Orographic Precipitation Modeling with Multiple Linear Regression

S. Naoum, M.ASCE,[1] and I. K. Tsanis, M.ASCE[2]

Abstract: A multiple linear regression (MLR) model, in conjunction with Geographic Information Systems technology, was used to derive the relationship between annual precipitation and elevation, longitude, and latitude. The island of Crete, in Greece, was used as the case study. A multiscale precipitation analysis was performed on areas ranging from large areas (the whole island and the northern, southern, and eastern parts of the island), to medium areas (watersheds), to small areas (sub-basins). While the MLR annual precipitation estimates (which used elevation, latitude, and longitude information) were found to be more reasonable than estimates obtained using elevation only when applied to the whole island, the difference between the MLR estimates and the elevation-only estimates was smaller when applied to individual watersheds. The MLR provides realistic estimates for mean areal precipitation for the island of Crete: 700±100, 950±150, and 1,300±200 mm for dry, average, and wet years, respectively. Elevation-rainfall gradients are: 0.45–0.6, 0.6–0.9, and 0.9–1.3 mm/m for dry, average, and wet years, respectively. Of this, 44% falls on the northern, 33% on the southern, and 23% on the eastern parts of the island for a typical average year.

DOI: 10.1061/(ASCE)1084-0699(2004)9:2(79)

CE Database subject headings: Geographic information systems; Multiple regressions; Precipitation; Statistic analysis; Greece.

Introduction

The spatio-temporal variation in precipitation is due to seasonal and/or geographical variation. The tropics, for example, (from about 10°N to 10°S) may or may not have seasons. Rather, precipitation varies with the sun's position, which determines the spatial distribution of the heat energy it supplies to the atmosphere and oceans. Rainfall tends to be heaviest near the equator and diminishes as the air flows toward higher latitudes. This is caused by abundant warm moist air drifting continually from tropical seas and flowing north and south (Singh 1992). Other parts of the world have a more definite rainy season, which may occur in summer or in winter. Southern Africa has its rainy season mainly during the summer, whereas Greece has one in winter (Chow et al. 1988). The distribution of precipitation, on the other hand, depends upon latitude, orographic factors, and the distance the air mass moves away from the source of moisture. The continentality of an area is determined by the distance from the shore. Although there is no simple relation between the distance from the shore and continentality, it is clear that temperature ranges are greater over continental masses than elsewhere. These temperature variations also affect precipitation and the type of precipitation. The direction of prevailing wind can be determined by the direction of storm movements. The direction of prevailing wind also indicates the susceptibility of an area to the influence of the storm systems. Topography refers to the nature and elevation of land masses. Mountains, valleys, and all variations of these features are part of topography. A low pressure system picks up moisture over the ocean and transports that moisture over land. If that system is forced to move from sea level up over mountains, it will encounter cooler temperature at higher elevations, which will cause loss of moisture. If a low-pressure system is moving from west to east and has to cross several high elevations (or mountains) that are situated north–south, it will give out some of its moisture on each one of those mountains. If there is no additional source of moisture, the far east side of the elevations will receive little, if any, moisture. If those mountains are oriented east–west, however, the effect will cause less moisture to be lost. This is known as the orientation of topography factor. If two stations are on the same elevation on a mountain slope but on opposite sides of the mountain, the aspect factor will determine which station will get the greater amount of precipitation. The station with less precipitation is recognized as a "rain shadow" station. If the station is situated on a mountain slope, its aspect relative to the main axis of the mountain is measured in directional units of 10° clockwise from north.

[1]Research Associate, Dept. of Civil Engineering, McMaster Univ., 1280 Main St. West, Hamilton ON, Canada L8S 4L7.

[2]Professor, Dept. of Civil Engineering, McMaster Univ., 1280 Main St. West, Hamilton ON, Canada L8S 4L7 (corresponding author). E-mail: tsanis@water.eng.mcmaster.ca

Note. Discussion open until August 1, 2004. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on March 18, 2002; approved on October 21, 2002. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 9, No. 2, March 1, 2004. ©ASCE, ISSN 1084-0699/2004/2-79–102/$18.00.

Historical Documentation of Areal Precipitation Estimators

Most methods for estimating areal and gridded precipitation from point data would relate to any of these three major groups: graphical, topographical, and numerical. Graphical methods involve mapping of precipitation data, sometimes in combination with precipitation–elevation analyses, and include isohyetal mapping (Reed and Kincer 1917; Peck and Brown 1962) and Thiessen polygons (Thiessen 1911). Topographical methods involve the correlation of point precipitation data with an array of parameters such as slope, exposure, elevation, location of barriers, and wind speed and direction (Spreen 1947; Burns 1953; Schermerhorn

J. Hydrol. Eng. 2004.9:79-102.

1967; Houghton 1979). Numerical methods involve the spatial interpolation techniques in which a numerical function is used to weight irregularly spaced point data to estimate a regularly spaced prediction grid. In this paper, a grid refers to a two-dimensional array of regularly spaced grid cells. A grid cell refers to a single pixel that has dimensions equivalent to the resolution of the grid. For example, each cell in a 1-km resolution grid is 1 km $\times$ 1 km in size. A value assigned to a pixel, such as a precipitation estimate, is positioned at the cell center. However, it is not a point value; rather it represents an average value over the entire cell. Inverse distance weighting (IDW) is an example of a simple numerical interpolation method. In this case, the weighting of the data points is prescribed to decrease as the distance between the points increases. Spline (or thin plate) interpolation is another method, which is distinguished by two main features: (1) the surface must pass exactly through the data points, and (2) the surface must have minimum curvature. Kriging is a geostatistical approach (Matheron 1971) that has been well received and recognized as a tool for the interpolation of many types of data, including precipitation (Chua and Bras 1982; Dingman et al. 1988; Phillips et al. 1992). In kriging, a semivariogram model that best fits the data is developed to derive optimum station weights for interpolation. A potential drawback of kriging is that it implicitly relies on the data to directly represent the spatial variability of the actual precipitation field (Daly et al. 1994). If the data are not representative, as is often the case on a complex terrain, the accuracy of the resulting interpolated field will be in question. In addition, more than one semivariogram may be needed to estimate precipitation at various time periods. Another drawback, is the difficulty of developing a semivariogram for sparse networks. Linear trend surface is another interpolation method that uses a polynomial regression to fit a least-squares surface to the input points. It allows the user to control the order of the polynomial used to fit the surface. The surface generated will seldom pass through the original data points since it performs a best fit for the entire surface.

Attempts also were made to use digital elevation models (DEMs) to predict the physical influence of topographic factors on precipitation patterns (Peck and Schaake 1990; Fan and Duffy 1991; Hay et al. 1991; Daly et al. 1994). The PRISM model, which was developed by Daly et al. (1994), somewhat incorporates some general aspects of the physical nature of the orographic influence on precipitation in addition to the analytical nature of the numerical approaches. However, it is highly unlikely to correctly identify the aspect (and the elevation) of each precipitation gauge using a DEM, especially if only relatively coarse elevation grids are available. Big complex terrain areas are better studied, in this case, when divided into smaller regions each of general aspect (facet).

Orographic Effects

The orographic influence occurs only in the proximity of high ground in the case of stable atmosphere. There are three main mountain effects: orographic lifting, thermal forcing, and obstacle effects which include mountain blocking, flow deflection, and the production of lee-side flow disturbances.

Precipitation–Elevation Relationships

On a given mountain slope, precipitation typically increases with elevation (Alter 1919; Barrows 1933; Spreen 1947; Schermerhorn 1967; Hibbert 1977; Smith 1979). This phenomenon, commonly called the orographic effect, is evident worldwide. Depending on its size and orientation, a mountain or range of mountains can increase the intensity of cyclonic precipitation by retarding the rate of movement of the storm and causing forced uplift of the air mass (Barry and Chorley 1976; Marwitz 1987). In summer, the orographic effect may trigger a conditional or convective instability in an otherwise stable air mass, producing a local redistribution of precipitation over the higher grounds (Daly et al. 1994). Under some conditions the relationship between precipitation and elevation may be best described by log-linear or exponential functions, but the linear form is easy to use and appears to be an acceptable approximation in most situations (Daly et al. 1994).

Spatial Scale of Orographic Effects

Because we use a DEM as the source of spatially gridded elevation data, the scale at which orographic effects are observed is of great importance. There is an implicit mismatch in scale when using relationships between station point elevations and precipitation to estimate precipitation at DEM grid cells. The elevation at the center of a DEM grid cell does not represent the elevation at that point, but reflects an average elevation representing the entire grid cell. Therefore, the DEM cell elevation will rarely match the station elevation. Generally, the finer the resolution of the DEM, the more closely the elevation of the grid cell will match that of the point (Daly et al. 1994).

Also, the orographic scale depends on the scale of the prevailing storm type. Large-scale frontal systems have inherently larger scales than localized convective cells. The orographic scale illustrated in the data may also depend on the relationship between the scale of the topographic features involved and the density and placement of the data points (Daly et al. 1994). Another potential factor is the temporal resolution of the data. Small-scale orographic effects may be more likely to be resolved in short term interval data than in data averaged over a long time period (Daly et al. 1994). The best DEM resolution therefore becomes a function of the scale of orographic effects, data density, and temporal resolution of the data.

Spatial Patterns of Orographic Regimes

The characteristics of the relationship between measured precipitation and topographic factors can vary considerably from hillslope to hillslope, and are influenced, in addition to factors previously mentioned, by differences in steepness of the terrain, upwind barriers, and slope orientation. A usable relationship between precipitation and topographic factors is difficult to be determined, unless rainfall stations are grouped into regions that deal with these factors. To effectively model the spatial pattern of orographic precipitation in complex terrain, it is essential that topographic regions be recognized and isolated (Daly et al. 1994). A mountainous landscape can be divided into regions, each of which is assumed to experience a different orographic regime (as will be discussed later in ‘‘Analysis’’).

To summarize, although the demand for precipitation fields on a regular grid is growing dramatically, the lack of data for mapping orographic precipitation has always represented a problem when trying to develop precipitation grids for complex terrain. The purpose of this study is to develop a method for distributing point measurements of annual orographic precipitation to regularly spaced grid cells at different spatial scales by relating a spatial dependent variable (annual precipitation) to spatial independent or predictor variables (described above) by means of the multiple linear regression method.

J. Hydrol. Eng. 2004.9:79-102.

Role of Geographic Information System

Geographic Information System (GIS) technology has been increasingly used in supporting water resources management applications. These systems can be linked to/integrated with different, environmental simulation models that provide functions for data storage, calculation of required input parameters, data manipulation, and output processing (Maidment 1993; Ross and Tara 1993; Goodchild et al. 1995; Greene and Cruise 1995; Brimicombe and Bartlett 1996; Tsanis et al. 1996; Lichy 1998; Naranjo and Larsen 1998). Geographic Information System technology is also central to these models because it provides the system with spatial data management, analysis, display, and interface functions (Boyle et al. 1998; Boyle and Tsanis 1998). The GIS sophisticated spatial analysis tools, graphic generation and visualization capabilities have resulted in a rapid expansion in the number of areas in which these systems can be applied. An example of these systems, integrated with a statistical package, is being employed in this study. Initially, *ArcView* GIS 3.2 [Environmental Systems Research Institute (ESRI) 2000] together with the S-Plus extension (S-Plus for *ArcView* GIS) was extensively used for the purpose of this paper. The system provided a good tool for spatial analysis, math applications (including linear regression), and graphics generation. The programming language for *ArcView* (AVENUE) provided a well-defined mechanism for allowing user-written routines to be called from within the normal user interface of the GIS package. This resulted in the development of the *ArcView* Regression Utility (AVRU) (Naoum and Tsanis 2004). AVENUE was used to generate a menu-driven graphic interface, which made it possible to estimate total volume and average areal values of rainfall by using the multiple linear regression (MLR) models and spatial interpolation techniques.

Regression Method

Statistical methods can be used to describe the behavior of a set of observations by focusing attention on the observations themselves rather than on the physical processes that produced them. One of those statistical methods is regression. Regression analysis is a technique for analyzing raw data and searching for the messages they contain, hence providing certain insights into how to plan the collection of data when the opportunity arises. In any system in which variable quantities change, it is of interest to examine the effects that some variables exert (or appear to exert) on others. There may, in fact, be a simple functional relationship between variables. In most physical processes, however, this is the exception rather than the rule. Often a functional relationship exists that is too complicated to describe in simple terms. In this case we may wish to approximate to this functional relationship by some simple mathematical function, such as a polynomial, which contains the appropriate variables and which graduates or approximates to the true function over some limited ranges of the variables involved. By examining such a graduating function we may be able to learn more about the underlying true relationship and appreciate the separate and joint effects produced by changes in certain important variables. Even where no sensible physical relationship exists between variables, we may wish to relate them by some sort of mathematical equation. While the equation might be physically meaningless, it may, nevertheless, be extremely valuable for predicting the values of some variables from knowledge of other variables, perhaps under certain stated restrictions (Draper and Smith 1998). In this paper, one particular method of obtaining a mathematical relationship is used. This involves the initial assumption that a certain type of relationship, linear in unknown parameters, holds. The unknown parameters are then estimated under certain other assumptions with the help of available data so that a fitted equation is obtained. The method of analysis used is the method of least squares (LS), which is simply a minimization of the sum of squares of deviation of the estimated values from the true values. The ordinary LS was favored over the weighted LS due to the few number of outliers in the area of study and their low departure. The model function is of a specified form that involves both the predictor variables and the parameters. The distribution of the random errors (random error $=$response variable$-$model function) is assumed to be normal distribution with mean zero, and errors are usually assumed to be independent. Various checks are performed thereafter, such as model significance, parameters significance, residual analysis, etc.

Inferences on Individual Parameters

Testing the significance of the model parameters (β_i) that is assigned to predictor variable (x_i) is done as follows: if we assume that the variations of the observations about the best-fit regression line are normal—that is, the errors ε_i are all from the same normal distribution, $N(0,\sigma^2)$, it can be shown that we can assign $100(1-\alpha)\%$ confidence limits for β_i by calculating

$$\hat{b}_i \pm \frac{t_{(n-p,1-0.5\alpha)}s}{\{\Sigma(x_i-\overline{x})^2\}^{1/2}} \tag{1}$$

where $\hat{b}_i$=estimate of parameter β_i; s=estimate of standard deviation σ; n=sample size; $t_{(n-p,1-0.5\alpha)}=100(1-0.5\alpha)$ percentage point of a t-distribution, with $(n-2)$ degrees of freedom (the number of degrees of freedom on which the estimate of the variance s^2 is based); p=number of parameters in the model; and $\overline{x}$=average value of x_i. On the other hand, if a test is appropriate, we can test the null hypothesis (H_0) that β_i is equal to zero $(H_0:\beta_i=0)$ against the alternative hypothesis (H_1) that β_i is not equal to zero $(H_1:\beta_i\neq 0)$ by comparing the $|t|$ with $t_{(n-p,1-0.5\alpha)}$ from a t-table with $(n-p)$ degrees of freedom. The test is a two-sided test conducted at the $100\alpha\%$ level of significance. The null hypothesis cannot be rejected if it happens that the observed $|t|$ value is smaller than the critical value and the parameter is insignificant; or reject the null hypothesis, however, if the observed $|t|$ value is greater than the critical value, making the parameter significant.

Analysis of Variance

The analysis of variance is an attempt to answer the question of how much of the variation in the data has been explained by the regression model. For any given linear model:

$$\begin{aligned}&(\text{sum of squares about the mean})\\ &\quad=(\text{sum of squares due to regression})\\ &\quad\quad+(\text{sum of squares about regression})\end{aligned} \tag{2}$$

This shows that some of the variation in the data about their mean can be attributed to the regression and some to the fact that the actual observations do not all lie on the regression line. If they all did, the sum of squares (SS) about regression (residual sum of squares) would be zero, which yields a perfect regression model.

Generally, it is favorable to see that the SS due to regression is much greater than the SS about regression. Any sum of squares is associated with a number called its degrees of freedom. This number indicates how many independent pieces of information involving the independent numbers are needed to compile the sum of squares. Due to regression, the SS is associated with a $(df_{\text{Reg}}=p-1)$ degrees of freedom, while the SS about regression is associated with a $(df_E=n-p)$ degrees of freedom.

Model Significance

The F-test is used to test the significance of regression. Since the sample variables are random variables, any function of them is also a random variable. Two particular functions are: (1) the mean square due to regression

$$\left(\text{MS}_{\text{Reg}}=\frac{\text{SS}_{\text{Reg}}}{df_{\text{Reg}}}=\frac{\text{SS}_{\text{Reg}}}{p-1}\right)$$

and (2) the mean square due to residual variation

$$\left(\text{MS}_E=S^2=\frac{\text{SS}_E}{df_E}=\frac{\text{SS}_E}{(n-p)}\right)$$

and the ratio

$$\left(F=\frac{\text{MS}_{\text{Reg}}}{\text{MS}_E}\right)$$

which follows an F distribution with $(p-1)$ and $(n-p)$ degrees of freedom. If the test is appropriate, we can test the null hypothesis (H_0) that β_i is equal to zero $(H_0:\beta_i=0)$ against the alternative hypothesis (H_1) that at least some β_i are not equal to zero $(H_1:$ at least some $\beta_i\neq 0)$. The ratio

$$\left(F=\frac{\text{MS}_{\text{Reg}}}{\text{MS}_E}\right)$$

is compared to the 100(1−α)% point of the tabulated $F_{(p-1,n-p)}$ distribution in order to determine whether or not the model is significant. It should be noted that just significant regressions may not predict well. Suppose we decide a specified risk level "α." The fact that the observed mean square ratio exceeds $F_{(p-1,n-p,1-\alpha)}$ means that a "statistically significant" regression has been obtained. In other words, the proportion of the variation in the data, which has been accounted for by the fitted equation, is deemed greater than would be expected by chance in similar sets of data with the same values of n and x. This does not necessarily mean that the fitted equation is useful for predictive purposes. Unless the range of values predicted by the fitted equation is considerably greater than the size of the random error prediction will often be of no value even though a "significant" F value has been obtained, since the equation will be "fitted to the errors" only. For the purpose of this paper, a ratio $(F_{\text{Calculated}}/F_{\text{Table}}\geqslant 2)$ is set as a criterion to reflect the so-called "descriptive model," while a ratio $(F_{\text{Calculated}}/F_{\text{Table}}\geqslant 4)$ reflects a "predictive model." Note that a predictive model is, in general, a better model than a descriptive one for estimating rainfall at ungauged locations (point estimates).

Extra Sum of Squares Principle

In this study, the stepwise regression was used. The stepwise regression procedure starts off by choosing an equation containing the single best predictor variable and then attempts to build the model up with subsequent additions of predictor variables one at a time as long as these additions are worthwhile. After a variable has been added, the equation is examined to see if any variable should be deleted. The extra sum of squares principle simply answers that question by considering the extra portion of the regression sum of squares, which arises due to the fact that the terms under consideration are in the model. The mean square derived from this extra sum of squares can then be compared with the estimate s^2 to see if it appears significantly large. If it does, the terms should have been included; if it does not, the terms would be judged unnecessary and could (though not necessarily) be removed. Generally, for the two models,

$$\text{Model 1}\ (E1):\quad y=\beta_1x_1+\cdots+\beta_qx_q+\varepsilon_1 \tag{3}$$

$$\text{Model 2}\ (E2):\quad y=\beta_1x_1+\cdots+\beta_qx_q+\beta_{q+1}x_{q+1}+\cdots+\beta_yx_y+\varepsilon_2 \tag{4}$$

there are a few extra terms in Model 2 which, if adequate, should be able to explain some of the residual sum of squares of Model 1, in this case, the null hypothesis $(H_0:\beta_{q+1}=\ldots=\beta_y=0)$ against the alternative hypothesis $(H_1:\beta_{q+1},\ldots,\beta_y\neq 0)$. The extra sum of squares is explained by $(\beta_{q+1},\ldots,\beta_y)$ is $(\text{SS}_{E1}-\text{SS}_{E2})$ with a $(y-q)$ degrees of freedom. If that holds, the ratio

$$\left(\frac{(\text{SS}_{E1}-\text{SS}_{E2})/(y-q)}{\text{SS}_{E2}/(n-y)}\right)$$

should be greater than, or equal to, the 100(1−α)% point of the tabulated $F_{(y-q,n-y)}$ distribution. For the purpose of this paper, α is assigned the value 0.05 (a value that is commonly used in statistical analysis).

R^2 Statistic

The coefficient of multiple determination (R^2) is often used as a convenient measure of the success of the regression equation in explaining the variation in the data. It is expressed as the percentage ratio of the SS caused by the regression to the total sum of squares. Ideally, the value of R^2 could be 100%, however, practically, it should not be no matter how good the model is. The R^2 value could increase in some cases where the extra sum of squares test disapproves the extra terms. This takes place when the degrees of freedom are consumed due to the increase in the number of parameters in the model, therefore approaching the saturation point, i.e., number of parameters $(p)\approx$sample size (n).

Case Study

The island of Crete, in Greece, was chosen for the purpose of this paper in order to study the spatial variation in precipitation in small and medium size mid-latitude areas, where orographic effects tend to increase both frequency and intensity of winter precipitation. The island of Crete occupies the southern part of the country of Greece (as shown in Fig. 1) with an area of 8,265 km^2, which is almost 6.3% of the area of Greece. Crete has a mean elevation of 482 m and an average slope of 228 m/km. The island

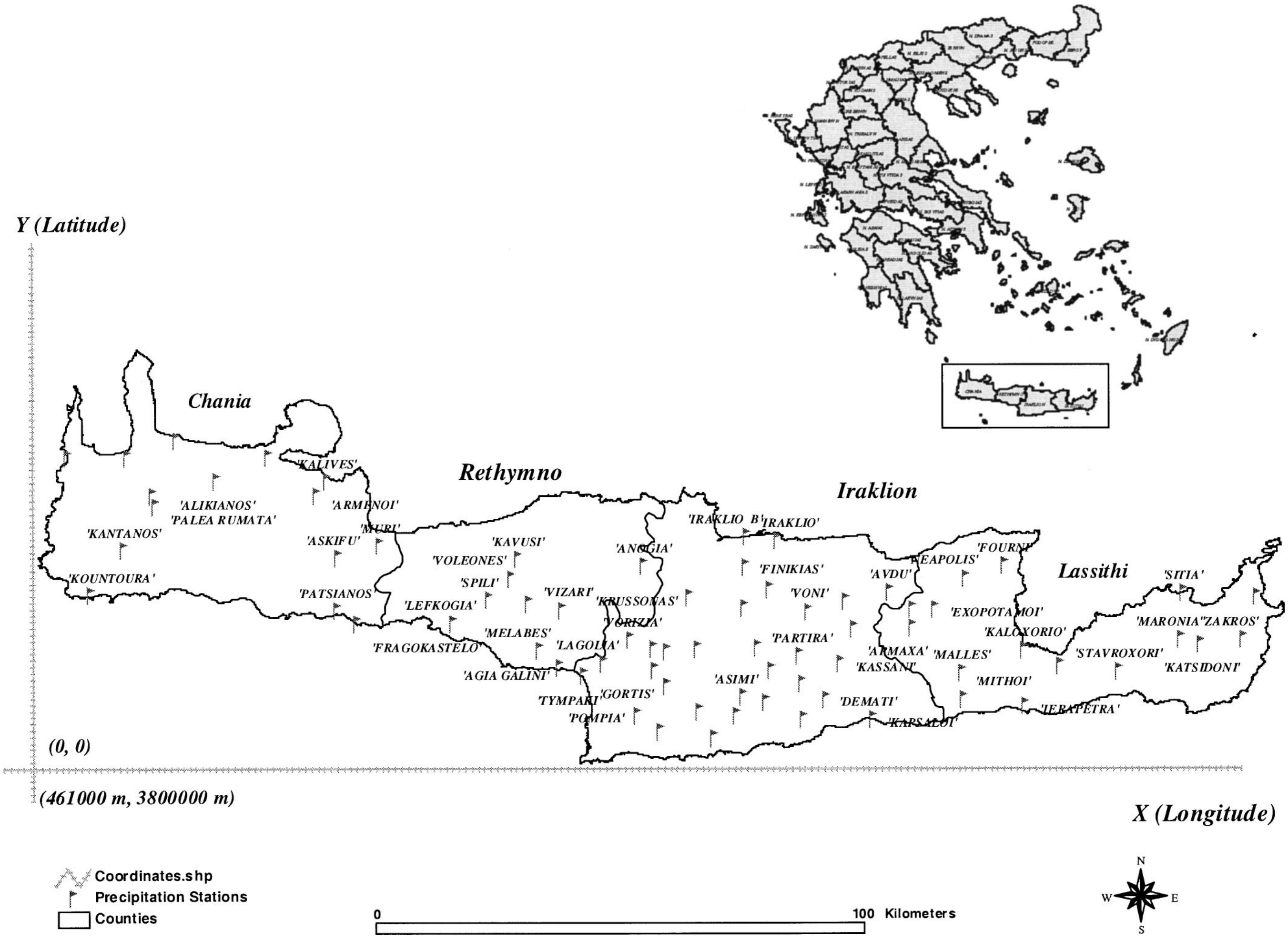


Fig. 1. Precipitation stations in four prefectures of Crete

is divided into four counties: Lassithi, Iraklion, Rethymno, and Chania. Most agricultural activity takes place in the county of Iraklion (2,626 km^2). The remaining counties, in order of agricultural activity, are: Lassithi (1,810 km^2), Chania (2,342 km^2), and Rethymno (1,487 km^2).

Monthly precipitation data (which is then aggregated to annual amounts) was compiled by the Hellenic National Meteorological Service for 74 precipitation stations representing the island (refer to Fig. 1). The gauges used are of the cylindrical weighing type, with a diameter of 10 cm. A small amount of oil is added to the gauge to prevent excessive evaporation losses. The location of these gauges was provided in degrees, minutes, and seconds (DMS) (geographic projection). The stations mainly cover the eastern part of the island, which has a higher level of agricultural and tourism activity than the western part. The counties of Iraklion, Lassithi, Chania and Rethymno were covered by 34, 16, 15, and 9 gauges, respectively. Six gauges (two in Iraklion, one in Lassithi, and three in Chania) recorded only for 4 or 5 years and were excluded from the analysis, while the rest recorded for 12–50 years. The gauges were located at elevations that ranged from mean sea level (MSL), in the county of Iraklion, to 905 m above MSL, in the county of Lassithi. The GIS coverages used in this study were based on maps developed by the Greek Army Geographical Service. The software *ArcView* GIS was used to create a small database to handle spatial data. The database included information for spot elevations for the island, shoreline, counties, location of stations (x,y,z), and monthly precipitation values. A GIS layer was generated for the precipitation gauges where their locations were identified (after conversion from DMS) by decimal degrees. The layers were then projected, using the Transverse Mercator projection and the GRS 80 Spheroid with a central meridian of 24°, scale factor of 0.9996, and false easting of 500,000. This projection is also referred to as EGSA units. A DEM, with a cell size of 30 m and a grid of 3,027 rows and 8,659 columns, was generated from the spot elevation values and a 20 m contour map. The DEM was used to check the elevation of the gauges and to delineate watersheds. As shown in Fig. 2, a chain of high elevations lies across the island, receiving fairly high amounts of precipitation. The precipitation and ensuing runoff are greater in the north than in the south because the prevailing wind direction is northwesterly. Generally, the western part of the island receives more precipitation than the eastern part. Using *ArcView* GIS, watersheds were then delineated using the DEM (as shown in Fig. 3) through a multistep procedure.

Analysis

Most natural phenomena such as rainfall vary continuously. The observations (collected data) that describe these phenomena represent a finite number (sample) of the infinity of possible locations (population). Although these processes (phenomena) are

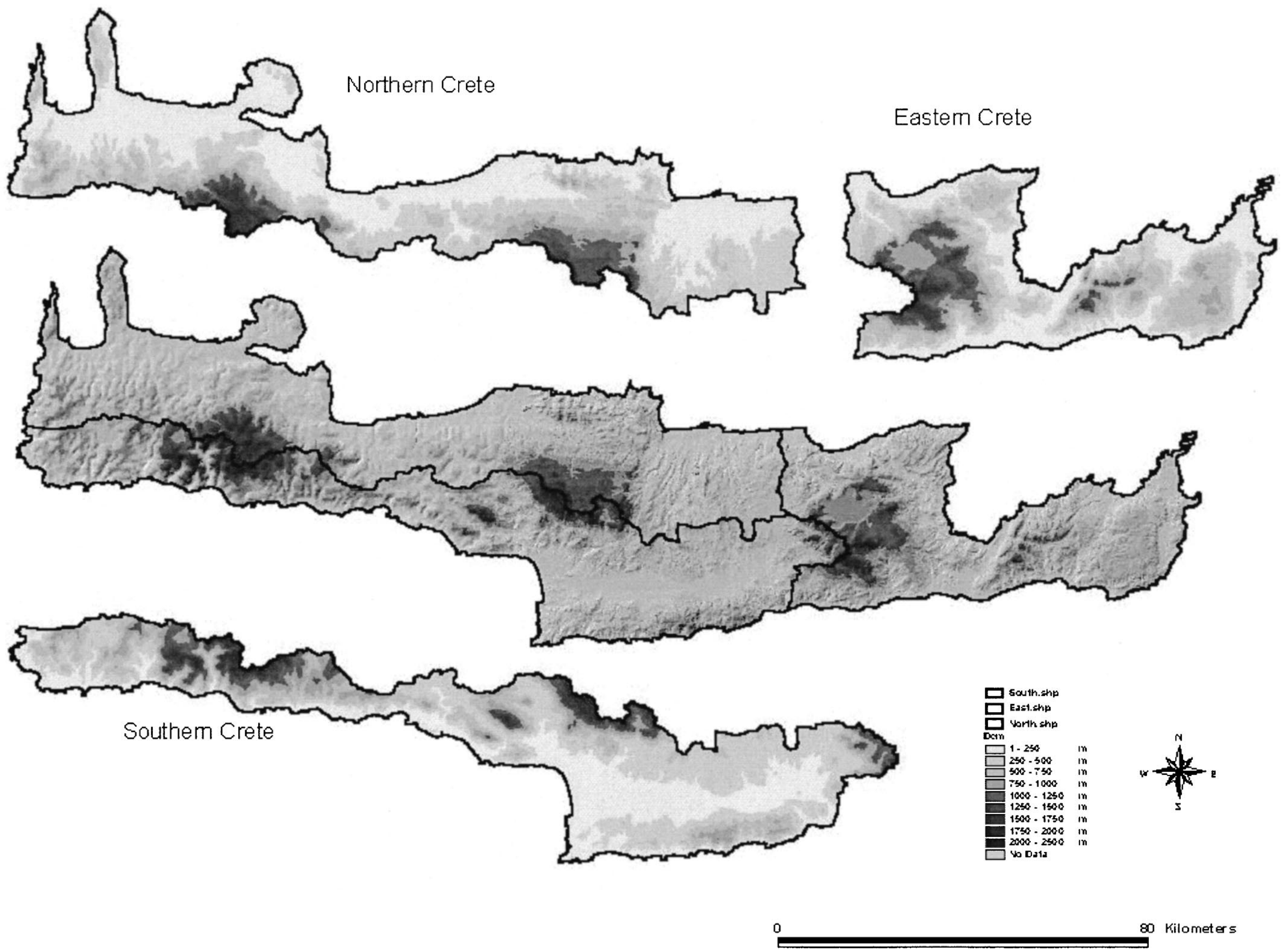


Fig. 2. Digital elevation model of Crete and its three divisions/regions

continuous and the values at sites that are close together in space are more likely to behave similarly than those farther apart, exceptions do exist. For instance, in the Geropotamou Basin (located in southern Crete, as shown in Fig. 3) the stations of Vorizia and Pompia record more rain than the neighboring stations of Zaros and Agios Kirillos, respectively. The stations Zaros and Agios Kirillos are considered "rain shadow" stations due to the nature and orientation of topography at that area and considering that the prevailing wind direction is northwesterly.

The annual records were used to study the variation in precipitation in time and space. Temporally, according to Naoum and Tsanis (2003), it was clear that the year 1977–1978 was the wettest, while the year 1989–1990 was the driest. Spatially, it was evident that precipitation is of an orographic type. In other words, on a small scale, precipitation can be well explained by the elevation of the gauge. However, additional independent variables were introduced in order to construct efficient models for relatively larger areas, as in the case of Gerapotamou Basin. For the average year 1974–1975, for example, a value of 43.6% for the coefficient of simple determination for simple regression (r^2) was obtained when using Eq. (5)

$$P=536.0+0.65E \quad 150\ \text{m}<E<517\ \text{m} \tag{5}$$

where P=precipitation (mm/year); and E=gauge elevation (m); while an R^2 value of 87.8% was then obtained according to Eq. (6)

$$P=-290.51+0.47E-3.42L_o+15.54L_a \tag{6}$$

where $L_0=x$ coordinate (longitude) (km); and $L_a=y$ coordinate (latitude) (km).

The two new variables introduced were the latitude and longitude of the gauges. The analysis in this paper is thus extended to cover the spatial variation of precipitation over the whole island, rather than small watersheds, by introducing additional predictor variables. The analysis is performed on two wet years (1977–1978 and 1986–1987), a dry year (1989–1990), an average year (1974–1975), a short-term average (1974–1975 to 1984–1985), and a long-term average (1969–1970 to 1994–1995).

In order for the aspect and orientation of topography factors to be considered, the island was divided into regions (divisions) for which the observations of the respective stations would reflect a general regional pattern. Four plots of latitude versus precipitation for the wet year of 1977–1978 were constructed, as shown in Fig. 4. Each of the plots represents a prefecture (county), starting from Chania west to Lassithi east. The horizontal axis covers 80 km south to north (i.e., latitude values start from 3,860,000 m south to 3,940,000 m north). Generally, and being of orographic type, the plots show similar behavior: precipitation is low at the south and north ends of the island and peaks in the middle. Unlike the other three prefectures, the prefecture of Lassithi showed a rather unique behavior (probably for the lack of representativeness of the gauges in that region or the less obvious orographic effects),

J. Hydrol. Eng. 2004.9:79-102.

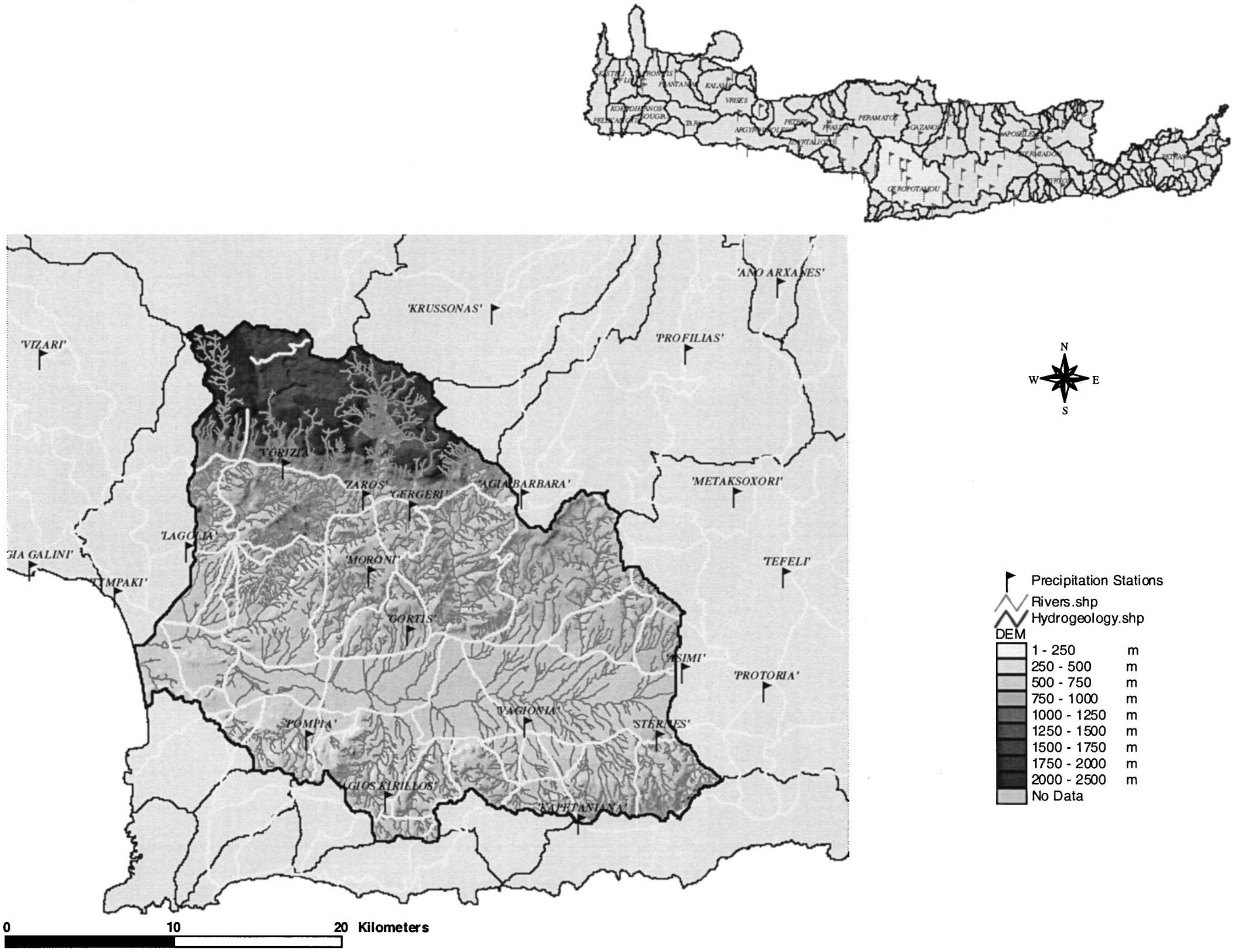


Fig. 3. Map of Gerapotamou Basin showing digital elevation model, roads, streams, and rain gauges

which suggested that the prefecture represents an independent division. Therefore, and as shown in Fig. 2, the island was divided into three divisions: namely ‘‘northern Crete,’’ ‘‘southern Crete,’’ and ‘‘eastern Crete.’’ This process is, as will be seen later, most favorable for studying the quantitative influence of the different environmental and geographical variables on the mean annual precipitation. Satisfactory results would not, normally, be obtained if the area studied was large or complex, since the factors vary inconsistently. The northern division occupies around 40% of the island and covers an area of 3,354 km^2. It has a mean elevation of 446 m and an average slope of 210 m/km. The southern division covers an area of 2,765 km^2 (34% of Crete), with a mean elevation of 536 m and an average slope of 250 m/km. The eastern part covers an area of 2,146 km^2 (26%), with a mean elevation of 471 m and an average slope of 230 m/km. The database and DEM are used to generate and apply the multiple regression models. To simplify calculations, a relative coordinate system was established by locating an origin (0,0) at the lower left corner of the island at latitude 3,800,000 m and longitude 461,000 m (refer to Fig. 1). Therefore, all Y coordinates employed in the regression were the result of subtracting 3,800,000 from the original latitudes of the different stations and dividing by 1,000 to obtain latitudes in km. Similarly, all X coordinates were the result of subtracting 461,000 from the original longitudes of the different stations and dividing by 1,000 to obtain longitude values in km. This manipulation of coordinates is important when performing the regression analysis. Using large numbers for latitude and longitude could have resulted in small values for model parameters (β_i) and any small error could have resulted in significant changes in the model output. It also is more practical in multiple regression to put all variables into approximately the same order of magnitude (i.e., so that elevation, longitude, and latitude values have similar order of magnitude) to ensure that they receive appropriate weighting in the multiple regression analysis.

Elevation was considered to be the primary predictor variable for the regression models and accounts for the topography factor. Fig. 5 indicates that there is a positive correlation between elevation and precipitation. The r^2 value ranged from 31% in the case of the dry year of 1989–1990 to 46% in the case of the wet year of 1977–1978. The two new variables (latitude and longitude) were then added to the models and the R^2 values were significantly increased, a possible sign of model improvement. Developing separate regression equations and still using elevation as the only predictor variable for the three divisions of Crete provided even more descriptive models, as shown in Fig. 6, which shows the wet year of 1977–1978 as an example. The three models of the divisions show a stronger positive elevation–precipitation correlation and were better than one model of the

J. Hydrol. Eng. 2004.9:79-102.

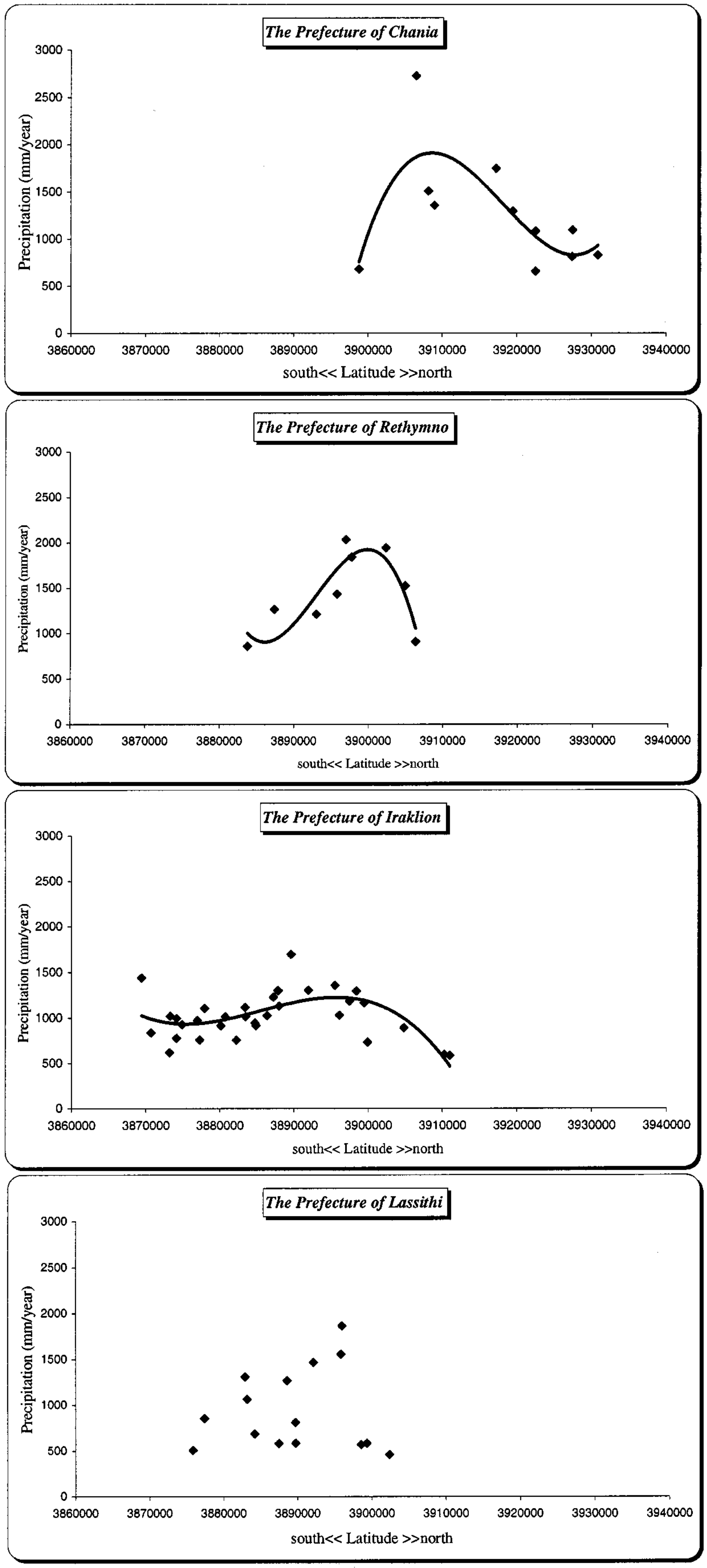


Fig. 4. Variation of precipitation with latitude in four prefectures

J. Hydrol. Eng. 2004.9:79-102.

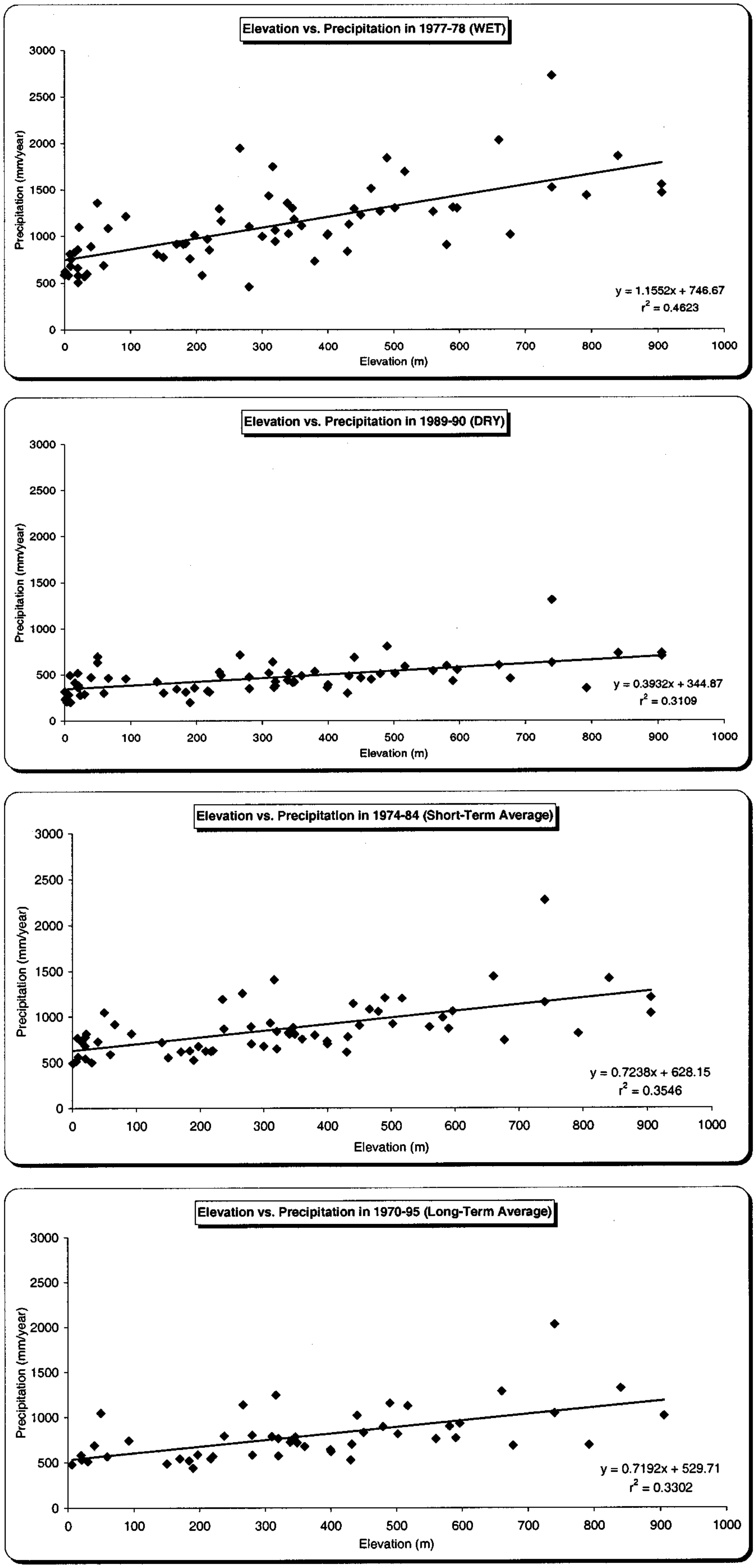


Fig. 5. Elevation–precipitation correlation for different data sets for Crete

J. Hydrol. Eng. 2004.9:79-102.

(a)

(b)

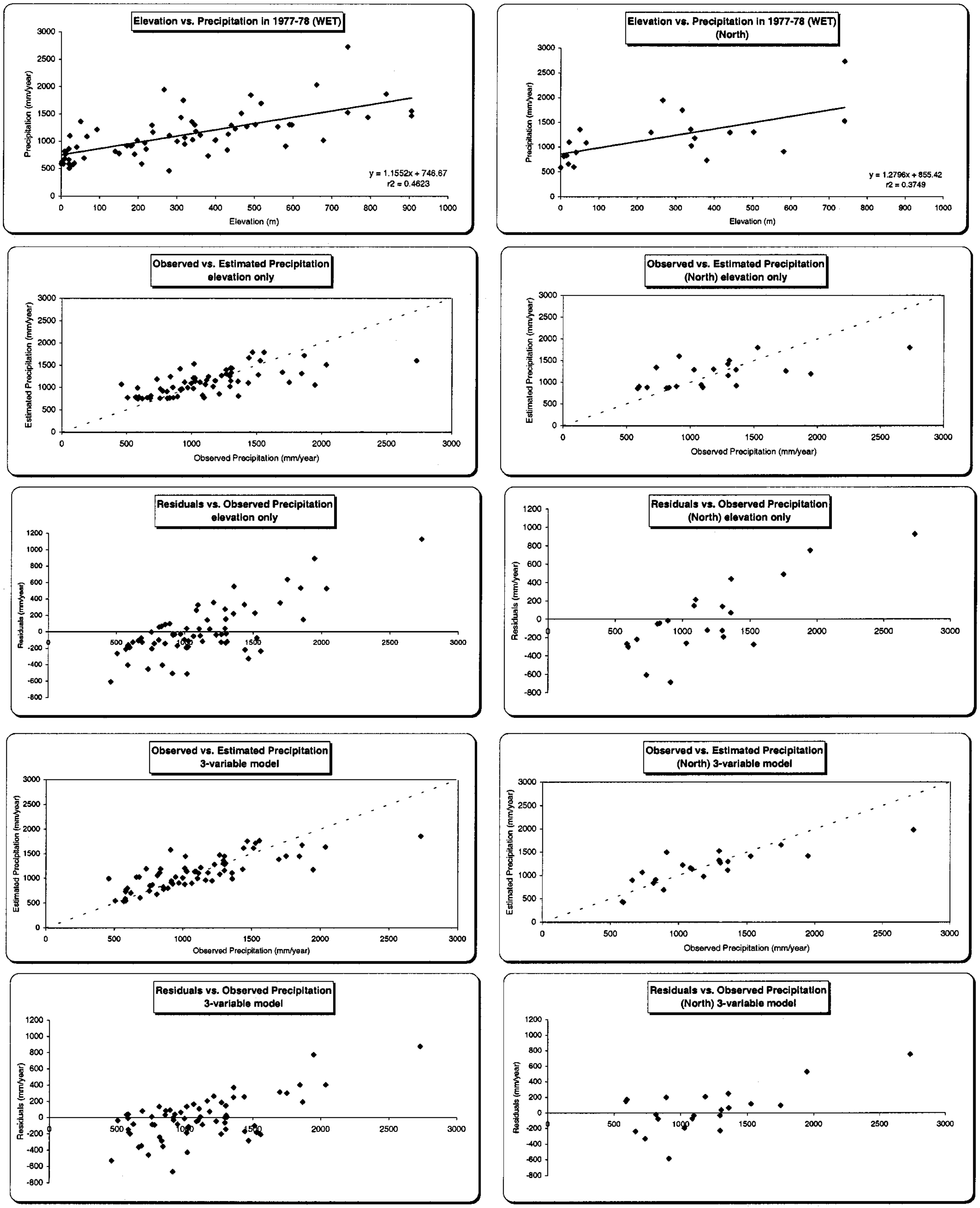


Fig. 6. Using water year of 1977–1978 as example, observed versus estimated precipitation plots as well as residual plots were generated for two cases of elevation only and three-variable models for: (a) Crete; (b) northern Crete; (c) southern Crete; and (d) eastern Crete

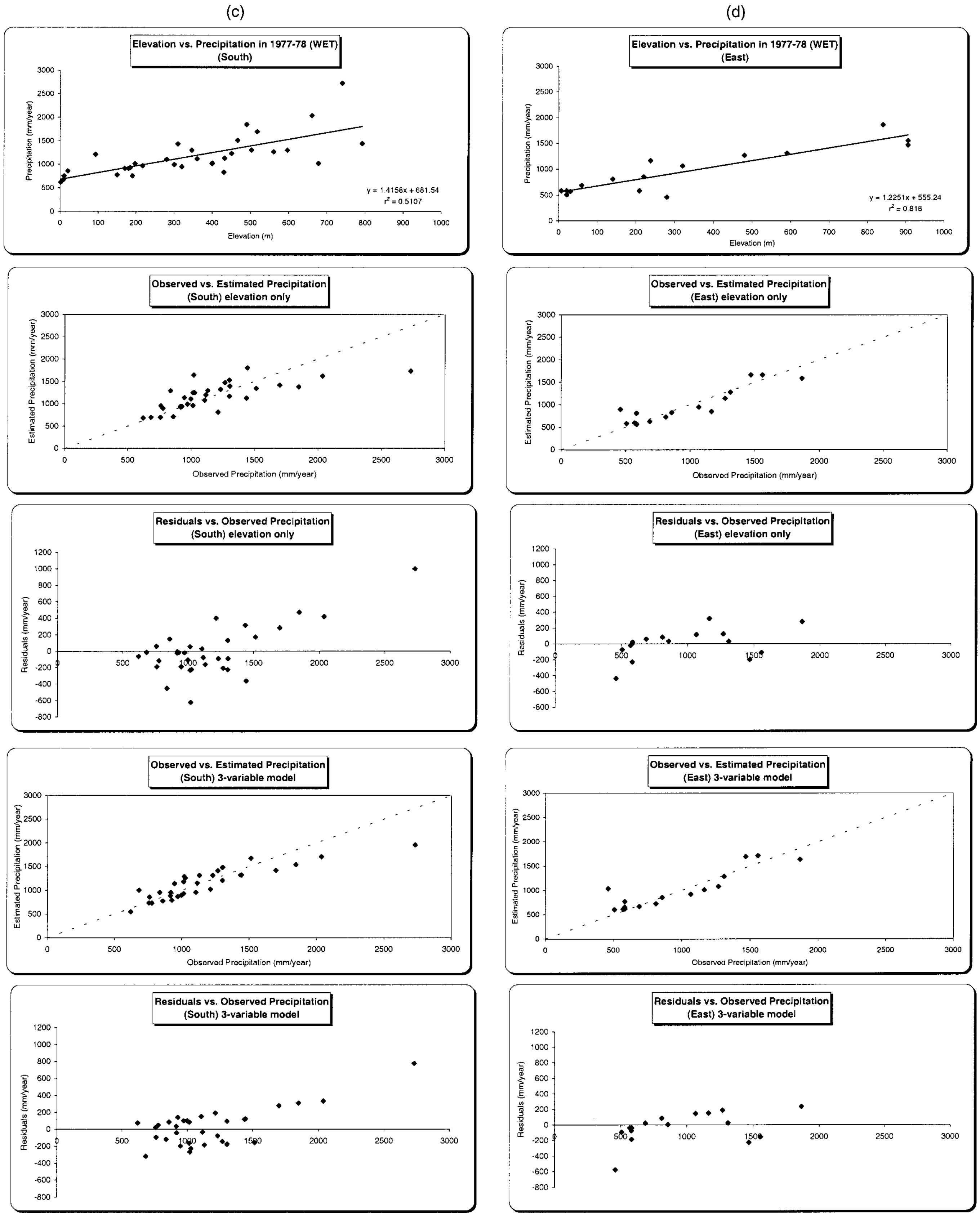


Fig. 6. (*Continued*)

J. Hydrol. Eng. 2004.9:79-102.

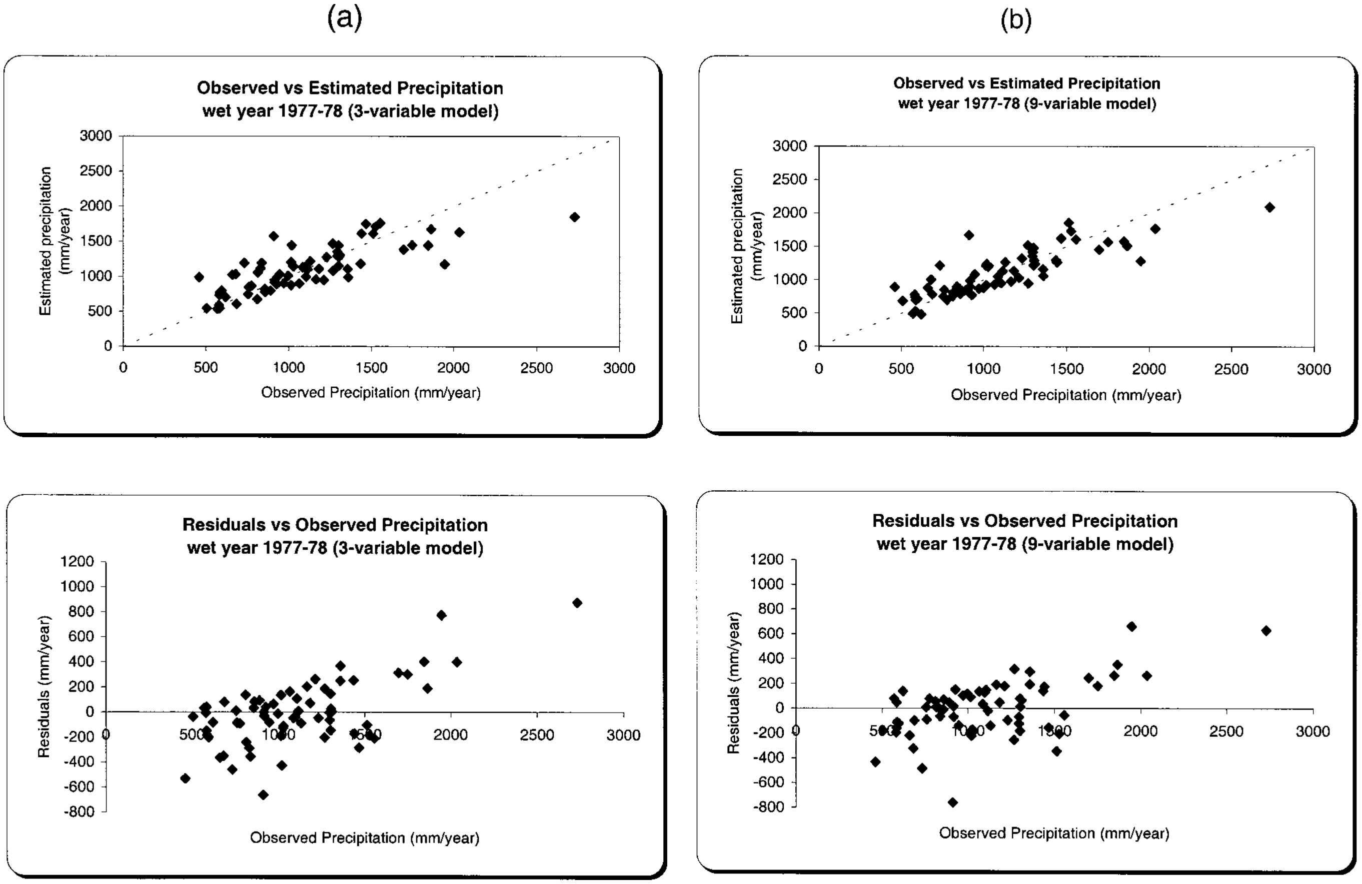


Fig. 7. For northern Crete, and based on year 1977–1978, comparison between output of: (a) four-parameter model and (b) ten-parameter model is illustrated

whole island. In the residuals versus observed precipitation plot, which represents the whole island case, there is a linear relation between observed precipitation and model residuals. This is an indication of one (or more) missing predictor variable(s) that could further describe the phenomenon. Residual plots should show no sign of association with the observed sample and must be purely random, as in the three plots of the three divisions.

The four parameters $\{\beta_0$ (intercept), β_1 (elevation), β_2 (longitude), and β_3 (latitude)$\}$, have been proposed and adopted to successfully establish multiple linear regression models. However, interaction effects between the variables can be considered. These would be represented in the models by adding three more terms. The use of a second-order model is also a good possibility. Second-order models with, as in this case, three predictor variables are particularly used in response surface studies, where it is desired to graduate, or approximate to, the characteristics of some unknown response surface by a polynomial of low order. Note that all possible second-order terms will be in the model, thus leading to ten-parameter models for Crete and its divisions for the different data sets. The general form of the final product is

$$P=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_1^2+\beta_5 x_2^2+\beta_6 x_3^2+\beta_7 x_1 x_2 + \beta_8 x_1 x_3+\beta_9 x_2 x_3 \quad (7)$$

where P=precipitation (mm/year); x_1=elevation (m); $x_2=x$ coordinate (longitude) (km); and $x_3=y$ coordinate (latitude) (km). Although the models improved by introducing the interaction and second-order terms by observing the R^2 statistic value in each of these cases, this improvement took place at the expense of the loss of degrees of freedom. This, in turn, raises the question: was it worthwhile adding the new terms to the models?

Results and Discussion

Large-Scale (Island and Its Three Regions)

Regression analysis results show that for the eastern part of Crete, where the number of gauges is relatively small, using the four-parameter (three-variable) models is more feasible. This is especially true if there is no noticeable improvement in the coefficient of determination and model significance. As for the northern and southern parts of the island, the coefficient of determination increased significantly (a 34% increase was reported in the case of the wet year 1986–1987 for northern Crete). Fig. 7 compares the four-parameter model output and the ten-parameter model output for Northern Crete for the year 1977–1978. Notice the improvement in the estimated values as the residuals tend to scatter over a smaller range around the x axis for the second case. The estimate for the station of Askifu (an extreme observation) improved from 1,824 mm (using the four-parameter model) to 2,386 mm (using the ten-parameter) model, thus approaching the observed value of 2,594 mm. Results for the whole island show that the increase in R^2 ranged from 7 to 14%, which suggests that using the ten-parameter model rather than the four-parameter model for Crete can be regarded as a justifiable gain with a moderate loss in degrees of freedom.

Table 1. Different Multiple Linear Regression Models for Crete and Its Divisions for Year 1974–1975

Data set	Spatial extent	No. of Stations	Boundary conditions																	Is the model significant as a ... model?	
			Elev	Longitude		Latitude															
			($X1$)	($X2$)	Long (m)	($X3$)	Lat (m)	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	R_2 (%)		Descriptive	Predictive
1974–1975 (average year)	Crete	66	$6 \leqslant E(X1) \leqslant 905$ m	$4 \leqslant X2 \leqslant 244$ km	$464{,}917 \leqslant \mathrm{Lo} \leqslant 705{,}496$	$69 \leqslant X3 \leqslant 131$ km	$3{,}869{,}421 \leqslant \mathrm{La} \leqslant 3{,}930{,}8$	50.6	0.74	−0.84	6.04	—	—	—	—	—	—	68.4		YES	YES
								−8.4	0.92	−0.80	6.39	−0.0002	—	—	—	—	—	68.8		YES	YES
								−33.2	0.92	−0.56	6.54	−0.0002	−0.0009	—	—	—	—	68.8		YES	YES
								−1,637.8	0.75	−1.43	43.71	−0.0001	0.0008	−0.1968	—	—	—	72.2		YES	YES
								−1,315.4	1.11	−0.24	33.89	0.0001	−0.0003	−0.1437	−0.0037	—	—	75.6		YES	YES
								−153.7	−0.17	−0.25	12.09	0.0001	−0.0006	−0.0448	−0.0027	0.0127	—	77.1		YES	YES
								−3,766.0	0.27	11.75	73.12	0.0003	−0.0085	−0.2979	−0.0036	0.0072	−0.1027	79.1		YES	YES
	North	19	$8 \leqslant E(X1) \leqslant 740$ m	$13 \leqslant X2 \leqslant 153$ km	$473{,}829 \leqslant \mathrm{Lo} \leqslant 613{,}595$	$88 \leqslant X3 \leqslant 131$ km	$3{,}887{,}874 \leqslant \mathrm{La} \leqslant 3{,}930{,}8$	3,458.1	0.62	−6.39	−20.18	—	—	—	—	—	—	82.9		YES	YES
								4,144.8	−0.10	−7.00	−25.26	0.0009	—	—	—	—	—	84.7		YES	YES
								3,814.0	0.08	−3.85	−23.30	0.0006	−0.0170	—	—	—	—	85.3		YES	YES
								−1,594.0	−0.22	−7.49	83.09	0.0010	0.0004	−0.4997	—	—	—	90.7		YES	YES
								1,910.0	0.70	−3.35	9.70	0.0006	−0.0107	−0.1475	−0.0070	—	—	93.1		YES	YES
								11,135.0	−5.33	−1.89	−147.10	0.0011	−0.0205	0.5118	−0.0055	0.0518	—	93.8		YES	YES
								13,764.0	−5.97	−9.05	−188.20	0.0011	−0.0153	0.6720	−0.0047	0.0571	0.0554	93.8		YES	YES
	South	30	$9 \leqslant E(X1) \leqslant 792$ m	$4 \leqslant X2 \leqslant 162$ km	$464{,}917 \leqslant \mathrm{Lo} \leqslant 622{,}750$	$69 \leqslant X3 \leqslant 108$ km	$3{,}869{,}421 \leqslant \mathrm{La} \leqslant 3{,}908{,}1$	−543.6	0.62	0.38	11.60	—	—	—	—	—	—	72.5		YES	YES
								−605.9	0.44	0.59	12.32	0.0002	—	—	—	—	—	72.8		YES	YES
								−915.1	0.43	6.41	13.63	0.0002	−0.0318	—	—	—	—	83.6		YES	YES
								1,223.0	0.26	8.59	−42.08	0.0003	−0.0381	0.3394	—	—	—	84.8		YES	YES
								449.0	0.40	7.96	−22.76	0.0005	−0.0332	0.2217	−0.0019	—	—	85.1		YES	YES
								2,440.0	−3.20	5.86	−50.15	0.0006	−0.0341	0.2981	0.0046	0.0337	—	88.7		YES	YES
								−1,763.0	−4.66	33.01	15.11	0.0006	−0.0664	0.0669	0.0102	0.0438	−0.2433	90.2		*YES*	YES
	East	18	$6 \leqslant E(X1) \leqslant 905$ m	$160 \leqslant X2 \leqslant 244$ km	$621{,}158 \leqslant \mathrm{Lo} \leqslant 705{,}496$	$76 \leqslant X3 \leqslant 102$ km	$3{,}875{,}851 \leqslant \mathrm{La} \leqslant 3{,}902{,}3$	364.3	0.59	−1.15	3.98	—	—	—	—	—	—	88.3		YES	YES
								393.3	0.49	−1.20	3.90	0.0001	—	—	—	—	—	88.5		YES	YES
								−1,754.0	0.59	18.21	5.95	0.0000	−0.0475	—	—	—	—	89.5		YES	YES
								−1,433.0	0.60	17.88	−0.59	0.0000	−0.0466	0.0363	—	—	—	89.5		YES	YES
								−2,300.0	1.47	24.99	0.33	−0.0002	−0.0615	0.0301	−0.0037	—	—	89.8		YES	NO
								−2,983.0	1.00	29.90	4.72	−0.0004	−0.0713	−0.0088	−0.0061	0.0122	—	90.2		YES	NO
								−3,118.0	1.10	30.96	5.15	−0.0004	−0.0707	0.0044	−0.0062	0.0114	−0.0138	90.2		YES	NO

J. Hydrol. Eng. 2004.9:79-102.

Table 2. Three-Variable Multiple Linear Regression Models for Crete and Its Divisions for Different Years

Data set	Spatial extent	Number of stations	β_0	β_1	β_2	β_3	R_2 (%)	Conditional Average rain (mm)	Conditional Total rain (giga M3)	Unconditional Average rain (mm)	Unconditional Total rain (giga M3)
1977–1978	Crete	67	746.67	1.16	—	—	46.2	1,360.9	11.27	1,450.1	12.01
(Wet year)			655.80	1.24	−2.15	3.76	61.8	1,437.9	11.91	1,533.6	12.70
	North	20	5,761.00	0.74	−9.94	−35.43	63.4	1,365.8	4.56	1,421.0	4.74
	South	30	−978.30	1.05	0.62	19.91	71.9	1,464.8	3.94	1,558.7	4.20
	East	18	669.50	1.09	−2.28	4.71	89.6	1,236.4	2.63	1,304.7	2.78
1986–1987	Crete	65	809.00	0.57	—	—	13.5	1,109.4	9.19	1,153.0	9.55
(Wet year)			−1,121.20	0.88	0.82	18.17	53.1	1,238.9	10.26	1,306.9	10.82
	North	21	3,763.00	0.75	−7.27	−19.58	47.7	1,361.9	4.54	1,417.7	4.73
	South	28	−1,340.50	0.73	0.42	21.45	65.7	1,038.5	2.80	1,103.9	2.97
	East	17	−920.40	0.67	1.67	15.36	61.2	1,164.2	2.48	1,206.2	2.57
1989–1990	Crete	64	344.87	0.39	—	—	31.1	553.9	4.59	584.3	4.84
(Dry year)			−240.80	0.49	−0.01	5.89	54.6	601.2	4.98	639.0	5.30
	North	21	1,505.00	0.46	−2.43	−7.79	46.6	666.1	2.22	700.4	2.34
	South	27	−681.60	0.32	1.10	10.07	61.2	509.0	1.37	537.7	1.45
	East	17	−122.50	0.44	0.03	4.57	90.2	536.5	1.14	564.1	1.20
1974–1975	Crete	66	536.00	0.65	—	—	43.6	882.1	7.31	932.4	7.72
(Average year)			50.60	0.74	−0.84	6.04	68.4	944.2	7.82	1,001.3	8.29
	North	19	3,458.10	0.62	−6.39	−20.18	82.9	995.1	3.32	1,041.2	3.47
	South	30	−543.60	0.62	0.38	11.60	72.5	886.3	2.39	941.8	2.54
	East	18	364.30	0.59	−1.15	3.98	88.3	816.8	1.74	853.8	1.82
Short-term	Crete	62	628.15	0.72	—	—	35.5	1,013.0	8.39	1,068.9	8.85
average			−88.00	0.90	−0.92	8.35	64.3	1,109.8	9.19	1,179.3	9.77
	North	18	3,949.00	0.89	−8.31	−22.68	70.4	1,187.0	3.96	1,253.3	4.18
	South	28	−738.50	0.66	0.24	14.95	74.6	1,000.9	2.70	1,060.0	2.85
	East	17	85.00	0.75	−0.02	5.17	83.4	958.0	2.04	1,005.1	2.14
Long-term	Crete	49	529.71	0.72	—	—	33.0	912.0	7.55	967.6	8.01
average			−516.50	0.66	−1.50	14.16	74.0	1,061.2	8.79	1,112.2	9.21
	North	11	2,504.00	0.62	−7.32	−9.32	67.9	1,185.4	3.95	1,231.6	4.11
	South	25	−503.10	0.60	−2.18	15.07	83.4	980.3	2.64	1,034.0	2.78
	East	14	−83.50	0.72	0.29	5.78	84.4	890.0	1.90	935.1	1.99

From a statistical perspective, the different statistical diagnostic tests revealed that the effect of adding new terms to the models varied greatly. Using the southern Crete model as an example, adding the term "x_2^2" had almost no effect on the model when applied to the long-term average (1969–1970 to 1994–1995) data, but the term did have a noticeable effect on the model when applied to the other datasets. Adding the term "$x_1 x_2$" had moderate effects in the case of all northern Crete models. Generally, the parameter "β_4" (associated with the variable "x_1^2") was classified as "not significant" for all models in all cases. Also, the loss of degrees of freedom is very critical. For example, results of the ten-parameter models in the case of eastern Crete cannot be considered reliable, since the loss of degrees of freedom is great when compared to the sample size (number of stations). Generally, the analysis results supported the significance of neither the majority of ten parameter models for eastern Crete nor, in the case of long-term average, the ten parameter model for northern Crete. These models proved to be insignificant and, according to the "extra sum of squares" principle, the extra terms were not needed in spite of the high R^2 value, which is deceptive in this case. Adding the terms did not exhibit any cost problems. In other words, adding the interaction and second-order terms did not impose any additional efforts for collecting new data and the high cost that is normally associated with it. Only available data were used in a different form. However, losing degrees of freedom due to the new terms indirectly suggests an increase in the number of precipitation stations which would be needed to enhance the models, specifically in eastern Crete and the far western part of the island.

To summarize, the ten-parameter models can be used with care, especially when they are derived from not-a-dense-enough network of gauges. A sample of the different models is presented in Table 1 for the average year of 1974–1975. The models are also classified as "descriptive" or "predictive."

Using the GIS module (AVRU) (Naoum and Tsanis 2004), estimates for average and total volume of rainfall for Crete and its three regions for the different cases were obtained, as shown in Table 2. The four-parameter models were used for Crete, as well as the three divisions, while elevation-only models were used for Crete for comparison purposes. Moving from the one variable model to the three variable model resulted in a noticeable improvement in R^2 for Crete in all cases. Knowing that the highest

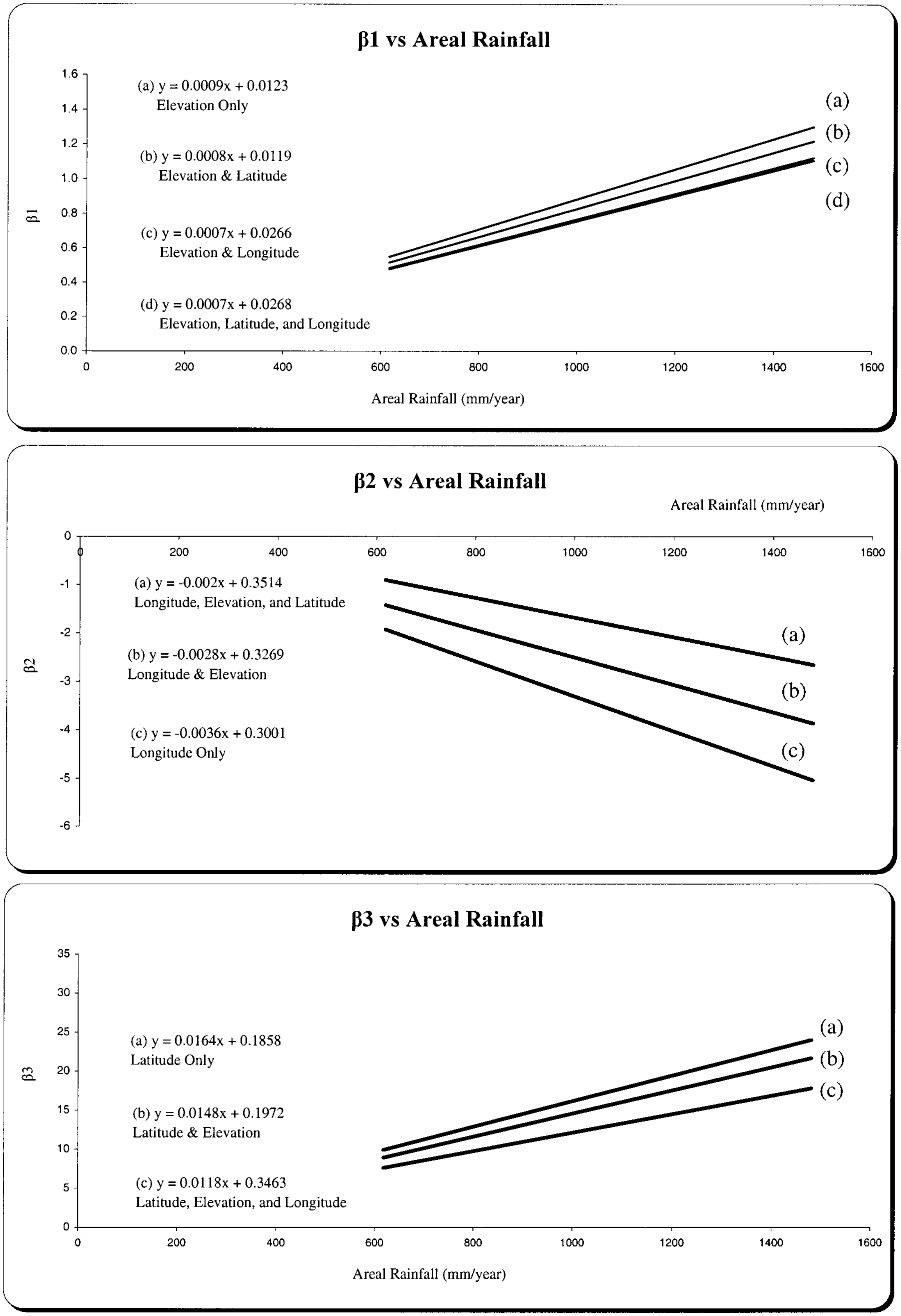


Fig. 8. Relation between different parameters and areal rain

Table 3. Examining Individual Influence of Different Predictor Variables on Estimating Average and Total Rain for Years 1974–1975

Spatial extent	Number of stations	β_0	β_1	β_2	β_3	R_2 (%)	Conditional Average rain (mm)	Conditional Total rain (giga M3)
Crete ($x1$)	66	536.00	0.65	—	—	43.6	882.10	7.31
Crete ($x1$,$x2$)		735.99	0.69	−1.65	—	59.5	913.34	7.56
Crete ($x1$,$x3$)		−234.72	0.74	—	7.92	65.6	945.02	7.83
Crete ($x2$,$x3$)		527.07	—	−0.86	3.50	13.9	773.22	6.40
Crete ($x1$,$x2$,$x3$)		50.60	0.74	−0.84	6.04	68.4	944.20	7.82
North ($x1$)	19	651.83	0.79	—	—	47.4	1,042.07	3.48
North ($x1$,$x2$)		843.27	1.00	−2.95	—	71.4	1,103.79	3.68
North ($x1$,$x3$)		−243.41	1.03	—	7.55	53.5	1,109.93	3.70
North ($x2$,$x3$)		5,438.31	—	−8.36	−35.12	69.7	837.06	2.79
North ($x1$,$x2$,$x3$)		3,458.10	0.62	−6.39	−20.18	82.9	995.10	3.32
South ($x1$)	30	472.85	0.65	—	—	46.2	844.14	2.27
South ($x1$,$x2$)		647.08	0.70	−1.64	—	56.0	891.75	2.40
South ($x1$,$x3$)		−416.86	0.63	—	10.57	72.3	890.41	2.40
South ($x2$,$x3$)		−806.53	—	1.63	15.41	34.0	730.57	1.97
South ($x1$,$x2$,$x3$)		−543.60	0.62	0.38	11.60	72.5	886.30	2.39
East ($x1$)	18	472.37	0.66	—	—	84.7	834.39	1.78
East ($x1$,$x2$)		718.93	0.61	−1.15	—	86.4	820.91	1.75
East ($x1$,$x3$)		117.67	0.64	—	3.98	86.7	828.35	1.76
East ($x2$,$x3$)		928.06	—	−4.37	6.98	40.0	684.66	1.46
East ($x1$,$x2$,$x3$)		364.30	0.59	−1.15	3.98	88.3	816.80	1.74

rain gauge (the Germiado gauge in the prefecture of Lassithi) is installed at an elevation of 905.4 m, two values for average and total rainfall were estimated. The first value is the ‘‘conditional estimate,’’ which is based on the assumption that elevations higher than the 905.4 m value are considered 905.4 m when using the MLR models. The ‘‘unconditional estimate’’ is obtained by using elevation values as they are. The relatively larger influence of elevation on rainfall estimates, in addition to the lack of information about precipitation behavior above the 905.4 m level, were the reasons behind the differentiation. In other words, although the models were derived based on information obtained from gauges installed on elevations of 0–905.4 m, they still are considered applicable on elevations higher than the 905.4 m limit. According to the generated DEM, about 15% of the elevations are above the 905.4 m level. The mean elevation above 905.4 m is 1,310 m. The GIS module provides the user with the choice of a conditional or unconditional estimate based on a DEM of a resolution of 1 km×1 km. The results show that the difference between conditional and unconditional estimates is obviously caused by the number of cells with elevations that are greater than 905.4 m in the DEM. The unconditional volume estimates are, on average, higher by 4.6, 5.9, and 4.9% for northern, southern, and eastern Crete, respectively. For the whole island, there is also an increase of 5.5% using elevation-only models, and 5.9% using three-variable models. Results also show that the north receives approximately 45% of the total rain, while the south receives 31%, and the east 24%.

Based on the parameter significance analysis alone, new models were derived to examine the effect of eliminated (insignificant) parameters on the results. The results show that, although some models provided comparable results to the original models for estimating areal rainfall, they failed to provide reliable models for northern Crete for the different cases. These models are not recommended for point (i,j) estimates (which is regarded as the strength of the MLR models) due to the fact that negative results were obtained at some locations.

Upon examining the individual influence of the different variables on the models, it was found that due to the unique shape of the island and the new coordinate system (refer to Fig. 1) high values for the latitude parameter (β_3) as well as high negative values for the parameter (β_0) are obtained. The models adjust to this condition in the western part by introducing positive values for (β_3), adding a large negative number for (β_0) to equalize that effect. This was clearly noticed in the case of the wet year (1986–1987) models, where latitude is a key variable in explaining rainfall for that year. Two types of wet years can be distinguished for Crete. One type can be referred to as ‘‘conservative,’’ as in the case of 1977–1978, since it conserves the elevation–precipitation correlation. In other words, the increase in precipitation amounts collected is homogeneously distributed and correlated over the island with the variation in elevation of the recording gauge. An r^2 of 46% is obtained and precipitation–elevation gradient (slope) of 1.2 mm/m is estimated. The second type can be referred to as ‘‘aspect-driven,’’ as in the case of 1986–1987, due to the fact that the gauges in northern Crete recorded almost double the precipitation of those in southern Crete. An r^2 of 14% for the elevation–precipitation correlation was obtained and precipitation–elevation gradient of 0.6 mm/m was estimated.

Based on results reported by Naoum and Tsanis (2003), for a typical dry year, the island of Crete would receive up to 800 mm of rainfall. For an average year, between 800 and 1,100 mm and for a wet year the island would receive a rainfall that is greater than 1,100 mm. Based on these values, and as shown in Fig. 8, the range of elevation–precipitation gradient (β_1) for a dry year is 0.6–0.75 mm/m, rising to 1.0 mm/m for an average year and up to 1.4 mm/m for a typical wet year. The range of longitude-precipitation gradient (β_2) for a dry year is −1.9 to −2.6 mm/km, up to −3.6 mm/km for an average year and −5.1 mm/km for a typical wet year. The range of latitude-precipitation gradient (β_3) for a dry year is 10–13.3 mm/km, rising to 18.0 mm/km in an average year and up to 25.0 mm/km in a typical wet year. Details for wet and dry years are shown in the following:

wet years	1967–1968	1968–1969	1975–1976	1977–1978	1980–1981	1981–1982	1984–1985	1986–1987
areal rainfall (mm)	1,206	1,346	1,211	1,480	1,163	1,185	1,174	1,285
dry years	1969–1970	1973–1974	1976–1977	1985–1986	1989–1990	1992–1993		
areal rainfall (mm)	791	778	827	736	618	769		

Fig. 8 also shows the degree of interaction between the parameters (i.e., the influence of a particular parameter decreases when additional parameters are added to the model). For example, introducing latitude in addition to elevation reduces the parameter (β_1). For a certain value of areal rainfall, (β_2) is at its lowest value when the elevation and latitude variables are added to the model. It is at its highest value when only longitude models are used.

Table 3 shows the influence of each individual variable on the model output for an average year (1974–1975) for Crete and its divisions when employing elevation only; elevation and longitude; elevation and latitude; longitude and latitude; and elevation, longitude, and latitude. The results show that there is a general increase in the coefficient of determination from the one-variable model to the three-variable model. Although the three-variable model provided more reasonable estimates, the models based on elevation and longitude provided a more physically meaningfulinterpretation of the effect of each variable. For example, looking at the model of Crete, an average rain (β_0) of 736 mm is expected over the island, with an increase of 69 mm/100m of elevation and a decrease of 1.65 mm/km of longitude (i.e., from west to east). Very poor R^2 values are obtained when using only longitude and latitude models.

To compare the different interpolation techniques, the so-called ''fictitious-point'' method (Delhomme 1978; Tabios III and Salas 1985) has been used. This is done by suppressing one sampling point (station) and values for that point are interpolated based on the remaining ($n-1$) points. Then, the interpolated values are compared with those observed for that point. In this case the MLR models, as a spatial interpolation technique, are consid

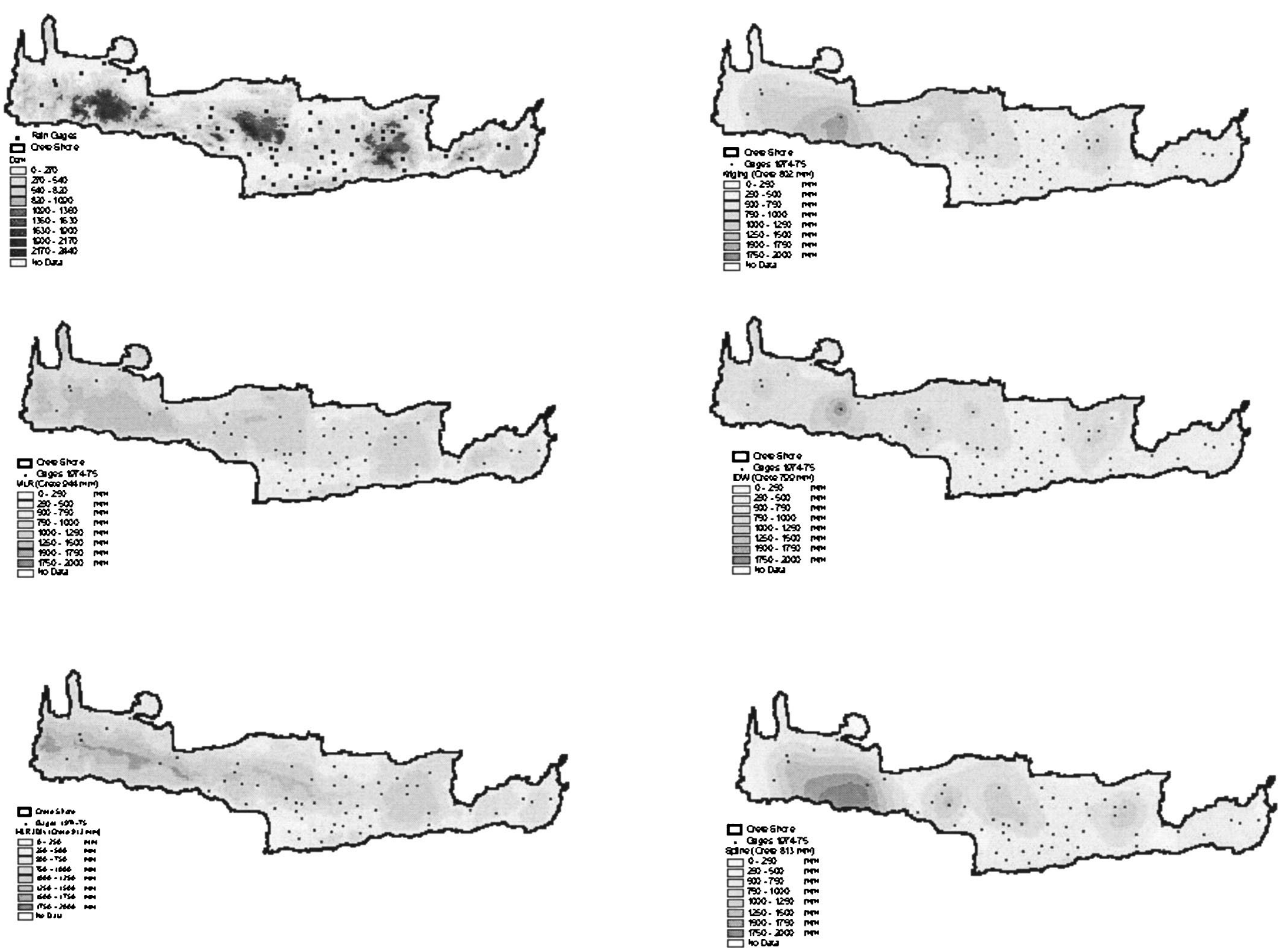


Fig. 9. Comparison between multiple linear regression models and spatial interpolation techniques

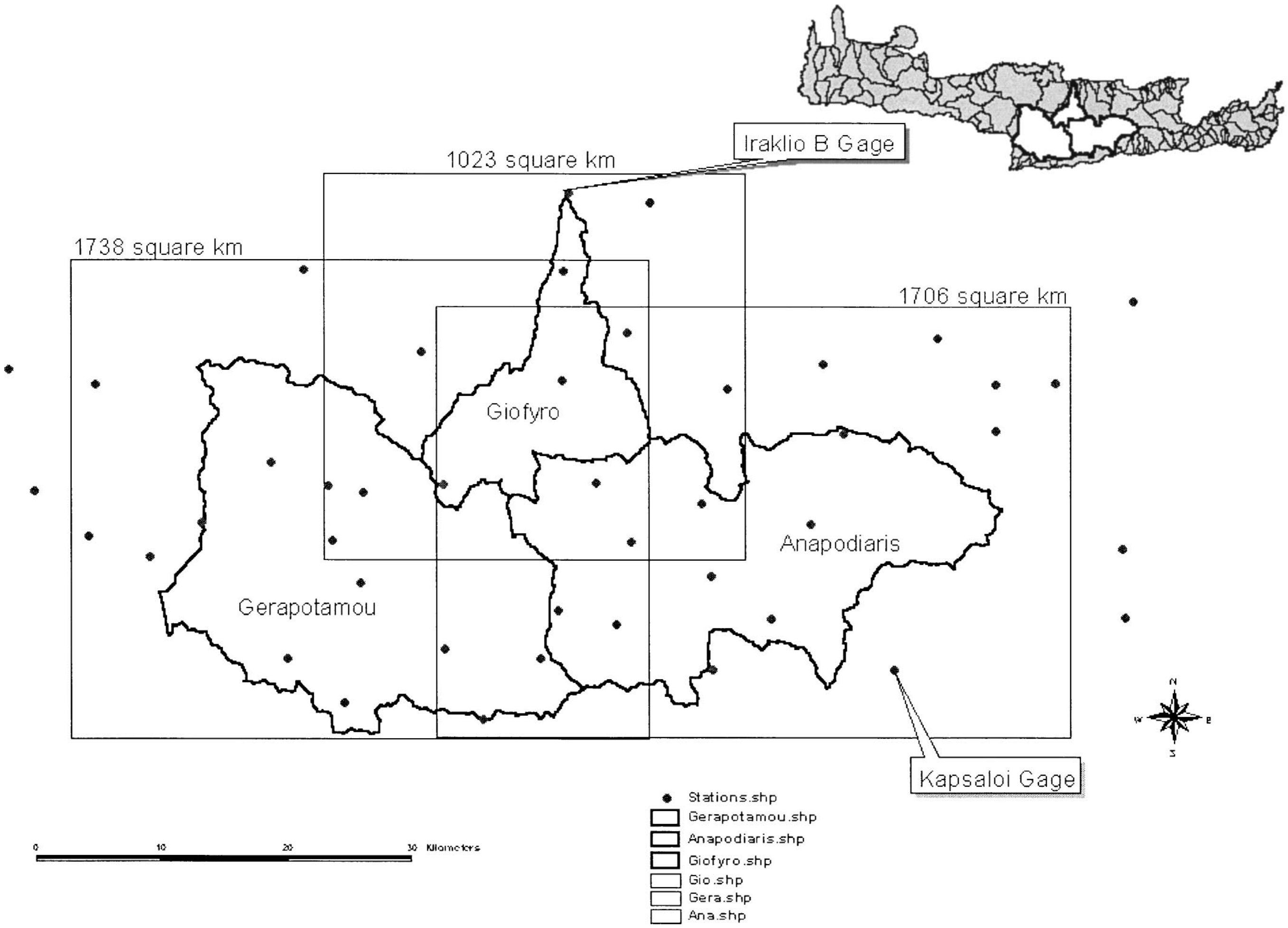


Fig. 10. Illustration of spatial extents of basins and their respective rain gauges

ered superior to the conventional spatial interpolation techniques such as Spline, IDW, and kriging. Fig. 9 illustrates the difference in average estimates according to the different methods as well as the spatial distribution of rainfall across the island. The estimates ranged from 799 mm, using IDW, to 944 mm using MLR. The spatial distribution of the MLR for three divisions is considered to be a reasonable one, yielding a mean annual precipitation of 913 mm for the average year 1974–1975.

Medium Scale (Watersheds)

As previously indicated in "Analysis," introducing latitude and longitude in addition to elevation as independent variables improved the estimates when studying the Gerapotamou Basin. Three basins were used as test cases to compare the MLR model results with the spatial interpolation techniques and to examine the effect of spatial extent on the output. The selected basins were used to represent different divisions of Crete with relatively dense rain gauge networks. As shown in Fig. 10, the basins are Gerapotamou (with nine gauges and area of 602 km^2) and Anapodiaris (with nine gauges and area of 520 km^2), which both represent the south, and Giofyro (with three gauges and area of 189 km^2), which represents the north. No basins represented the far east or the far west due to the few gauges available. A comparable analysis was held by using four different MLR models derived from data sets representing: the whole island (Crete), the division where the basin is located, the gauges located inside the basin, and the gauges located inside and outside the basin. The models were applied to the basins to estimate conditional and unconditional areal rain and rain volume. For Gerapotamou Basin, the larger extent is repre-sented by the rectangle in Fig. 10, with an area of 1,738 km^2 including 24 gauges; for Anapodiaris Basin, the larger extent is of an area of 1,706 km^2, containing 23 gauges (excluding the Kapsaloi gauge); and the larger extent for Giofyro Basin occupies an area of 1,023 km^2 surrounding 14 gauges (excluding the Iraklio B gauge). The results are summarized in Table 4. As in the case of larger scale analysis, the models based on elevation and longitude mostly provide more physically meaningful interpretations of the effect of each variable. The precipitation–elevation gradient is higher in the north than in both south and east. Both conditional and unconditional estimates for Giofyro Basin are the same because the highest elevation in the basin is 869 m, which is less than the 905.4 m limit. The difference between the lowest and highest estimates for average rainfall for the three basins ranged from 13 to 28%, with the values estimated by the three-variable models derived from the larger extent are in an intermediate position.

Fig. 11 shows a comparison between MLR models and spatial interpolation techniques for the Gerapotamou Basin. When the gauges located inside the basin were considered, the estimates ranged from 698 mm using IDW to 782 mm using MLR. Estimates for gauges located inside the larger extent ranged from 701 mm using IDW to 763 mm using MLR. Notice the slight difference in the spatial distribution between the two cases of MLR

Table 4. Multiple Linear Regression Models for Some Basins for Year 1974–1975

Spatial extent	Area (sq km)	Mean elevation (m)	Average slope (m/km)	Models used	Number of stations	β_0	β_1	β_2	β_3	R_2 (%)	Conditional		Unconditional	
											Average rain (mm)	Total rain (mega M3)	Average rain (mm)	Total rain (mega M3)
Gerapotamou	602	471	170	Crete ($x1$)	66	536.00	0.65	—	—	43.6	843	508	897	541
				Crete ($x1$,$x2$)		735.99	0.69	−1.65	—	59.5	860	519	918	554
				Crete ($x1$,$x3$)		−234.72	0.74	—	7.92	65.6	765	461	827	499
				Crete ($x1$,$x2$,$x3$)		50.60	0.74	−0.84	6.04	68.4	790	477	852	514
				South ($x1$)	30	472.85	0.65	—	—	46.2	778	469	832	502
				South ($x1$,$x2$)		647.08	0.70	−1.64	—	56.0	778	469	836	504
				South ($x1$,$x3$)		−416.86	0.63	—	10.57	72.3	745	449	798	481
				South ($x1$,$x2$,$x3$)		−543.60	0.62	0.38	11.60	72.5	745	449	796	480
				inside ($x1$)	9	371.52	0.97	—	—	53.8	827	449	908	548
				inside ($x1$,$x2$)		1,495.17	0.84	−8.76	—	63.8	813	490	884	533
				inside ($x1$,$x3$)		−823.57	0.48	—	16.87	86.5	784	473	824	497
				inside ($x1$,$x2$,$x3$)		−290.51	0.47	−3.42	15.54	87.8	782	471	821	495
				inside and out ($x1$)	24	492.08	0.63	—	—	50.7	788	475	841	507
				inside and out ($x1$,$x2$)		853.03	0.67	−2.96	—	55.3	804	485	860	519
				inside and out ($x1$,$x3$)		−246.74	0.58	—	8.80	75.8	748	451	797	480
				inside and out ($x1$,$x2$,$x3$)		106.25	0.62	−2.86	8.75	80.2	763	460	815	491
Anapodiaris	520	455	167	Crete ($x1$)	66	536.00	0.65	—	—	43.6	880	457	921	478
				Crete ($x1$,$x2$)		735.99	0.69	−1.65	—	59.5	851	442	894	464
				Crete ($x1$,$x3$)		−234.72	0.74	—	7.92	65.6	817	424	863	448
				Crete ($x1$,$x2$,$x3$)		50.60	0.74	−0.84	6.04	68.4	815	423	861	447
				South ($x1$)	30	472.85	0.65	—	—	46.2	816	423	856	444
				South ($x1$,$x2$)		647.08	0.70	−1.64	—	56.0	770	399	814	422
				South ($x1$,$x3$)		−416.86	0.63	—	10.57	72.3	794	412	833	432
				South ($x1$,$x2$,$x3$)		−543.60	0.62	0.38	11.60	72.5	805	418	844	438
				inside ($x1$)	9	390.03	0.72	—	—	82.5	771	400	816	423
				inside ($x1$,$x2$)		353.51	0.72	0.25	—	82.6	771	400	816	423
				inside ($x1$,$x3$)		−132.22	0.38	—	7.59	85.2	700	363	729	376
				inside ($x1$,$x2$,$x3$)		−232.47	0.16	−1.96	13.20	86.8	647	336	657	341
				inside and out ($x1$)	22	433.09	0.67	—	—	69.9	789	409	830	431
				inside and out ($x1$,$x2$)		39.37	0.60	2.81	—	73.6	788	409	825	428
				inside and out ($x1$,$x3$)		−189.46	0.62	—	7.48	83.7	759	394	797	414
				inside and out ($x1$,$x2$,$x3$)		−175.73	0.62	−0.18	7.61	83.7	758	393	797	414
Giofyro	189	341	163	Crete ($x1$)	66	536.00	0.65	—	—	43.6	846	151	846	151
				Crete ($x1$,$x2$)		735.99	0.69	−1.65	—	59.5	839	150	839	150
				Crete ($x1$,$x3$)		−234.72	0.74	—	7.92	65.5	873	156	873	156
				Crete ($x1$,$x2$,$x3$)		50.60	0.74	−0.84	6.04	68.4	861	154	861	154
				North ($x1$)	19	651.83	0.79	—	—	47.4	1,028	184	1,028	184
				North ($x1$,$x2$)		843.27	1.00	−2.95	—	71.4	913	164	913	164
				North ($x1$,$x3$)		−243.41	1.03	—	7.55	53.5	962	172	962	172
				North ($x1$,$x2$,$x3$)		3,458.10	0.62	−6.39	−20.18	82.9	957	171	957	171
				inside ($x1$)	3	591.73	0.67	—	—	86.3	911	163	911	163
				inside ($x1$,$x2$)		—	—	—	—	—	—	—	—	—
				inside ($x1$,$x3$)		—	—	—	—	—	—	—	—	—
				inside ($x1$,$x2$,$x3$)		—	—	—	—	—	—	—	—	—
				inside and out ($x1$)	13	454.66	0.78	—	—	55.3	828	148	828	148
				inside and out ($x1$,$x2$)		1,259.00	0.65	−5.51	—	67.8	807	145	807	145
				inside and out ($x1$,$x3$)		−358.36	1.09	—	7.59	62.5	881	158	881	158
				inside and out ($x1$,$x2$,$x3$)		371.40	1.00	−6.33	9.40	78.6	870	156	870	156

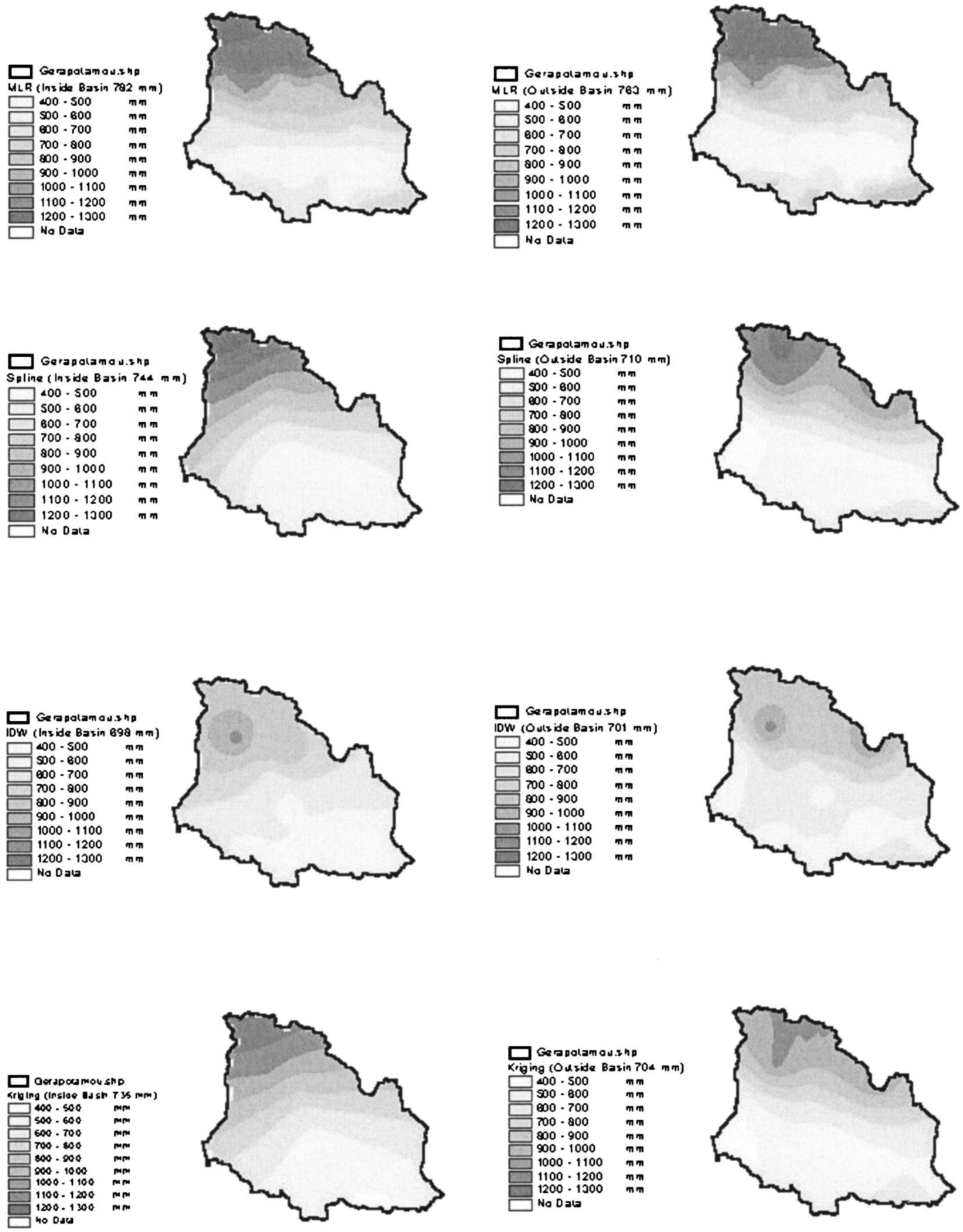


Fig. 11. Comparison between multiple linear regression models and conventional spatial interpolation techniques for smaller scale areas (basins)

compared to the two cases for each of the other spatial interpolation techniques.

Fig. 12 shows the 13 gauges located in and around the Anapodiaris Basin, from which annual data for 30 years (from 1967–1968 to 1996–1997) were obtained and used to derive MLR models. The models were then applied for each year to estimate the areal rain and volume for that particular basin based on the two scenarios of conditional and unconditional assumptions. The values obtained show a cyclic behavior on a downward trend for the areal estimates of −3.67 mm/year, which corresponds to −1.9 Mm^3/year change in the rain volume (refer to Fig. 13). The shaded area of this figure represents the output of 30 MLR models (one for each year) developed for this basin. The unshaded areas represent the de-trended precipitation values. Based on the conditional and unconditional assumptions, the average rain over the 30 years, for the Anapodiaris Basin, is 805–855 mm/year (or 420–445 Mm^3/year).

Small-Scale (Sub-Basins)

Following the delineation of sub-basins using *ArcView* GIS, another form of analysis was carried out for the Giofyro Basin. The MLR models, together with the GIS module, were used to estimate both areal rain and the mean elevation for the sub-basins. It

J. Hydrol. Eng. 2004.9:79-102.

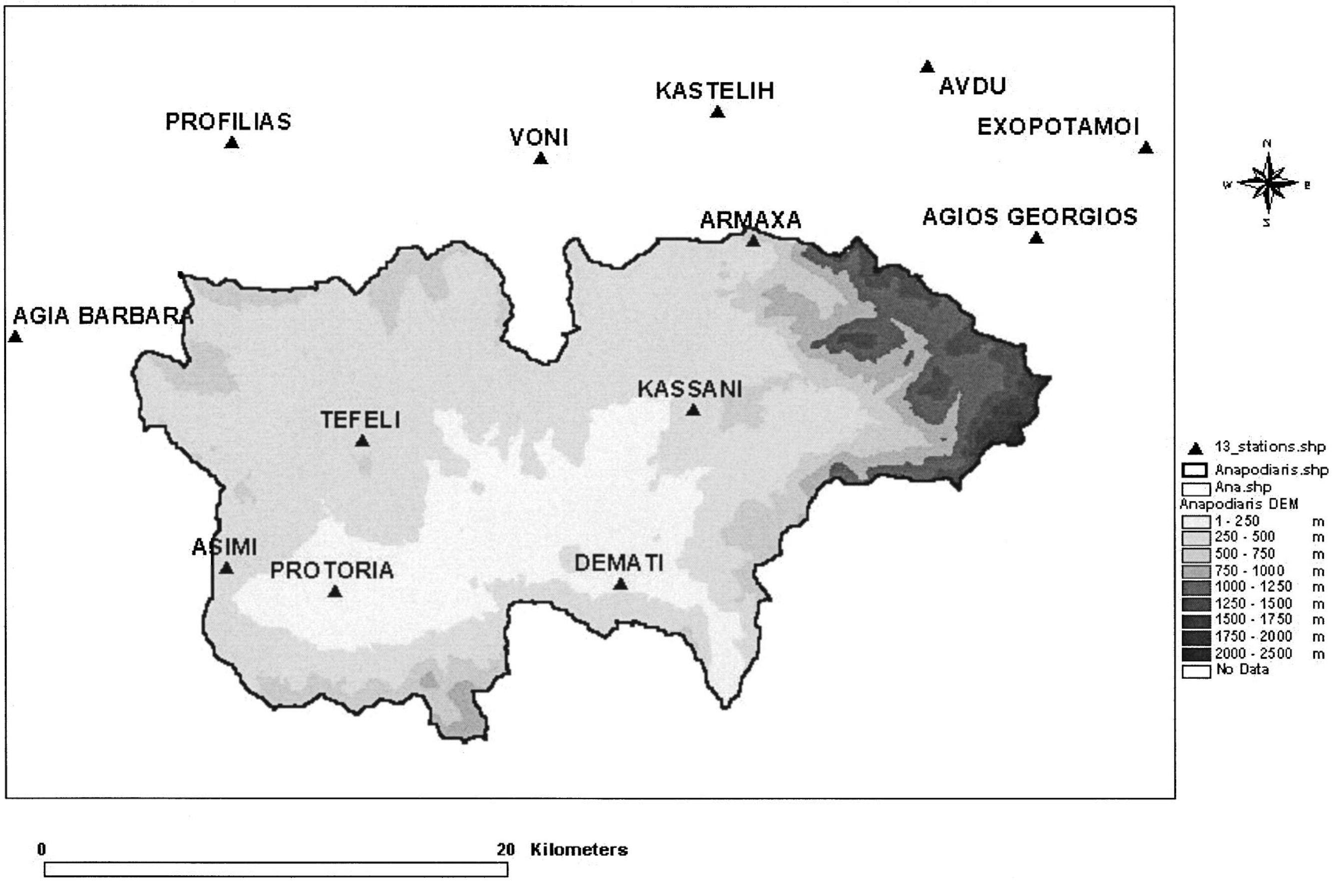


Fig. 12. Anapodiaris Basin and 13 gauges with common set of data for 30 years (1967–1968 to 1996–1997)

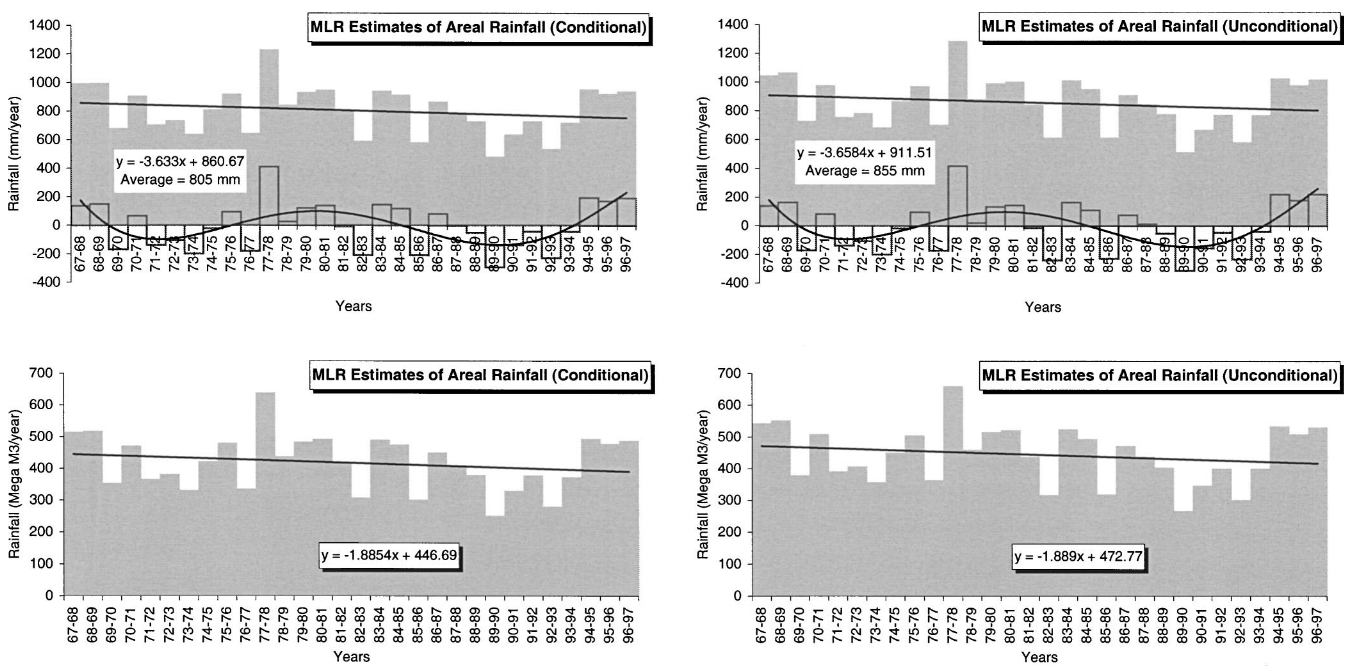
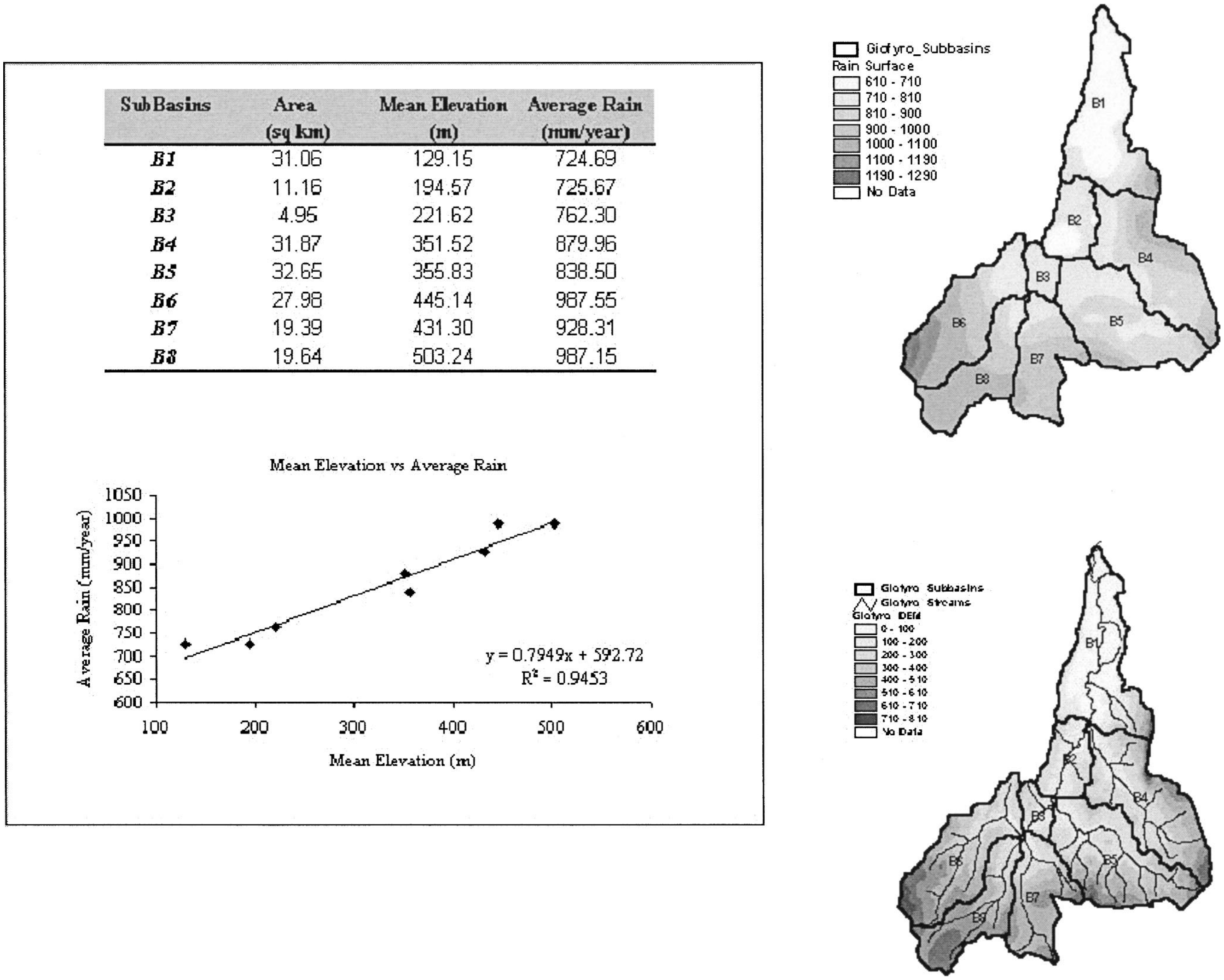


Fig. 13. Areal rain estimated yearly for Anapodiaris Basin

J. Hydrol. Eng. 2004.9:79-102.



Sub Basins	Area (sq km)	Mean Elevation (m)	Average Rain (mm/year)
B1	31.06	129.15	724.69
B2	11.16	194.57	725.67
B3	4.95	221.62	762.30
B4	31.87	351.52	879.96
B5	32.65	355.83	838.50
B6	27.98	445.14	987.55
B7	19.39	431.30	928.31
B8	19.64	503.24	987.15

Fig. 14. Areal rain estimated for Giofyro sub-basins using multiple linear regression models for average year

is clear that, for an average year, mountainous sub-basins receive more rain than coastal ones. Fig. 14 shows that sub-basins (B6, B7, and B8), which occupy the highest elevations, collect greater amounts of rain than the rest of the sub-basins. A strong positive correlation existed that relates the mean elevation and the average rain. The slope of the line (0.79 mm/m) is comparable to the results obtained for Giofyro Basin shown in Table 4.

In addition to estimating areal rain for areal extents (Crete, northern Crete, southern Crete, eastern Crete, Anapodiaris Basin, Gerapotamou Basin, Giofyro Basin, and sub-basins), the developed GIS module also provides a means for estimating areal rain for ranges of elevation, as shown in Table 5. In all three basins, elevations that are higher than 600 m occupied around 28% of the area of the basin. However, Gerapotamou Basin collects more rain for that range due to the fact that its highest elevation is 2,300 m. The highest elevation for Anapodiaris Basin is 2,067 m, and for Giofyro Basin is 809 m. Based on an average year, the elevation range of 300–600 m in the Gerapotamou Basin receives almost 45% of the amount of rain that falls on the elevation range of 600 m and higher. Generally, most of the precipitation falls between the elevations 300–600 m.

Table 5. Estimated Areal Rain for Basins for Different Elevation Ranges for Average Year

Elevation range		Giofyro Basin	Anapodiaris Basin	Gerapotamou Basin
$E \leqslant 300$ M	% of basin area	15%	12%	32%
	Average rain	680 mm	575 mm	564 mm
300 M $< E <$ 600 M	% of basin area	57%	60%	40%
	Average rain	830 mm	700 mm	720 mm
$E \geqslant 600$ M	% of basin area	28%	28%	28%
	Average rain	1,045 mm	1,090 mm	1,230 mm

J. Hydrol. Eng. 2004.9:79-102.

Conclusion

The MLR method has been used to develop models to estimate the amount and the spatial distribution of orographic precipitation for complex terrain such as that of the island of Crete in Greece using elevation, latitude, and longitude. This can be considered a new spatial interpolation technique that provided comparable results to those of the conventional spatial interpolation techniques such as spline, IDW, and kriging. The MLR models were derived and applied using a GIS. The use of MLR in conjunction with GIS provides flexibility and accuracy in estimating both areal and point precipitation for different scales with minimum user intervention. The MLR technique was found to be superior for estimating precipitation at point (ungauged) locations. Negative values for point precipitation estimated using the other conventional spatial interpolation techniques are not uncommon, especially if the rain gauges network is not dense enough (sparse). This problem can be avoided by using the appropriate MLR models which yield not only positive but reasonable values as well.

The three-variable models are easier and more acceptable, especially when dealing with a relatively small number of rain gauges. The more parameters used, the more the degrees of freedom are compromised. In some cases, however, the ten-parameter models are regarded as more informative. As for the case study, the island of Crete, different models were derived that were used for different scales that ranged from large (the island), medium (watersheds), and small (sub-basins). Results show that Crete could receive around 950 mm (8×10^9 m^3) of rain for an average year, with the north sharing 44%, the south 33%, and the east 23% of that amount. For a wet year, the island could receive around 1,440 mm of rain with the north sharing 40%, the south 36%, and the east 24% of that amount. For a dry year, the island could receive around 600 mm of rain with the north sharing 46%, the south 30%, and the east 24% of that amount. These analyses should be of value for the purpose of water resources planning and management in those regions where the dominating type of rain is orographic.

References

Alter, J. C. (1919). ''Normal precipitation in Utah.'' *Monthly Weather Rev.,* 47, 633–636.

Barrows, H. K. (1933). ''Precipitation and runoff and altitude relations for Connecticut River.'' *Trans., Am. Geophys. Union,* 14, 396–406.

Barry, R. G., and Chorley, R. J. (1976). *Atmosphere, weather, and climate*, 3rd Ed., Methuen, London.

Boyle, S. J., and Tsanis, I. K. (1998). ''IDOR^{2D}GIS—A closely coupled hydrodynamic/pollutant transport GIS model.'' *Proc., 18th Annual ESRI Int. User Conf.*, San Diego.

Boyle, S. J., Tsanis, I. K., and Kanaroglou, P. S. (1998). ''Developing geographic information systems for land use impact assessment in flooding conditions.'' *J. Water Resour. Plan. Manage.,* 124(2), 89–98.

Brimicombe, A. J., and Bartlett, J. M. (1996). ''Linking geographic information systems with hydraulic simulation modeling for flood risk assessment: The Hong Kong approach.'' *GIS and environmental modeling*, M. F. Goodchild, ed., Oxford University Press, New York, 165–168.

Burns, J. I. (1953). ''Small-scale topographic effects on precipitation distribution in San Dimas experimental forest.'' *Trans., Am. Geophys. Union*, 34, 761–768.

Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology*, McGraw-Hill, New York.

Chua, S. H., and Bras, R. L. (1982). ''Optimal estimators of mean areal precipitation in regions of orographic influence.'' *J. Hydrol.,* 57, 23–48.

Daly, C., Neilson, R. P., and Phillips, D. L. (1994). ''A statistical-topographic model for mapping climatological precipitation over mountainous terrain.'' *J. Appl. Meteorol.,* 33(2), 140–158.

Delhomme, J. P. (1978). ''Kriging in the hydrosciences.'' *Adv. Water Resour.,* 1(5), 251–266.

Dingman, S. L., Seely-Reynolds, D. M., and Reynolds, R. C., III (1988). ''Application of kriging to estimate mean annual precipitation in a region of orographic influence.'' *Water Resour. Bull.,* 24, 329–339.

Draper, N. R., and Smith, H. (1998). *Applied regression analysis*, Wiley, New York.

Environmental Systems Research Institute (ESRI). (2000). *ArcView GIS 3.2*, Redlands, Calif.

Fan, Y., and Duffy, C. J. (1991). ''Estimating space-time precipitation and temperature fields in mountainous terrain: Wasatch Range, Utah.'' *Proc., American Geophysical Union Fall Meeting Program and Abstracts*, San Francisco.

Goodchild, M. F., Parks, B. O., and Steyaert, L. T. (1995). *GIS and environmental modeling: progress and research issues*, GIS World Books.

Greene, R. G., and Cruise, J. F. (1995). ''Urban watershed modeling using geographic information systems.'' *J. Water Resour. Plan. Manage.*, 4(121), 318–325.

Hay, L. E., Branson, M. D., and Leavesly, G. H. (1991). ''The effects of scale on precipitation estimates from an orographic precipitation model.'' *Proc., American Geophysical Union Fall Meeting Program and Abstracts*, San Francisco.

Hibbert, A. R. (1977). ''Distribution of precipitation on rugged terrain in central Arizona.'' *Hydrology Water Resources Arizona Southwest*, 7, 163–173.

Houghton, J. G. (1979). ''A model for orographic precipitation in the North-Central Great Basin.'' *Monthly Weather Rev.,* 107, 1462–1475.

Lichy, C. (1998). ''A multidirectional and multifunctional gateway between GIS and hydrodynamic models.'' *Proc., Hydroinformatics 98*, Copenhagen, Denmark.

Maidment, D. R. (1993). ''GIS and hydrologic modeling.'' *Environmental modeling with GIS*, M. F. Goodchild, B. O. Parks, and L. T. Steyaert, eds., Oxford University Press, New York, 147–167.

Marwitz, J. D. (1987). ''Deep orographic storms over the Sierra Nevada. Part I: Thermodynamic and kinematic structure.'' *J. Atmos. Sci.,* 44, 159–173.

Matheron, G. (1971). ''The theory of regionalized variables and its applications.'' Cahiers du Centre de Morphologie Mathematique, Ecole des Mines, 211 pages.

Naoum, S., and Tsanis, I. K. (2003). ''Temporal and spatial variation of annual rainfall on the island of Crete, Greece.'' *Hydrolog. Process.,* 17(10), 1899–1922.

Naoum, S., and Tsanis, I. K. (2004). ''A multiple linear regression GIS module using spatial variables to model orographic rainfall.'' *J. Hydroinformatics,* 6(1), 39–56.

Naranjo, E., and Larsen, H. (1998). ''An integrated model to assess pollution loads with the use of GIS and numerical models.'' *Proc., Hydroinformatics 98*, Copenhagen, Denmark.

Peck, E. L., and Brown, M. J. (1962). ''An approach to the development of isohyetal maps for mountainous areas.'' *J. Geophys. Res.*, 67, 681–694.

Peck, E. L., and Schaake, J. C. (1990). ''Network design for water supply forecasting in the west.'' *Water Resour. Bull.,* 26, 87–99.

Phillips, D. L., Dolph, J., and Marks, D. (1992). ''A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain.'' *Agric. Forest Meteorol.,* 58, 119–141.

Reed, W. G., and Kincer, J. B. (1917). ''The preparation of precipitation charts.'' *Monthly Weather Rev.,* 45, 233–235.

Ross, M. A., and Tara, P. D. (1993). ''Integrated hydrologic modeling with geographic information systems.'' *J. Water Resour. Plan. Manage.,* 119(2), 129–140.

Schermerhorn, V. P. (1967). ''Relations between topography and annual precipitation in western Oregon and Washington.'' *Water Resour. Res.,* 3, 707–711.

Singh, V. P. (1992). *Elementary hydrology*, Prentice-Hall, Englewood Cliffs, N.J.

Smith, R. B. (1979). ''The influence of mountains on the atmosphere.'' *Advances in geophysics*, Vol. 21, Academic, New York, 87–230.

Spreen, W. C. (1947). ''A determination of the effect of topography upon precipitation.'' *Trans., Am. Geophys. Union*, 28, 285–290.

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

Thiessen, A. H. (1911). ''Precipitation average for large areas.'' *Monthly Weather Rev.*, 39, 1082–1084.

Tsanis, I. K., Valeo, C., Wu, J., and Boyle, S. (1996). ''Managing contaminated sediments using a hydrodynamic model and a GIS.'' *Environ. Technol.*, 17, 877–883.

J. Hydrol. Eng. 2004.9:79-102.