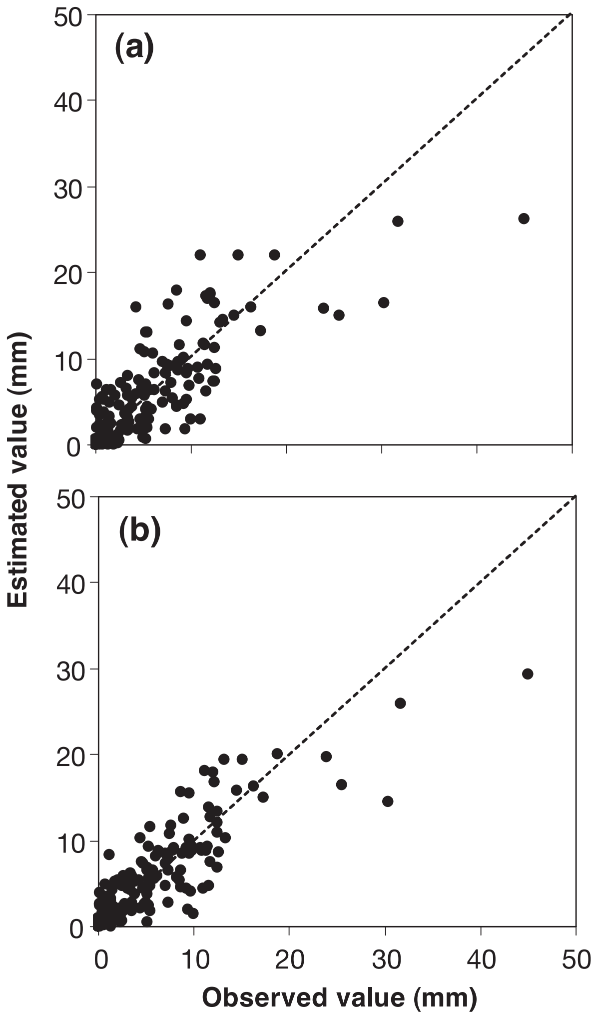

Fig. 4. Scatter plots of observed values versus (a) kriging and (b) IDW-n1 estimated values

Table 8. Correlation coefficient of observed and estimated values for 4 seasons during the 1980–1987 study period

Season	IDW-n1	Kriging
Northeast monsoon	0.91	0.91
Inter-monsoon 1	0.81	0.72
Southwest monsoon	0.88	0.88
Inter-monsoon 2	0.87	0.80

eastern Sri Lanka and the spatial correlation of rainfall may be high in that period. On the other hand during the northeast monsoon period the northern area receives heavy rainfall and the spatial interpolation methods work well.

To check the stability of each method, the same 2 interpolation methods were applied to 2 more stations in the dry zone, Vavuniya and Maha Illuppallama, for 2 different time periods, 1970–1979 and 1990–1999.

Table 9. Results of spatial interpolation at Vavuniya (Group A) and Maha Illuppallama (Group B)

	Vavuniya		Maha Illuppallama	
	RMSE	r	RMSE	r
1970–1979				
Kriging	3.8	0.81	2.7	0.88
IDW-n1	3.9	0.80	2.7	0.88
IDW-n2	4.0	0.79	2.7	0.89
1990–1999				
Kriging	3.9	0.79	3.2	0.88
IDW-n1	3.9	0.75	3.4	0.87
IDW-n2	4.0	0.74	3.3	0.87

The first study assumed Vavuniya as the unobserved station and 7 other stations (Group A) were used in the interpolation process. In the second study the same 8 stations (Group B) were used considering Maha Illuppallama as the unobserved station. The distances from the unobserved station to each of the stations in both groups are <115 km. Inter-station distances for Group A are very large and are all >46 km. This means no close stations were considered for spatial interpolation at Vavuniya. Inter-station distances in Group B range from 11 to 115 km.

It was assumed that the fitted semivariogram model of the previous study applies to all the stations in the considered area in the dry zone. Therefore the same parameters were applied in kriging interpolation for Groups A and B. The inverse distance method was applied for power parameter $n = 1$ and 2. The results are shown in Table 9.

The estimated values for Maha Illuppallama are correlated to the observed values better than for Vavuniya. The reason could be the spatial arrangement of the stations (see Fig. 2). In IDW, the weighting factor is proportional to the inverse of the distance between the observed and estimated point and in other words is based on the assumption that nearby values contribute more to the interpolated values than distant observations. Similarly kriging uses weights which assign more influence to the nearest data points in the interpolation. Thus, the weights and consequently the interpolated values change according to the spatial arrangement of the observational data.

Correlation coefficients are maximum and RMSE are minimum for kriging for both periods 1970–1979 and 1990–1999 in Group A. In Group B, the correlation coefficient is maximum and RMSE is minimum for IDW-n2 for the period 1970–1979 and kriging for the period 1990–1999. When the same analysis was carried out separately for the 4 seasons, it was revealed that the strength of the correlation depends on the rainfall season, with the best result being for the northeast monsoon season. This confirms the result obtained for the 1980–1987 study period. As expected, Group B performed better in all seasons compared with Group A, which has large inter-station distances. Although there is a higher percentage of missing values in the 1990–1999 period compared to 1970–1979, the results obtained for Groups A and B showed no significant adverse effects due to missing values.

5. CONCLUSIONS

The PCA based on monthly rainfall records during the period 1901 to 1990 was successful in identifying rainfall regions of Sri Lanka. The study clearly identified the 2 dominant rainfall regions, the wet zone and the dry zone, which are related to the 2 main monsoon seasons in the northeast and southwest. It was not necessary to extract higher resolution than this due to the focus of this study. The results of the principal component analysis were used to select a region in the dry zone of Sri Lanka for the spatial interpolation study.

Considering the overall outcome of the spatial interpolation results at Anuradhapura, the inverse distance weighted method (IDW) with power parameter 1 was most representative of the original data. Higher correlations between observed and estimated values for 4 seasons favored IDW. IDW with power 1 shows the minimum RMSE value and IDW with power 2 was minimum for MRE. Both kriging and IDW use a linear combination of weights at known points to estimate the value at an unknown point. The spatial arrangement of the observed stations does not affect the weights in the IDW method. Kriging uses a semivariogram, a measure of spatial correlation between 2 points, so the weights change according to the spatial arrangement of the observations. Therefore the spatial arrangement and the number of stations affect the performance of kriging. The predictions showed some difference in the performance of the 2 methods for various seasons. The reason could be the effect of corresponding atmospheric processes on rainfall values in each season. During the first inter-monsoon period, cumulus clouds associated with mesoscale convection produce intense rain, sometimes leading to marked contrasts over few kilometers and weakening the spatial correlation. In the second inter-monsoon, unlike the convectional weather, heavy rainfall is produced by the depressions and cyclonic systems which develop in the Bay of Bengal. On the other hand, during the monsoon periods, particularly during the northeast monsoon, prediction performance is high due to widespread rainfall in the northern area of the country.

The spatial interpolation at Vavuniya and Maha Illuppallama showed that kriging and the inverse distance weighting method perform similarly for spatial interpolation of rainfall data. Since the majority of prediction errors of kriging occur due to improper parameter estimation, tuning or refining the parameters can increase the accuracy of kriging. Although the inverse distance method is simpler and estimates reasonable values, kriging is useful since it allows quantification of the quality of predictions via the kriging variance. It can also be used when preparing irregularly scattered rainfall data to construct a contour surface which is a 2-dimensional representation of a 3-dimensional surface.

The analysis presented here illustrates the performance of spatial interpolation methods in the North Central part of Sri Lanka which receive less rainfall than the wet and mountainous regions. One can apply both kriging and inverse distance weighting to such regions combined with additional parameters such as elevation and humidity to improve the accuracy of the predictions.

Acknowledgements. Financial assistance by IPPS, Uppsala University, Sweden (research grant number SRI:01/1), and the National Science Foundation, Sri Lanka (grant number RG/2000/P/01) are acknowledged.

LITERATURE CITED

Castell RB (1966) The scree test for the number of factors. Multivar Behav Res 1:245–276

Comrie AC, Glenn EC (1998) Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. Clim Res 10: 201–215

Domroes M, Ranatunge E (1993) A statistical approach towards a regionalization of daily rainfall in Sri Lanka. Int J Climatol 13:741–75

EasyKrig2.1 (2000) The GLOBEC kriging software package. ftp://globec.whoi.edu/pub/software/kriging/easy_krig/V 2.1/

Jackson IJ (1992) Climate, water and agriculture in the tropics. Longman, London

Kaiser HF (1959) The varimax criterion for analytic rotation in factor analysis. Educ Psychol Meas 19:413–420

Perera HKWI (2003) Exploring the spatial and temporal variations of rainfall in Sri Lanka. M Phil thesis, Department of Physics, University of Colombo

Punyawardena BVR, Kulasiri D (1998) Spatial interpolation of rainfall in the dry zone of Sri Lanka. J Nat Sci Counc Sri Lanka 26(3):247–262

Raven MR (1994) The application of exploratory factor analysis in agricultural education research. J Agric Educ 35(4):9–14

Schwarb M (2000) The Alpine precipitation climate: elevation of a high resolution analysis scheme using comprehensive rain-gauge data. Dissertation, ETH 13911, http://www.geo.umnw.ethz.ch/staff/homepages/sharb/

Suppiah R, Yoshino MM (1984) Rainfall variations of Sri Lanka Part 1: Spatial and temporal patterns. Arch Meteorol Geophys Biocl Ser B34:329–340

Yoshino MM, Suppiah R (1983) Climate and paddy production: a study on selective districts in Sri Lanka. Climatological Notes, Tsukuba, Japan, 33:33–50

Editorial responsibility: Otto Kinne, Oldendorf/Luhe, Germany

Submitted: November 11, 2004; Accepted: August 23, 2005
Proofs received from author(s): September 5, 2005