Geostatistical Methods for Prediction of Spatial Variability of Rainfall in a Mountainous Region

A. Sarangi, C. A. Cox, C.A. Madramootoo

Abstract. *Reliable estimation of rainfall distribution in mountainous regions poses a great challenge not only due to highly undulating surface terrain and complex relationships between land elevation and precipitation, but also due to non-availability of abundant rainfall measurement points. Prediction of rainfall variability over mountainous islands is a logical step towards meaningful land use planning and water resources zoning. In this context, geostatistical techniques were developed for mapping the rainfall variability over the island of St. Lucia in the Caribbean, using the elevation information extracted from a Digital Elevation Model (DEM) and long-term mean monthly rainfall (MMR) data of 40 raingauge stations spread over 616 km*^{2}*. The ordinary co-kriging (OCK) and collocated co-kriging (CCK) methods of interpolation were applied for the standardized rainfall depths associated with elevation, as the primary variate, and the surface elevation values as the secondary variate. The best semivariogram model algorithm generated, using either of the above co-kriging (CK) methods, was used to predict standardized values for the elevation points extracted from the DEM for which the rainfall depths were not known. The predicted values were further destandardized to generate the rainfall depth at the unmeasured locations. Ordinary kriging (OK) was then performed for the destandardized and observed rainfall depths to generate the prediction map of MMR over the entire island. These sequential steps were repeated for the MMR data of all twelve months to generate rainfall prediction maps over the island. The spherical semivariogram model fit well (*$0.84 < R^2 < 0.98$*) for both the OCK and OK methods. The cross-validation error statistics of OCK presented in terms of coefficient of determination (*R^2*), kriged root mean square error (KRMSE), and kriged average error (KAE) were within the acceptable limits (KAE close to zero, *R^2* close to one, and KRMSE from 0.55 to 1.45 for 40 raingauge locations) for most of the months. The exploratory data analysis, variogram model fitting, and generation of MMR prediction map through kriging were accomplished through use of ArcGIS and GS+ software.*

Keywords. *ArcGIS, Collocated ordinary co-kriging, Geostatistical analysis, GS+, Ordinary co-kriging, Ordinary kriging, Rainfall interpolation, Spatial variability, St. Lucia.*

The spatial variability of precipitation is highly influenced by meteorological conditions and land morphological factors. Mountainous regions are characterized by complex precipitation patterns due to undulating terrain. Moreover, sparse raingauge networks add to the complexity of accurate estimation of rainfall data for ungauged locations. These features make it difficult to adopt a model to describe the spatial variability of rainfall in mountainous regions. Phenomena like the enhancement of precipitation on windward sides and the rain-shadow effect are typical for mountainous regions with varied topography (Christel and Reed, 1999). In regions with strong orographic rainfall effects, the proper estimation of rainfall may contribute significantly to development of rainfed agricultural production systems. The spatial and temporal variability of precipitation affects both soil infiltrability and the production and successive propagation of surface runoff. In addition, determination of long-term spatial variation in rainfall is important for crop agro-ecological zoning and demarcation of potential surface and ground water availability regions.

This study was undertaken to model the spatial variability of rainfall over the eastern Caribbean island of St. Lucia (fig. 1). The Caribbean Archipelago stretches from southern North America to northern South America. Most of the islands are mountainous due to volcanic genesis. The Greater Antilles, which include Jamaica, Hispaniola, and Puerto Rico, are dominated by mountainous ranges at elevations between 1000 and 3000 m. The Lesser Antilles is comprised of low-lying coralline islands in the northern Leeward group, while the islands of the Windward group toward the south are volcanic, with elevations rising above 1000 m. Agriculture is an integral part of the economy of most of the islands. In the absence of any established irrigation network, agriculture is solely dependent on rainwater. Moreover, in the relatively flatter valley regions, the spatial variability of rainfall being less, rainwater is a dependable water resource for practicing agriculture. However, due to high spatial variability of rainfall in the mountainous region, the availability of rainfall over a larger space becomes uncertain, and agricultural

Article was submitted for review in November 2003; approved for publication by the Soil & Water Division of ASAE in April 2005.

The authors are **Arjamadutta Sarangi, ASAE Student Member,** Post-Doctoral Fellow, Department of Bioresource Engineering, Macdonald Campus, McGill University, Ste. Anne-de-Bellevue, Quebec, Canada; **Christopher A. Cox,** Chief Agricultural Planning Officer, Ministry of Agriculture, Forestry and Fisheries, Castries, St. Lucia; and **Chandra A. Madramootoo, ASAE Member Engineer,** James McGill Professor, Director of the Brace Centre for Water Resources Management, Macdonald Campus, McGill University, Ste. Anne-de-Bellevue, Quebec, Canada. **Corresponding author:** Arjamadutta Sarangi, Department of Bioresource Engineering, Macdonald Campus, McGill University, 21,111 Lakeshore Road, Ste. Anne-de-Bellevue, Quebec, Canada, H9X 3V9; phone: 514-398-7773; fax: 514-398-8387; e-mail: arjamadutta.sarangi@elf.mcgill.ca.

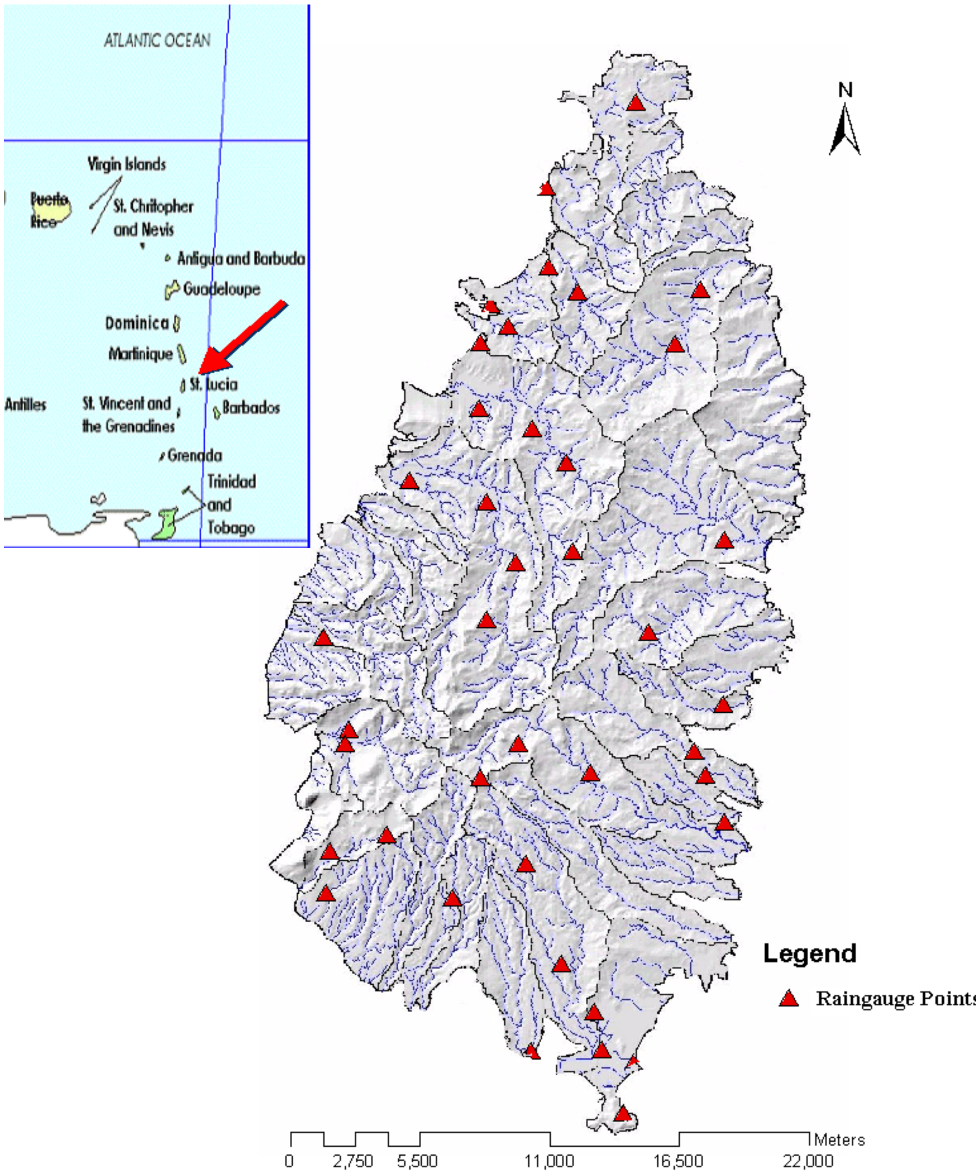

Figure 1. Location map of St. Lucia Island with raingauge locations.

activities are at low level in the mountainous regions. Reliable estimation of spatial and temporal rainfall patterns over the islands and the variability of temperature at higher elevations (FAO, 1996) are important for agricultural planning. Therefore, the present study aims at accurate estimation of spatial rainfall depths over St. Lucia for agro-ecological crop zoning, land use planning, and watershed management.

Background Information

To account for the spatial variability of rainfall, conventional techniques such as Thiessen polygons, the isohyetal method, and inverse distance weighting (IDW) were used for interpolation of rainfall data until the late 1980s (Goovaerts, 1997; Phillips et al., 1992) but were not always reliable in mountainous regions (Lebel et al., 1987; Hevesi et al., 1992). These methods are useful where there is an adequate distribution of sample points over the study area or where the surface terrain is uniform. However, higher raingauge density in steep mountain terrain is often not possible for several reasons: the costs of installation, operation, and maintenance are quite high, and these regions are also inaccessible to human climate observers. Therefore, a reliable method to generate spatial rainfall information for mountainous areas using point rainfall data is required. The Thiessen polygon method (Thiessen, 1911) and the IDW technique do not allow hydrologists to consider the morphological factors that can affect the rainfall depth at a gauging point. The isohyetal method (McCuen, 1998) is designed to overcome this deficiency by using the location and depth information for drawing isohyets. The amount of rainfall at the unsampled location is then estimated by interpolation within the isohyets. The limitation of this technique is that an extensive raingauge network is required to draw isohyets accurately.

In order to overcome these limitations and use the topographical information in interpolation, geostatistical methods are employed for estimation of rainfall variability. Geostatistics, based on the theory of describing the relationship between the spatially random variables (Goovaerts, 1997; Kitiandis, 1997; Goovaerts, 2000; Wackernagel, 2003), is increasingly employed because it utilizes the spatial correlation between neighboring observations to predict attribute values at unsampled locations. Several authors (Bacchi and Kottegoda, 1995; Christel and Reed, 1999; Goovaerts, 1999, 2000; Campling et al., 2001; Drogue et al., 2002) have shown that geostatistics provides better estimates of precipitation than conventional methods. Moreover, the co-kriging (CK) technique is advantageous for map generation using sparsely sampled observations of the primary attribute with more densely sampled secondary attributes. Dirks et al. (1998) found that the results of kriging depended on the sampling density, and for high-resolution networks (e.g., 13 raingauges over a 35 km^2 area) the kriging method did not show significantly greater predictive skill than simpler techniques, such as the IDW method.

Christel and Reed (1999) developed geostatistical methods for mapping extreme rainfall using data from 1003 raingauge stations spread over the mountainous regions of Scotland. Ordinary kriging (OK) and modified residual kriging techniques were used for mapping extreme rainfall due to their simplicity. It was also observed that inclusion of topographical information (i.e., elevation, slope, and aspect) in OK resulted in better prediction accuracy of cross-validation results.

Goovaerts (2000) compared the predictability of three geostatistical methods, kriging with external drift (KED), collocated co-kriging (CCK), and simple kriging with variable means, with the regression (between rainfall and elevation), IDW, Thiessen polygon, and OK methods to predict rainfall variability. Geostatistical techniques were used to incorporate the elevation from a Digital Elevation Model (DEM) to predict the spatial rainfall variability for 36 climatic stations within an area of 5000 km^2 in Portugal. It was observed that the IDW and Thiessen polygon methods gave the largest prediction errors of cross-validation, whereas the KED, CCK, and OK methods yielded minimal prediction errors. Moreover, OK outperformed all other methods with a moderate correlation coefficient value of 0.75 between the rainfall and elevation data.

Drogue et al. (2002) compared the multiple regression techniques of rainfall versus the geomorphologic parameters (using PLUVIA software) with two methods of geostatistics (KED and extended CCK) for mapping the rainfall over mountainous regions of northeastern France. It was inferred from the cross-validation statistics that the extended CCK methods generated better results in terms of mean absolute error and standard deviation. The PLUVIA model was linked with GIS to plot the kriged surface, and statistical software was used to perform the multiple regression calculations.

It was revealed that only one kriging technique, such as OK, KED, CK, and CCK, did not always give the best results in predicting rainfall in mountainous regions (Christel and Reed, 1999). Therefore, in this study, an effort was made to develop a combination of geostatistical techniques of kriging within a GIS environment to predict the spatial variability of rainfall for a mountainous region.

Study Objectives

The present study's goal was to use multivariate geostatistical methods to predict rainfall from elevation information and then use the univariate geostatistical method to interpolate rainfall and map the spatial variability. The specific objectives were: (1) to extract elevation points from the DEM for analysis and develop a mathematical association (standardization) between the elevation and rainfall data for use in CK methods as a primary variate and elevation values as a secondary variate, (2) to attempt different CK methods and select the best method from the analysis of the cross-validation error statistics through cross-semivariogram models, (3) to use the selected CK method for prediction of the standardized rainfall values at unmeasured locations, and (4) to perform kriging for generation of interpolated rainfall maps using the predicted and observed mean monthly rainfall data of twelve months for St. Lucia.

Study Area and Data Collection

St. Lucia is a small volcanic island in the Windward Island group in the Lesser Antilles (fig. 1). The island is 616 km^2 and measures 45 km by 23 km across its widest axes. The highest peaks are in excess of 800 m elevations, with steep gradients, particularly in the central region of the island. The climate is tropical maritime with mean annual temperatures ranging between 26°C and 32°C (Cox, 2002). Based on available historic records, it was observed that the mean annual rainfall varied from 1500 mm around the coastal low lands to 3800 mm in the central interior, which was due to orographic influences. The bulk of the annual rainfall occurs in the period from July through December due to tropical cyclonic weather systems prevailing during the Atlantic hurricane season.

Long-term monthly rainfall records are available for 42 raingauge locations in St. Lucia. Historic records for some stations exceed 100 years (Cox, 2002). Rainfall is collected daily and compiled to generate monthly totals. Most of the rainfall stations are non-recording. These data have been used in various applications to conduct analyses to support crop production systems planning and estimation of hydrologic parameters required to plan infrastructural development related to water extraction for drinking or irrigation. In these applications, data from the nearest stations are typically assumed to represent rainfall regimes within a local geographic region. However, improvements can be made in estimating rainfall depth at unsampled locations by interpolating between the nearest gauges. Data used in the analysis were derived from the U.N. Water Resources Preliminary Report (Migeot and Hadwen, 1986), the WEMP report (HTS, 1997), and the thematic data layers as listed in table 1. Data from 40 rainfall stations within St. Lucia with more than 10 years of record were used in the analyses (Cox, 2002). The IDW method was employed to interpolate the rainfall data and generate spatial variability in monthly rainfall (Cox, 2002). However, the IDW method assumes a weighted average within the unsampled region and therefore failed to provide a reliable prediction of the rainfall in the mountainous and hilly terrain of St Lucia (Cox, 2002). These inaccuracies can impact the hydrologic analysis and water resources planning of the region.

Interpolation Procedures

This section briefly deals with the procedures of data analysis using ArcGIS and GS+ software and the kriging methods of interpolation. The details of the kriging algorithms are presented in Johnston et al. (1996), Kitiandis (1997), Goovaerts (1997), and Wackernagel (2003).

Data Presentation and Analysis within GIS Environment

The delineated base map (i.e., polygon feature class) of St. Lucia and the location of rainfall gauging stations within the island (i.e., point feature class) were generated using ArcGIS as two different coverage feature classes. The point feature class coverage map, representing the rainfall locations, also contained the mean monthly rainfall depths for twelve months as attribute values. The base map and the rainfall point coverage map were overlaid to represent the raingauge locations within St. Lucia. These feature class maps and the DEM were projected to a standard geographic co-ordinate system representing the island of St. Lucia. The DEM of St. Lucia was used for extraction of approximately 400 elevations points, ranging from 3 to 870 m, covering the entire island, for use in kriging analysis. The geostatistical analysis extension module of ArcGIS 8.3 was used for analysis and development of kriged surfaces, and GS+ was used to understand the details of variogram models and statistical fitting parameters.

Kriging Techniques

The objective of kriging is to predict rainfall values at ungauged locations (x_0) within the system domain (D) using information available elsewhere in D ($x_1, x_2, ..., x_n$). This can be carried out by expressing $Z(x_0)$ {where $Z(x)$: $x \in D$} as a linear combination of the data $Z(x_1)$, $Z(x_2)$, ..., $Z(x_n)$, such that:

$$\ddot{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_0) \tag{1}$$

where λ_i is the kriging weight of the parameter value at $Z(x_0)$ for n nearby sample points to be used in estimation. The optimal weight (λ_i) is calculated such that the estimation of $\ddot{Z}(x_0)$ by $Z(x_0)$ is unbiased and the sum of squares of error is minimized. Of the different kriging techniques, the OK and CCK methods were used in the present study because of their simplicity and prediction accuracy in comparison to other kriging methods (Isaaks and Srivastava, 1989). Moreover, it is observed (e.g., Webster and Oliver, 2001; Johnston et al., 1996) that the OK of a single variable is most robust and is frequently used to account for data fluctuations and consideration of a global trend over the study region. Again, due to the low density of raingauges, i.e., 40 measured point locations in 5000 km^2, the omnidirectional variogram was used in the present analysis, which assumes the spatial variability to be identical in the X, Y, and Z directions.

Ordinary Kriging (OK)

Ordinary kriging is based on two assumptions. First, the mean of the process is assumed constant and is invariant within the spatial domain. This is expressed as:

$$E[Z(x+h) - Z(x)] = 0 \tag{2}$$

where E is the expectation, and $x \in D$ and $x + h \in D$, with h being the distance between two points.

Second, the variance of the difference between two values is assumed to depend only on the distance h between the two points, and not on the location x. The variance is given by:

$$\mathrm{var}[Z(x+h) - Z(x)] = 2\gamma(h) \tag{3}$$

where the function $\gamma(h)$ is the semivariogram.

Based on these assumptions, the kriging equation is given as (Deutsch and Journel, 1992):

$$\gamma(h,\alpha) = \frac{1}{2N(h,\alpha)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \tag{4}$$

where

$\gamma(h,\alpha)$ = semivariance as a function of both the magnitude of the lag distance or separation vector (h) and its direction (α)

$N(h,\alpha)$ = number of observation pairs separated by distance h and direction α used in each summation

$Z(x_i)$ = random variable at location x_i.

For OK, the weighing parameter λ_i shown in equation 1 is determined to fulfill conditions as presented in equations 2 and 3 by solving a system of linear equations such as:

$$\begin{cases} \sum_{i=1}^{n(x)} \lambda_i(x)\gamma(x_j - x_i) - \mu(x) = \gamma(x_j - x) & j = 1, 2, ..., n(x) \\ \sum_{i=1}^{n(x)} \lambda_i(x) = 1 \end{cases} \tag{5}$$

where $\mu(x)$ is the Lagrangian parameter accounting for the constraint on the weights. The information needed for equation 5 are the semivariogram values (γ), which are estimated using equation 4. The spherical model is the most widely used semivariogram model and is characterized by the linear behavior at the origin (Goovaerts, 2000). The spherical model is:

$$\gamma(h,\alpha) = \begin{cases} S\left(1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right) & h \le a \\ S & h > a \end{cases} \tag{6}$$

where $\gamma(h,\alpha)$ is the spherical semivariogram with range parameter (a) and sill (S) for lag distance h.

Ordinary Co-Kriging (OCK)

Ordinary co-kriging is the estimation of one variable based on measured values of two or more variables. It is a generalization of kriging in the sense that at every location there is a vector of many variables instead of one variable. OCK is the multivariate extension of kriging (Goovaerts, 1997). In this study, the primary variables are standardized values of rainfall and elevation data at measured locations, which are obtained by using the mathematical association of rainfall and elevation given as:

$$Z^S(x_i) = \sqrt{\frac{[RF(x_i)]^2 + [EL(x_i)]^2}{2}} \tag{7}$$

where $Z^S(x_i)$ is the primary variate for OCK, and $RF(x_i)$ and $EL(x_i)$ are rainfall and elevation values, respectively, for the ith variate (i = 1, 2, ..., N, where N = 40). All the available rainfall data corresponding to 40 raingauge stations are standardized using equation 7. This transformation resulted in higher correlation coefficients between the elevation and transformed values rather than the use of non-transformed rainfall and corresponding elevation data of the raingauge network. Moreover, the mean of the primary and secondary variables are equivalent to each other, which is an essential requirement of OCK (Goovaerts, 1997). The OCK estimate, $Z^*(x_i)$, of $Z^S(x_i)$ is given by:

$$Z^*(x_i) = \sum_{i=1}^{n} \lambda_i Z^S(x_i) + \sum_{j=1}^{m} \eta_j S(x_j) \quad (8)$$

where $\{S(x_j); j = 1, 2, ..., m\}$ are available data of elevation used as secondary variates. The parameter weights (λ_i and η_j) are obtained as solutions of the OCK system given by the set of equations:

$$\begin{cases} \sum_{j=1}^{n} \lambda_j \cdot \gamma_Z(x_i, x_j) + \sum_{k=1}^{m} \eta_k \gamma_{ZS}(x_i, x_k) + \mu_1 \\ \quad = \bar{\gamma}_Z(x_i, B(x)) \qquad i = 1 \ldots n \\ \sum_{j=1}^{n} \lambda_j \cdot \gamma_{ZS}(x_k, x_j) + \sum_{t=1}^{m} \eta_t \gamma_S(x_k, x_t) + \mu_2 \\ \quad = \bar{\gamma}_{ZS}(x_k, B(x)) \qquad k = 1 \ldots m \\ \sum_{j=1}^{n} \lambda_j = 1 \\ \sum_{j=1}^{m} \eta_j = 0 \end{cases} \quad (9)$$

where μ_1 and μ_2 are Lagrange parameters, $\gamma_Z(h)$ is the variogram function of standardized parameters, $\gamma_S(h)$ is the variogram function of elevation values, and $\gamma_{ZS}(h)$ is the cross-variogram function of standardized parameters and elevation. $B(x)$ denotes the support of the estimate at location x of point kriging, and the total study region is representative of summation of all these support locations. The bar over the variogram denotes the mean variogram between a point and its support $B(x)$. The estimated variance of an OCK system is:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \lambda_i \cdot \bar{\gamma}_Z(x_i, B(x)) + \sum_{k=1}^{m} \eta_k \bar{\gamma}_{ZS}(x_k, B(x)) + \mu_1 \quad (10)$$

Collocated Ordinary Co-Kriging (CCK)

Collocated ordinary co-kriging is a conditional estimator of CK where the neighborhood uses the secondary variable as a subset of locations where primary data are available along with the estimated locations (Wackernagel, 2003). The standardized primary variate of the 40 measured values of rainfall depth and the corresponding elevation values were used in the analysis. The collocated elevation, $S(x)$, in CCK tends to influence the farther elevation points. The CCK estimate is given as:

$$Z^*(x) = \sum_{i=1}^{N(x)} \Big\{ \lambda_i^{CK}(x) Z^S(x_i) + \lambda^{CK}(x) \times [S(x) - m_S - m_Z] \Big\} \quad (11)$$

where $Z^*(x)$ is the CCK estimator with weight parameter λ_i^{CK}, primary variate $Z^S(x_i)$, and standardized secondary variate $S(x)$, and m_S and m_Z are the mean estimator of standardized rainfall and elevation values, respectively. The standardization of the primary variate using equation 7 was done to ensure that the mean values of the primary and secondary variates were equivalent to each other, which would result in an unbiased estimation (Goovaerts, 2000). The CCK cross-semivariogram is:

$$\gamma_{ZS}(h) = \frac{1}{2N(h)} \times \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)] \times [S(x_i + h) - S(x_i)] \quad (12)$$

where $\gamma_{ZS}(h)$ is the cross-semivariogram representing the standardized primary variable and the secondary variable representing the elevation points.

KRIGING ERROR STATISTICS

The performances of the OK, OCK, and CCK algorithms were assessed and compared using cross-validation results (Isaaks and Srivastava, 1989). This was achieved by temporarily removing one datum at a time from the data set and re-estimating the deleted value from the remaining data using the kriging algorithms. All computations have been performed using the GS+ and ArcGIS software (Johnston et al., 1996). In the present study, the coefficient of determination (R^2), the kriged reduced mean square error (KRMSE), and the kriged average error (KAE) were the error statistics used (Campling et al., 2001; Kitanidis, 1997) for comparison of the model-predicted results with the observed values.

The KRMSE was used to check the consistency between the estimation errors and the standard deviation of the observed values:

$$\text{KRMSE} = \left(\frac{1}{N} \sum_{i=1}^{N} \left[\frac{zo_i - zp_i}{s} \right]^2 \right)^{1/2} \quad (13)$$

The KAE was used to test the predictability of the developed models:

$$\text{KAE} = \frac{1}{N} \sum_{i=1}^{N} \frac{(zo_i - zp_i)}{s} \quad (14)$$

where

zo_i = observed value at location i
zp_i = predicted value at i through CCK and OK methods
N = number of pairs of observed and predicted values
s = standard deviation of the observed values.

This KRMSE value should be within the range $1 \pm [2(2/N)^{1/2}]$ for the model to be acceptable (Ella et al., 2001). The KAE value should be close to zero for the model to be acceptable (Kitanidis, 1997).

INTERPOLATION STEPS USING ARCGIS AND GS+

The primary variate for co-kriging was standardized with the rainfall and elevation values using equation 7. The exploratory data analysis was performed for both primary and secondary variates to ascertain the correlation coefficient, mean value, normality in data trend, and presence of any outliers before semivariogram fitting.

Semivariogram Model Selection

The best semivariogram model was generated by observing the R^2 and residual sum square (RSS) values with a trial and error approach for different lag sizes and lag intervals (Goovaerts, 1997; Isaaks and Srivastava, 1989). However, the lag sizes and number of lags were varied based on a heuristic rule, generally called a rule of thumb, in which the lag size times the number of lags should be less than one-half of the largest distance in the database (Johnston et al., 1996). In this study, the lag size and number of lags were selected and confined within one-third to one-half the maximum distance on a trial and error basis to generate an optimized semivariogram model. Then the optimum semivariogram model parameters, such as sill, nugget, range, and the fitted model type corresponding to the highest R^2 value, were noted. The sample variance where the semivariogram stabilizes is known as the sill, the *y*-axis intercept of the model is called the nugget, and the separation distance along the *x*-axis where the model flattens out is called the range (fig. 5). In CK, the primary and secondary variates were fitted using different models available in GS+ and the ArcGIS geostatistical extension module, and the optimal cross-semivariogram was selected. The OCK analysis was performed for the 40 primary data points (standardized elevation and rainfall values) and the secondary variate (elevation points extracted from the DEM). We also ensured that the difference in the average of both primary and secondary variates was close to zero. The CCK analysis was performed for 40 primary data points (standardized elevation and rainfall values) and the secondary variate as the elevation of the point locations included in the primary variate. Both the OCK and CCK methods were compared based on the variogram fitting statistics and cross-validation error statistics. This was performed to select the best method for prediction of rainfall at unmeasured locations and to predict the spatial variability.

The cross-validation error statistics were estimated using equations 13 and 14 to select the best algorithm model. The OCK and CCK cross-validation error statistics were compared to select the best method in performing the CK. Then, the best CK algorithm (either OCK or CCK) was used to predict the standardized primary variate values for more locations within the study region. The database of these predicted variables corresponding to elevation locations were then destandardized to obtain the rainfall depths at predicted unmeasured locations, $RF(x_i)$, by solving equation 7. To this destandardized database of predicted rainfall depths, the observed rainfall depths of the 40 measured locations were added. Further, OK analysis was performed for the data set with rainfall depth as the interpolating variate. Exploratory data analysis was performed to ascertain the normality of the distribution of the variates and the presence of outliers in the data before fitting the semivariogram model. Similarly, the steps of variogram fitting as described for CK were followed for the OK method, and the best-fitted semivariogram was used for kriging. The cross-validation statistics in terms of KRMSE and KAE were estimated to ascertain the model algorithms, and finally the interpolated surface reflecting the variation of rainfall depth over the mountainous region of St. Lucia was generated using ArcGIS. These procedures were followed to generate the spatial variability map of long-term mean monthly rainfall for one month and replicated for all twelve months to generate twelve such maps.

Results and Discussion

The point rainfall data of the study region at 40 raingauge locations, the DEM (table 1), and the base map of St. Lucia were projected to the Transverse Mercator (TM) projection system with the ground distance represented in meters. Four hundred elevation points covering the entire island, excluding the raingauge points, were extracted from the DEM for prediction of rainfall values and determination of the interpolated surface (fig. 2).

Results of OCK and CCK for Prediction of Mean Monthly Rainfall Depths

The exploratory data analysis performed on the standardized primary variate and the secondary variate of elevation revealed that the data are normally distributed and free of outliers. The regression analysis of the primary variate (mean monthly rainfall depth) and secondary variate (elevation) resulted in poor R^2 values, ranging from 0.17 (June) to 0.41 (May). However, the regression analysis of the standardized primary values generated using equation 7 with the elevation values resulted in better R^2 values, ranging from 0.80 (October) to 0.97 (March). Therefore, the standardized primary variate was used with the elevation values for CK analysis. It was observed from the fitted semivariogram that using only the elevation points as a secondary variate for which the primary variate values (standardized) were known (i.e., CCK) resulted in R^2 values of the fitted spherical semivariogram ranging from 0.78 to 0.82 with optimal values of sill [118 $(m^2 + mm^2)^{0.5}$] and range (13.78 km) parameters. In comparison, when using more elevation points as the secondary variate (i.e., OCK), the R^2 of the fitted spherical semivariogram ranged from 0.91 to 0.99 with optimal sill [98$(m^2 + mm^2)^{0.5}$] and range (15 km) parameters. The lag size of 25 km and lag class interval of 3.3 km were observed to be the optimal values for obtaining the best-fit cross-semivariogram for all the data sets. Moreover, the cross-validation statistics performed for both the OCK and CCK methods of CK (table 2) revealed that the OCK method performed better than the CCK method.

The calculated KRMSE of cross-validation results for nine months were well within the acceptable range for the model obtained through the OCK method, whereas the

Table 1. GIS thematic data layers used in rainfall spatial modeling.

Data Layer	Data Representation	Data Sources
Digital elevation model (DEM)	50 × 50 m resolution raster	Topographic map sheets of 1:25000 scale acquired from Department of Overseas Surveys (DOS), St Lucia.
Base map	Boolean raster of island extents	Topographic map sheets of 1:25000 scale acquired from Department of Overseas Surveys (DOS), St Lucia
Mean annual rainfall	Interpolated raster from vector coverage	St. Lucia Development Atlas, 1:50000 map series (OAS, 1987)
Rainfall stations	Vector-point	U.N. Preliminary Water Resources Report (Migeot and Hadwen, 1986)

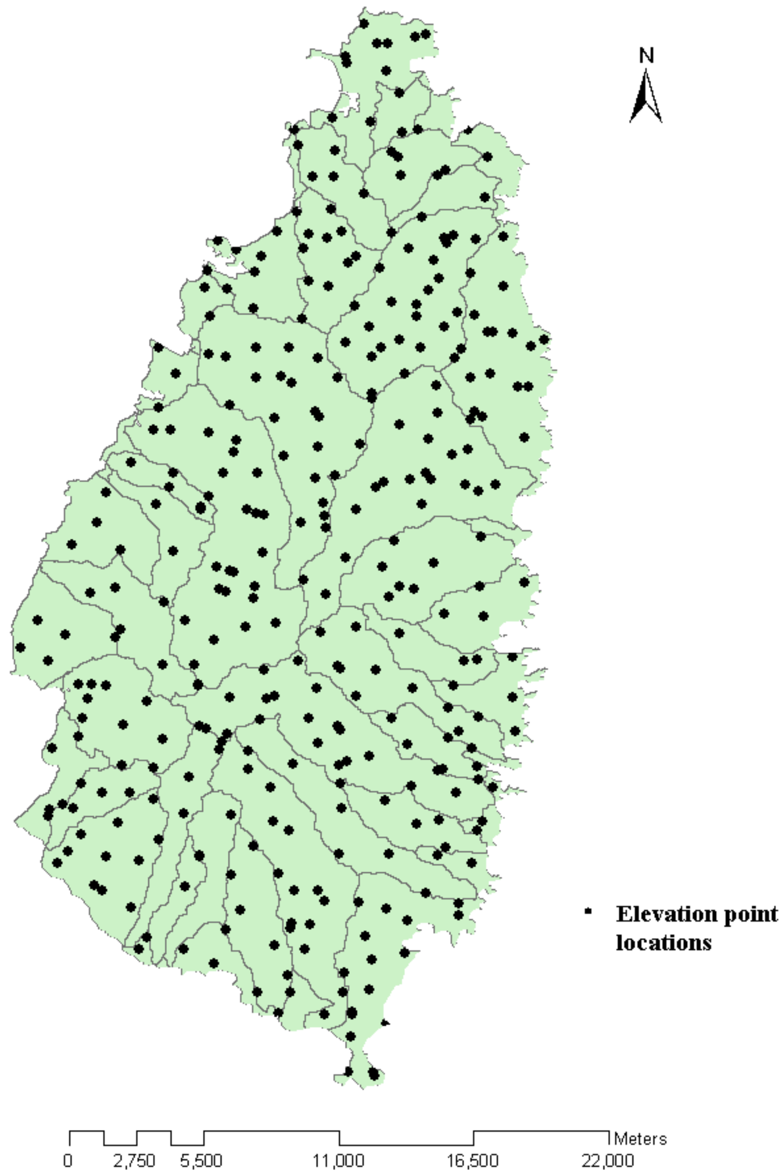

Figure 2. Location of extracted elevation points from DEM of St. Lucia.

Table 2. Comparison of OCK and CCK methods in terms of model fitting R^2 values and cross-validation statistics.

	Ordinary Co-kriging				Collocated Ordinary Co-kriging			
	R^2 of Cross-Semivariogram Model (Spherical)	Cross-Validation Statistics			R^2 of Cross-Semivariogram Model (Spherical)	Cross-Validation Statistics		
Month		R^2[a]	KAE[b]	KRMSE[c]		R^2[a]	KAE[b]	KRMSE[c]
January	0.95	0.97	−0.001	0.56	0.82	0.78	−0.012	0.18
February	0.98	0.95	0.013	0.42	0.84	0.85	0.018	0.32
March	0.98	0.94	0.017	0.75	0.84	0.84	0.037	0.86
April	0.98	0.97	−0.001	0.86	0.84	0.87	−0.09	0.18
May	0.94	0.97	0.006	0.62	0.82	0.87	0.12	0.17
June	0.85	0.95	−0.012	0.32	0.78	0.85	−0.13	0.20
July	0.94	0.99	0.004	0.76	0.84	0.89	0.08	0.62
August	0.89	0.98	−0.0003	0.58	0.84	0.88	−0.09	0.12
September	0.90	0.99	−0.007	1.01	0.76	0.79	0.17	0.1
October	0.90	0.99	−0.001	0.81	0.79	0.79	0.06	0.11
November	0.88	0.9	−0.012	0.29	0.80	0.83	0.031	0.56
December	0.94	0.97	−0.003	0.92	0.78	0.87	−0.043	0.16

[a] The R^2 value should close to 1 for better fit.
[b] The acceptable value of KAE is close to zero.
[c] The acceptable value of KRMSE ($1 \pm [2(2/N)^{1/2}]$) is 0.55 to 1.45 (N = 40).

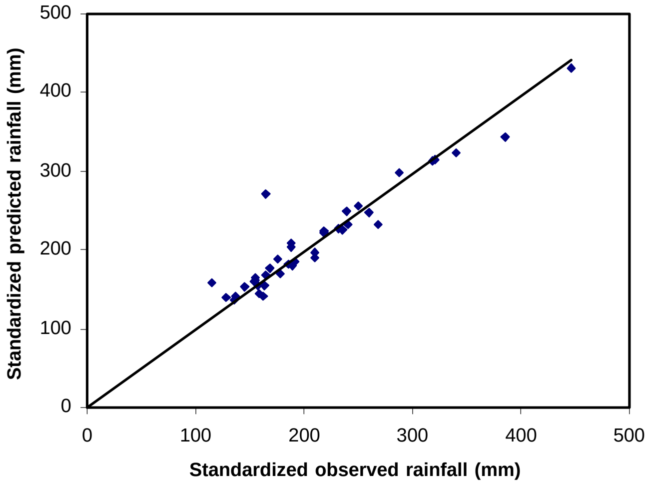

Figure 3. The 1:1 line of the predicted and observed rainfall values (standardized) for mean monthly rainfall of November (R^2 = 0.90).

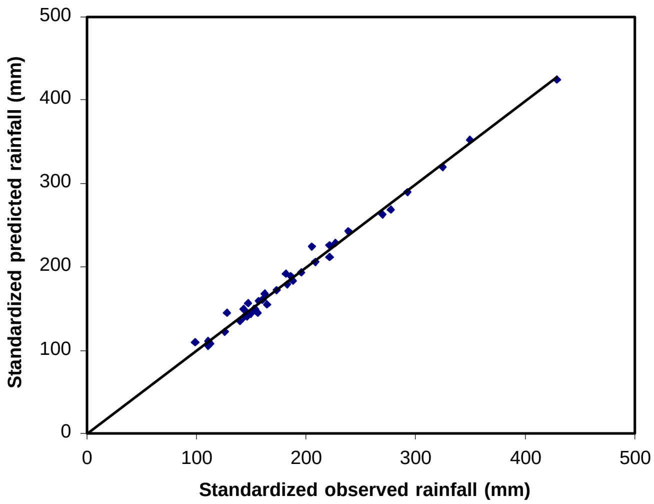

Figure 4. The 1:1 line of the predicted and observed rainfall values (standardized) for mean monthly rainfall of September (R^2 = 0.99).

KRMSE values of only three months were within acceptable range (table 2) when the CCK algorithm was used. Moreover, the KAE values calculated from the cross-validation results of the CCK and OCK algorithms methods for all months were close to zero. The 1:1 lines for the highest and lowest R^2 value of the cross-validation results of OCK are shown in figures 3 and 4. Hence, the OCK algorithm was used to predict the standardized values for 400 elevation point locations using the geostatistical extension tool in ArcGIS. The reason for using the limited elevation points representing the entire watershed by grid layer overlay technique was to minimize the prediction error and subsequent error in spatial interpolation. Finally, the predicted database was analyzed to extract the rainfall depths corresponding to the elevation values by solving equation 7.

It was observed that a few destandardized values of rainfall depth obtained by solving equation 7 resulted in negative values. To avoid this, the minimal difference in the mean values of the primary variate and the secondary variate were ensured by optimizing the number of elevation points through the trial approach (Goovaerts, 2000). Despite the above technique to minimize the negative rainfall depths, for all the analysis, the number of such negative values ranged from 1 to 8, and they were finally excluded from the data set before using the OK analysis to generate the rainfall spatial variability map.

Results of OK Analysis for Generation of Rainfall Variability Map

The predicted mean monthly rainfall values from the CK algorithm and the observed data of raingauge locations for all twelve months were subjected to ordinary kriging analysis. Exploratory data analysis was done to ensure normality of data trends and to determine if any outliers were present in the data. The data for all the months were observed to be distributed normally without any outliers. An empirical variogram model was fitted with the data set using different lag sizes and numbers of lag, and the fitted semivariogram model with the fitting statistics are listed in table 3. The lag size of 25 km and the lag class interval of 3.25 km were observed to be the optimal values for obtaining the best-fit variogram for all the data sets. It was estimated that the spherical semivariogram fit well with all the data sets, with the R^2 value varying from 0.84 (February) to 0.98 (October, November, and July). The fitted semivariograms for the months of February and November are presented in figures 5 and 6, respectively. The cross-validation error statistics with the KRMSE, KAE, and R^2 values are presented in table 4. It was observed that the KAE was close to zero for all the months, and the R^2 value varied from 0.71 (February) to 0.96 (June). The observed and predicted values with lowest and highest R^2 values are shown in figures 7 and 8, respectively.

Table 3. Fitted semivariogram parameter for the predicted and observed rainfall values in OK analysis.

Month	R^2 of Semivariogram Model (Spherical)	Nugget (mm^2)	Sill (mm^2)	Range (m)	Spherical Model
January	0.88	140	3749	11540	The isotropic spherical variogram model is given by:
February	0.84	372	3536	11480	
March	0.92	455	2975	11710	
April	0.89	492	2839	10930	
May	0.86	110	3766	9840	$\gamma(h) = \begin{cases} S\left(1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right) & h \le a \\ = S & \text{otherwise} \end{cases}$
June	0.97	1299	3692	11500	
July	0.98	1010	5615	11580	
August	0.93	1100	4348	10360	
September	0.94	840	4281	9810	where *S* is the sill and *a* is the range for lag distance *h*.
October	0.98	1760	7037	10790	(The parameters are estimated from this table.)
November	0.98	1990	6262	9850	
December	0.92	890	5148	10060	

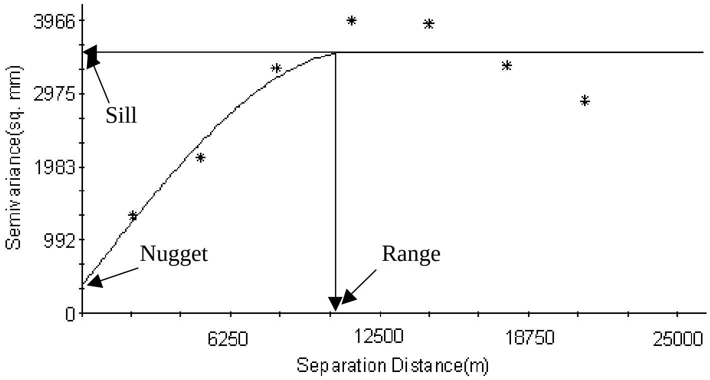

Figure 5. Semivariogram of ordinary kriging method for February (R^2 = 0.84).

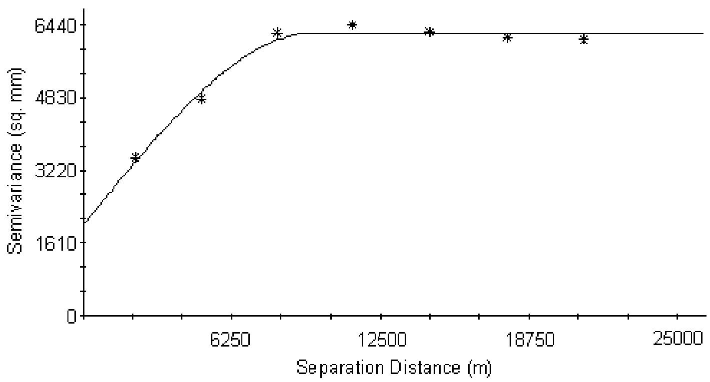

Figure 6. Semivariogram of ordinary kriging method for November (R^2 = 0.98).

The interpolated surface was generated for the study region for the twelve months, and the map of driest (March) and wettest (October) months of the year are presented in figures 9 and 10, respectively.

It was observed from the kriged surface that the spatial variability of rainfall from the leeward direction to the center of the island was very high for all the months in a year. The maximum rainfall for all the months occurred over the center of the island and the magnitude varied from month to month, with lowest rainfall depth occurring during March and the highest during the month of October. This observation was similar to observed rainfall values over St. Lucia.

CONCLUSIONS

It is understood that modeling rainfall spatial variability of a mountainous region with a sparse raingauge network poses a challenging task in terms of prediction accuracy. Furthermore, the short observation record of some stations and inconsistency in data recordings also influence the rainfall predictability over the mountainous region. This necessitates the use of exploratory data analysis techniques as a prerequisite before using the data for geostatistical modeling.

The combination of multivariate and univariate kriging approaches used in this study to predict the spatial rainfall variability for mountainous regions with orographic effects is a simple and reasonable approach, which can be applied to similar locations having sparse raingauge locations and undulating topography.

The standardization technique used in CK through mathematical association of rainfall depth and corresponding elevation values resulted in better prediction of rainfall variability over St Lucia. However, a similar approach of mathematical transformation needs to be applied to other mountainous regions for wide acceptability of this developed concept.

Table 4. Cross-validation error statistics of the fitted semivariogram models in OK.

	Ordinary Kriging			
	R^2 of Cross-Semivariogram Model (Spherical)	Cross-Validation Statistics		
Month		R^2[a]	KAE[b]	KRMSE[c]
January	0.88	0.79	0.023	0.57
February	0.84	0.71	0.032	0.81
March	0.92	0.88	−0.002	0.54
April	0.89	0.94	0.01	1.01
May	0.86	0.93	0.03	0.78
June	0.97	0.96	0.02	0.81
July	0.98	0.85	0.03	1.07
August	0.93	0.94	0.05	0.71
September	0.94	0.94	0.05	0.72
October	0.98	0.93	0.02	0.81
November	0.98	0.93	0.03	0.82
December	0.92	0.85	0.05	0.92

[a] The R^2 value should close to 1 for better fit.
[b] The acceptable value of KAE is close to zero.
[c] The acceptable value of KRMSE ($1 \pm [2(2/N)^{1/2}]$) is 0.86 to 1.14 (for N = 440)

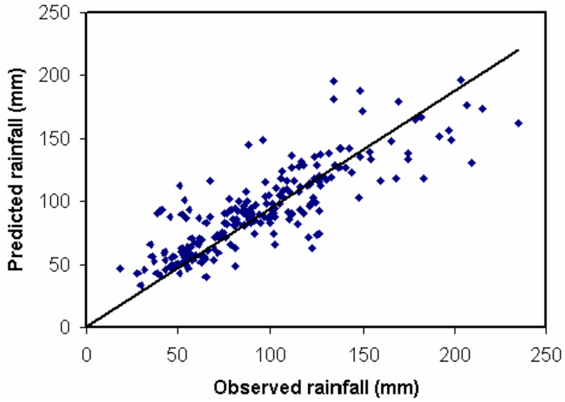

Figure 7. The 1:1 line of the predicted and observed rainfall values for mean monthly rainfall using ordinary kriging for February ($R^2 = 0.71$).

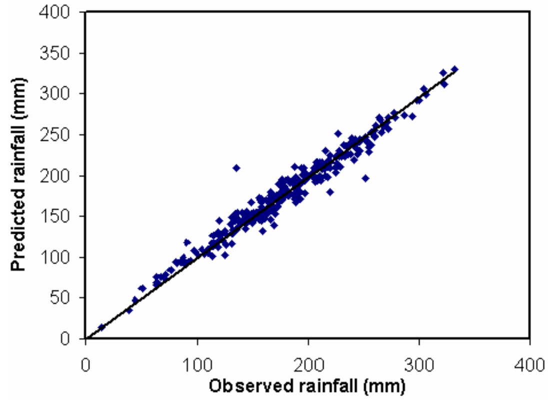

Figure 8. The 1:1 line of the predicted and observed rainfall values for mean monthly rainfall using ordinary kriging for June ($R^2 = 0.97$).

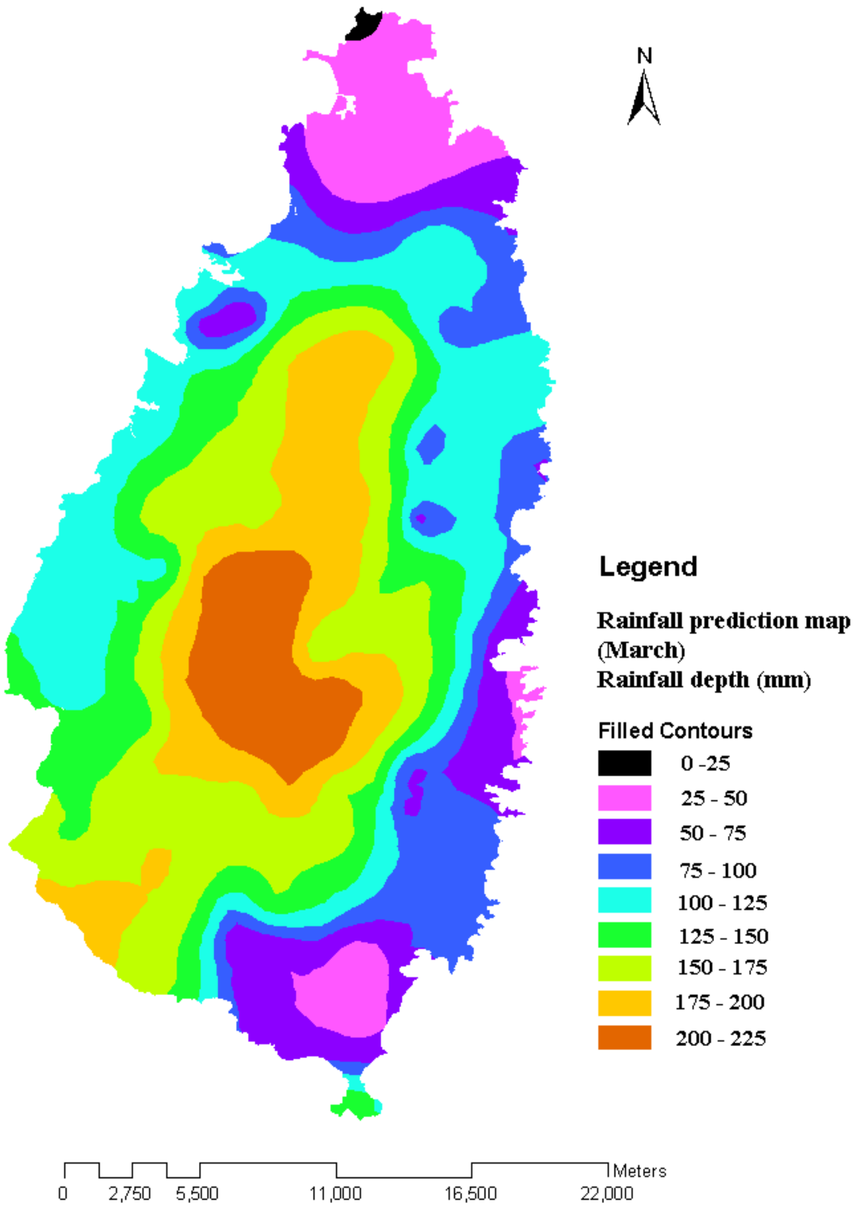

Figure 9. Rainfall prediction map of the driest month (March).

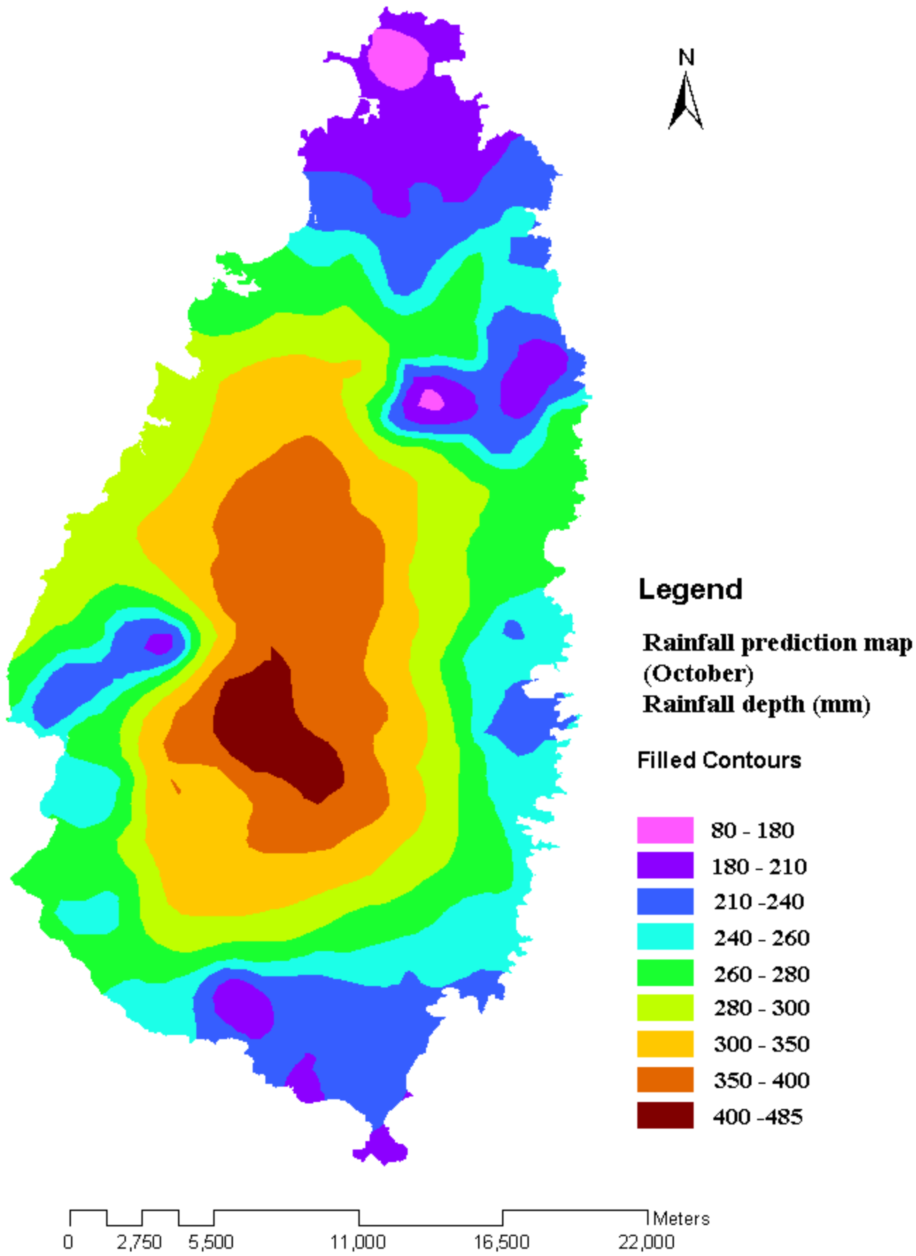

Figure 10. Rainfall prediction map of the wettest month (October).

Fitting the semivariogram model also depends on different lag sizes and lag class intervals, and the developed semivariogram affects the cross-validation results. Efforts should be made to generate an optimal variogram model, keeping in mind the fitting parameters. The R^2 value does not necessarily represent a best-fitted model. However, the KRMSE and KAE of the cross-validation results need to be considered to ascertain the best geostatistical algorithm.

Nonetheless, considering the study objectives, the modeled rainfall surfaces were considered reasonable, particularly when compared to the published annual rainfall surface (OAS, 1987). The developed methodology of geostatistical analysis and mathematical association of rainfall and elevation values to estimate a standardized value for use in CK method and the combination of OCK and OK approaches to generate rainfall prediction maps can be applied to account for the spatial variability of rainfall over mountainous regions.

References

Bacchi, B., and N. T. Kottegoda. 1995. Identification and calibration of spatial correlation patterns of rainfall. *J. Hydrology* 165: 311-348.

Campling, P., A. Gobin, and J. Feyen. 2001. Temporal and spatial rainfall analysis across a humid tropical catchment. *Hydrological Processes* 15(3): 359-375.

Christel, P., and D. W. Reed. 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: A case study in Scotland. *Intl. J. Climatology* 19(12): 1337-1356.

Cox, C. A. 2002. Geographic Information Systems-based hydrology modeling applications in support of watershed management planning for St. Lucia. PhD diss. Montreal, Quebec: McGill University, Department of Bioresource Engineering.

Deutsch, C. V., and A. G. Journel. 1992. *GSLIB Geostatistical Software Library and User's Guide*. New York, N.Y.: Oxford University Press.

Dirks, K. N., J. E. Hay, C. D. Stow, and D. Harris. 1998. High-resolution studies of rainfall on Norfolk Island: Part II. Interpolation of rainfall data. *J. Hydrology* 208(3-4): 187-193.

Drogue, G., J. Humbert, J. Deraisme, N. Mahrb, and N. Freslonc. 2002. A statistical-topographic model using an omni-directional parameterization of the relief for mapping orographic rainfall. *Intl. J. Climatology* 22(5): 599-613.

Ella, V. B., S. W. Melvin, and R. S. Kanwar. 2001. Spatial analysis of NO_3-N concentration in glacial till. *Trans. ASAE* 44(2): 317-327.

FAO. 1996. Agro-ecological zoning: Guidelines. FAO Soils Bulletin 73. Rome, Italy: U.N. Food and Agricultural Organizations.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. New York, N.Y.: Oxford University Press.

Goovaerts, P. 1999. Using elevation to aid geostatistical mapping of rainfall erosivity. *Catena* 34(3-4): 227-242.

Goovaerts, P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrology* 228(1-2): 113-129.

Hevesi, J. A., J. D. Istok, and A. L. Flint. 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics: Part I. Structural analysis. *J. App l. Meteorology* 31(7): 661-676.

HTS. 1997. *Watershed and Environmental Management Project, Final Report.* Vol. 3. Hemel Hampsted, U.K.: Hunting Technical Services.

Isaaks, E. H., and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. New York, N.Y.: Oxford University Press.

Johnston, K., J. M. V. Hoef, K. Krivoruchko, and N. Lucas. 1996. *Using ArcGIS Geostatistical Analysis*. New York, N.Y.: ESRI.

Kitanidis, P. K. 1997. *Introduction to Geostatistics: Application to Hydrology*. Cambridge, U.K.: Cambridge University Press.

Lebel, T., G. Bastin, C. Obled, and J. D. Creutin. 1987. On the accuracy of aerial rainfall estimation: A case study. *Water Resources Research* 23(11): 2123-2134.

McCuen, R. H. 1998. *Hydrologic Analysis and Design*. 2nd ed. Englewood Cliffs, N.J.: Prentice Hall.

Migeot, J., and P. Hadwen. 1986. *St. Lucia Water Resources: Preliminary Assessment*. Vols. 1 and 2. Castries, St. Lucia: Ministry of Agriculture.

OAS. 1987. *Saint Lucia Development Atlas*. Washington, D.C.: Organization of American States, Department of Regional Development.

Phillips, D. L., J. Dolph, and D. Marks. 1992. A comparison of geostatistical procedures for spatial analysis of precipitations in mountainous terrain. *Agric. Forest Meteorology* 58(1-2): 119-141.

Thiessen, A. H. 1911. Precipitation averages for large areas. *Monthly Weather Review* 39(7): 1082-1084.

Wackernagel, H. 2003. Multivariate geostatistics. In *Multivariate Geostatistics: An Introduction with Applications*, 145-169. 3rd ed. New York, N.Y.: Springer-Verlag.

Webster, R., and M. A. Oliver. 2001. Cross-correlation, coregionalization, and co-kriging. In *Geostatistics for Environmental Scientists*, 271-273. Chichester, U.K.: John Wiley and Sons.