SCALING METHODS IN SOIL PHYSICS

Edited by

Yakov Pachepsky
David E. Radcliffe
H. Magdi Selim

CRC PRESS

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Scaling methods in soil physics / edited by Yakov Pachepsky, David Radcliffe, H. Magdi Selim.
p. cm.
Includes bibliographical references and index.
ISBN 0-8493-1374-0 (alk. paper)
1. Soil physics. 2. Scaling laws (Statistical physics) I. Pachepsky, Y. II. Radcliffe, David Elliot, 1948- III. Selim, Hussein Magd Eldin, 1944-

S592.3 .S32 2003
631.4′3--dc21 2002191161

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Visit the CRC Press Web site at www.crcpress.com

International Standard Book Number 0-8493-1374-0
Library of Congress Card Number 2002191161
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

13 A Mix of Scales: Topography, Point Samples and Yield Maps

D.J. Timlin, Y.A. Pachepsky, and C.L. Walthall

CONTENTS

I. INTRODUCTION

Agricultural land management has requirements for many kinds of data, including information on soil hydraulic properties, soil texture, terrain attributes, and crop cover. The sources of these data are varied and range from manual collection of sparse point samples to acquisition of highly dense remotely sensed soil and crop canopy reflectance, and yield monitor data. Samples collected to characterize soil properties, however, are often collected manually and are likely to be discontinuous and represent localized sites and small scales. Elevation and topographic parameters may be useful to help interpolate discontinuous values of soil properties because they provide a more or less continuous distribution of measurements over large areas and usually represent a range in scales.[1,2] Terrain indices have also been shown to be related to moisture distributions in soils.[3,4]

Soil texture or water holding capacity data collected as point samples may provide an adequate statistical representation of a local variable but there may not be enough data to accurately represent the overall pattern of variability in order to map the data. Terrain attributes such as soil surface curvature, slope and elevation are related to soil texture, soil water content and crop yields.[5–9] Slope and soil surface curvatures have been shown to be good predictors of distributions of soil texture in the landscape.[7,10] Elevation can be easily measured and generally has a reasonably continuous variance over space in agricultural fields. By relating soil water holding capacity to topographic variables it may be possible to interpolate sparse samples of soil water holding capacity.[10] Thus

0-8493-1374-0/03/$0.00+$1.50

one could obtain a more accurate distribution of soil water holding capacity than from interpolation of the measured variables alone.

Regression analysis is commonly used to develop predictive equations for interpolation or to explain variability in terms of known variables. Regression is commonly used to develop pedotransfer functions to relate easily measured soil attributes to soil hydraulic properties.[11,12] The measured variables are often not collected with spatial information and are assumed to be independent. When ordinary least squares regression (OLS) is used on data that are spatially autocorrelated, the standard errors can be underestimated, tests of significance overinflated,[13] and prediction errors large.[14] Spatial autoregression can be a useful tool for quantifying large scale measurements, terrain variables, and point measurements where spatial autocorrelation exists.[13,15] Knowledge of spatial relationships can be used to "fill in" or interpolate sparse measurements of soil properties. Redundant information in the variables and local variances can be exploited to make better estimates.

Previous research at the site studied in this chapter[7] reported that the strongest relationship between soil water retention and topographic variables was observed at capillary pressures of 10 and 33 kPa, i.e., in the range where the soil reaches its "field capacity." Because the water content at 15 MPa (wilting point) did not vary substantially, the study proposed that the water holding capacity depended on landscape position and these available water holding capacities predicted from terrain variables could be used as interpretive attributes for yield maps in precision agriculture.

The objective of this study was to characterize soil water holding capacity (WHC) at the field scale using topographic variables as predictors and applying a spatial autoregressive response model to account for spatial dependence of the water holding capacities on terrain attributes. Several scales of interpolation of the elevation data were used to investigate the effect of interpolated DEM (digital elevation model) grid size on the relationship between calculated terrain variables and WHC. The spatially estimated water holding capacities were compared to measured yields to determine their usefulness as interpretive attributes for precision agriculture.

II. ANALYSIS OF DATA FROM VARIOUS SOURCES

A. Site Description

The experimental site was a 6-ha corn field located on the USDA, Henry A. Wallace Beltsville Agricultural Research Center in Beltsville, Maryland. The data set described here was collected during the growing seasons of 1997 and 1998. The investigation carried out on this field is part of a larger study entitled Optimizing Production Inputs for Economic and Environmental Enhancement (OPE3). The OPE3 project is addressing: 1) watershed scale fluxes of chemical inputs and biological agents from conventional, precision, and sustainable farming practices, 2) environmental impact of chemical inputs and biological agents on a wooded riparian wetland, 3) development of remotely sensed data products and analytical techniques for measuring and managing the spatial variability of crops and soils, and 4) long-term economic and environmental impacts and tradeoffs of precision and sustainable agricultural production practices.

The study site has a gentle slope running from the northwest part of the site to southeast with an elevation difference of approximately 4 m (Figure 13.1). Sandy loam soils predominate with silt and clay contents increasing down slope. The corn was planted north to south in 0.74-m wide rows. At the time of the data collection there were no known major pest infestations. The yields were recorded using an AgLeader 2000 yield monitor (Ag Leader Technology, Ames, Iowa).

A topographic survey of the site was obtained from a total of 555 photogrammetric mass points from airborne stereophotography (Air Survey Corp., Virginia). The average mass of sampled points was about one measurement per 140 m^2. Elevation values were obtained from a digital elevation model constructed by interpolation of the photogrammetric points' data to nodes of various grid sizes that ranged from 5×5 m to 55×55 m. The interpolation method of minimum curvature in

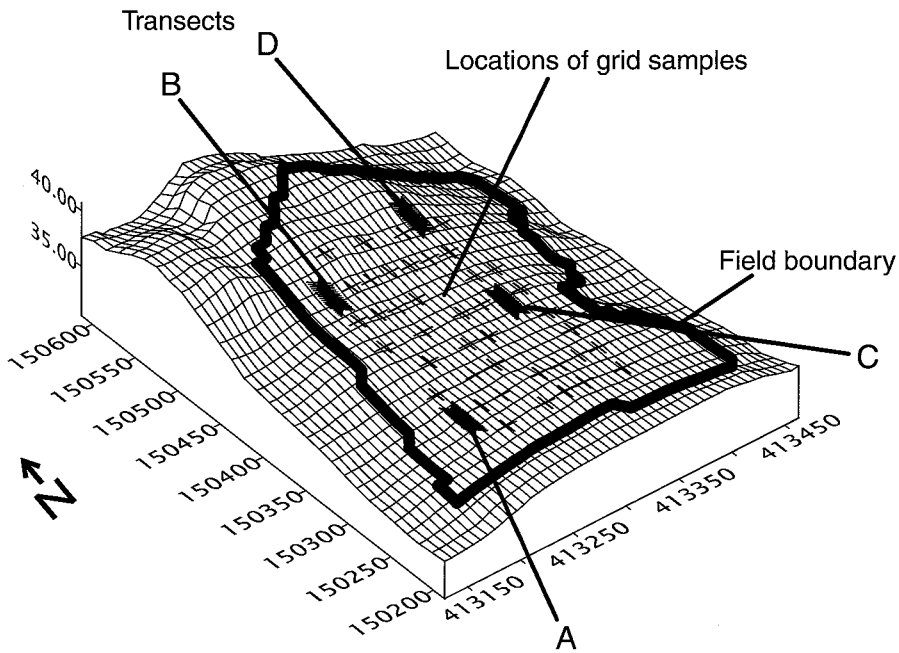

FIGURE 13.1 Elevation map of study site showing locations of transects, grid sampling locations, and field boundary.

Surfer (Golden Software, Golden, Colorado) was used to interpolate the grids. The method of minimum curvature was reported to give good results based on comparisons among five commonly used interpolation methods.[7]

Values of maximum slope, profile curvature and tangential curvature were used as topographic attributes for the study area. Profile curvature is defined as curvature of the surface cross section made in the direction of maximum slope. This is the uphill rate of change in slope. Negative (positive) values of profile curvature indicate convex (concave) flow paths where a surface flow accelerates (slows down). Tangential curvature is defined as curvature of the vertical surface cross section made perpendicular to the direction of maximum slope. Negative (positive) values of tangential curvature represent areas of divergent (convergent) flow.[16] Equations for calculation of the terrain parameters are given in Pachepsky et al.[7]

Soil was sampled along four transects positioned in different landscape elements (i.e., along slopes and at foot slopes) and at nodes of a 30- × –30-m grid as shown in Figure 13.1 . The sampling locations were 2 m apart within each transect. All samples were taken in duplicate from points 30 cm apart. There were 54 duplicated transect samples and 39 duplicated grid samples. Sampling depth was 4 to 10 cm. The same topographic variables were assigned to the duplicated samples.

A transect sampling scheme was used to provide detailed information on changes in soil properties along a gradient of slope, which would not have been possible with more widely spaced samples. However, the transects have been augmented by grid samples that provide a (pseudo) nested sampling structure.

B. Spatial Autoregression

The regressions were carried out using a spatial autoregressive response model. A more complete description of autoregressive models applied to spatial data can be found in References 13 and 17 through 19. In ordinary least squares regression, the dependent variable is a function of the independent variable as $Y = X\beta + \varepsilon$ where Y and X are $n \times 1$ vectors, β is a $1 \times k + 1$ vector of

coefficients and ε is error with mean of 0 and σ^2 = constant. There are n observations(dependent, Y) and k predictors (independent, X). If Y or X exhibit spatial autocorrelation, then they must be corrected for values of X or Y at nearby locations. Because the data are collected on a two-dimensional grid, a value of X or Y at a lag of $i - 1$ cannot be used to correct for autocorrelation as it can for a one-dimensional series of data. Instead, a connectivity matrix D is used.[18,20] The autoregressive model becomes:

$$Y - \alpha DY = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon \tag{13.1}$$

The value α is the autoregressive parameter and lies between 0 and 1. It is assumed here that α is a constant for the field. This may or may not be appropriate but is a good first approximation. The connectivity matrix, D, is an $n \times n$ weighting matrix with 0 on the diagonal so that only neighboring values and not the value itself are used for predictions. The independent variables are the three terrain attributes (slope, profile curvature and tangential curvature). The error, ε, is distributed as $\varepsilon \sim N(0,\sigma^2 I)$. We used an $n \times n$ connectivity matrix (D) described in Pace and Barry.[19] The properties of this matrix are 1) the diagonals are zero, and 2) the rows sum to one. The strength of the connection is related to distance between points and does not require uniform separation distances and can thus accommodate nonuniform grid spacings.

Figure 13.2 shows the relationship between a 3×3 grid (A) with equally spaced cells and the spatial weights (B) calculated by the nearest neighbor method. For example, the first row of the adjusted Y (αDY), Y_{1a}, would be calculated as $\alpha 0.5Y_2 + \alpha 0.5Y_4$. This is a two-dimensional analog of a time series model where the current Y value is adjusted for past values. The spatial connectivity matrix adjusts the Y value for its neighbors.

The autoregressive parameter, α, is solved using maximum likelihood computations.[19] The profile likelihood function is defined as:

$$L(\beta,\alpha,\sigma^2) = C + \ln|I - \alpha D| - \left(\frac{n}{2}\right)\ln(SSE) \tag{13.2}$$

where C is a constant and $ln|I - \alpha D|$ is the log determinant. The SSE is the sum of squared errors:

$$SSE = (Y - \alpha DY - X\beta_\alpha)^T (Y - \alpha DY - X\beta_\alpha) \tag{13.3}$$

where T denotes transpose of the matrix and $\beta_{-\alpha}$ is an $n \times 1$ vector of coefficients $\beta_0 - \beta_3$ in Equation 13.1). An efficient and rapid method to evaluate the maximum of the profile likelihood function and hence solve for α and β was used here.[19] The log determinant, $ln|I - \alpha D|$, is calculated for a number (usually 100) of values of α and then a lookup table is used to find the minimum of L in Equation 13.2. Matlab (The Mathworks, Natick, Massachusetts) programs from the Spatial-ToolBox, v1.1* were used to perform the calculations. The SpatialToolBox function FPAR1 was used to compute the spatial autoregression.

After α was determined, the error sums of squares for α were then calculated. Spatially adjusted values of Y (WHC) were calculated by subtracting αDY from Y. These adjusted Y values (Y_a) were next used in an OLS regression with the topographic variables ($X_1 - X_3$). The sums of squares for α were calculated by subtracting the total sums of squares for the OLS [Y_a – mean (Y_a)] from the total sums of squares for the AR model [Y – mean(Y)]. Next, the sums of squares for the OLS regression were obtained by subtracting the error sums of squares for the OLS ($Y_a - \hat{Y}_a$) from the total sums of squares for OLS [Y_a – mean(Y_a)].

* Spatial Statistics Toolbox 1.1. R. K. Pace and R. Barry, available at: http:\\spatial-statistics.com

A)

1	2	3
4	5	6
7	8	9

B)

Cell	1	2	3	4	5	6	7	8	9
1	0	1/2	0	1/2	0	0	0	0	0
2	1/3	0	1/3	0	1/2	0	0	0	0
3	0	1/2	0	0	0	1/2	0	0	0
4	1/2	0	0	0	0	0	1/2	0	0
5	0	1/2	0	0	0	0	0	1/2	0
6	0	0	1/2	0	0	0	0	0	1/2
7	0	0	0	1	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	1	0	0	0

FIGURE 13.2 Representation of the spatial weighting matrix (D) for a 3 × 3 grid. The weights sum to 1 in the rows. The Spatial Statistics Toolbox, v1.1 was used to obtain the weights.

1. Regression Methodology

As described above, terrain variables (slope and curvature) were calculated from the measured elevation data that were interpolated to different sizes of regular grids (5 × 5 m to 55 × 55 m). The calculated terrain variables were interpolated to the measured locations and the regressions were carried out using the WHC data in their measured locations, i.e., no gridding was performed for the WHC data. The same number of independent and dependent variables were used for each scale at which the regressions were carried out.

2. Prediction of WHC at Different Scales

In order to calculate WHC, predicted WHC, ($\hat{Y}$), as a function of terrain variables at different scales, Equation 13.1 can be rewritten as:

$$\hat{Y} = \frac{\beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3}{1 - \alpha D} \tag{13.4}$$

The value of α is determined from the spatial autoregression of the measured water holding capacities vs. the terrain variables calculated as a function of a specific grid size. The predictions are an interpolative process to fill in water holding capacities on regular grids at locations where measurements are missing. In actuality, Equation 13.4 is a smoothing interpolator for WHC because the predicted values are divided by the means of neighboring values. The grid sizes for interpolation range from 5 × 5 (4725 values) to 55 × 55 (20) values.

C. Yield Data

For comparison with the interpolated WHCs, the 6- × 2-m yield data were gridded to 20- × 20-, 25- × 25-, 30- × 30-, and 35- ×y 35-m grids by block kriging using Surfer. The yield data were then smoothed by using an average of the neighboring blocks (one grid cell on each side). The linear default semivariogram was used. The linear semivariogram provides a smoothing of the data similar to a moving average. Because the data were nearly continuous, the effect of the kriging was to smooth the data rather than interpolate and fill in missing locations. For this reason, an isotrophic, linear semivariogram was thought to be adequate. The WHC predicted from terrain attributes were interpolated to the yield measurement locations using Surfer, which uses a bilinear interpolation method.

III. INTERPOLATION AND MAPPING WITH AUXILIARY VARIABLES

A. Topography and Water Flow in the Landscape

The field slopes gently from the north to the southeast (Figure 13.1). This generates flow of water and materials toward the riparian area on the eastern border of the field. Zones of high positive tangential curvature [convergent flow (red)] are aligned with the flow vectors (Figure 13.3) so that zones are more continuous in the east to west direction than in the north to south direction.

These data suggest that landscape properties would have some relation to the correlation scales of crop yield and soil hydraulic properties. Semivariance for crop yields has a longer range in the east to west direction than in the north to south direction and suggests that the soil conditions that

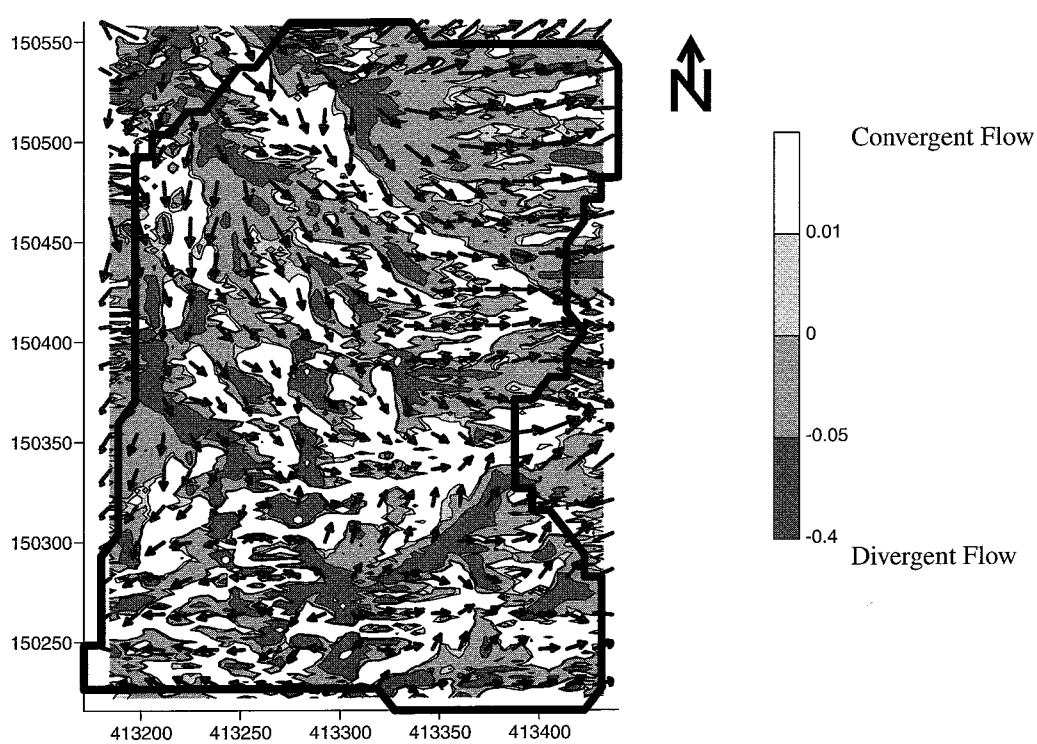

FIGURE 13.3 (See color insert following page 144.) Tangential curvature and water flow vectors calculated from a 5- × 5-m DEM grid.

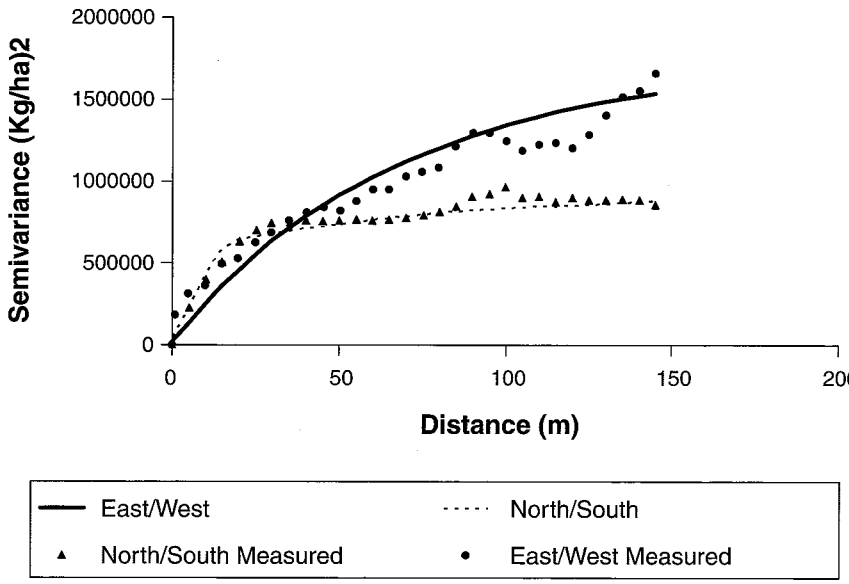

FIGURE 13.4 Semivariogram of the 1998 yield data for N–S and E–W directions

affect crop growth also have this correlation structure (Figure 13.4). Fine soil material with higher WHC and nutrients would tend to accumulate in the areas that are concave and converge water flow and be removed from areas that are convex and shed water.

B. Terrain Variables and Soil Properties

Results from the same field[7] showed that topographic variables complemented each other in distinguishing zones of different texture within the landscape. Sands and silts were separated reasonably well by slope, and tangential curvature discriminated transects by clay. Tangential curvature and slope were significantly related to water contents at 10 and 33 kPa.[7]

Figure 13.5 shows the relationship between WHC (the difference between the 1500 and 10 kPa water contents) and soil terrain variables: slope, tangential curvature, and profile curvature. There was only a weak relationship between profile curvature and WHC. The relationships between WHC and slope, and WHC and tangential curvature are stronger. In this site, the clay and silt contents of the soil increase from the north section of the field to the south as elevation decreases. Slow erosion of soil over a long period of time has resulted in movement of fines downslope and an increase in WHC. The relationship with slope is stronger than with tangential curvature. Tangential curvature rather than slope, however, was useful for distinguishing transects A and C, which differed with respect to soil texture. The contour map of WHC in Figure 13.6 illustrates the relationship with slope where WHC is increasing downslope with increasing silt content. Areas of equal WHC are distributed perpendicular to the slope.

C. Spatial Autoregression

Spatial autoregressions were carried out with terrain variables calculated from elevation grids for a range in scales. Because the relationship with profile curvature was weak, only slope and tangential curvature were used as regressors. The relationships shown in Figure 13.5 appear nonlinear. A linear relationship was used here as a first approximation because there was little prior information on how the linearity would be affected at different scales. The regression results in Table 13.1 show that a large part of the variation had a spatial component. The value of α, the spatial autoregression parameter, varied from 0.49 to 0.75 and was largest for terrain variables calculated on the finest

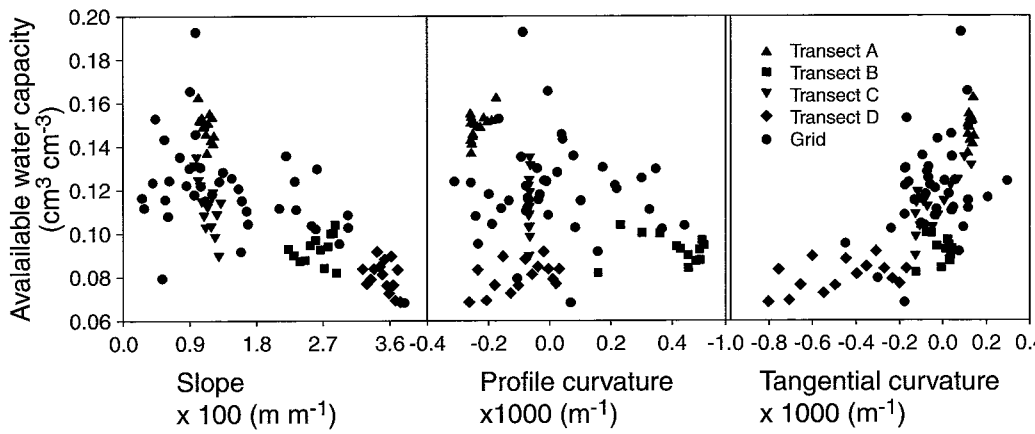

FIGURE 13.5 Relationships between WHC and the terrain attributes, slope, profile curvature, and tangential curvature.

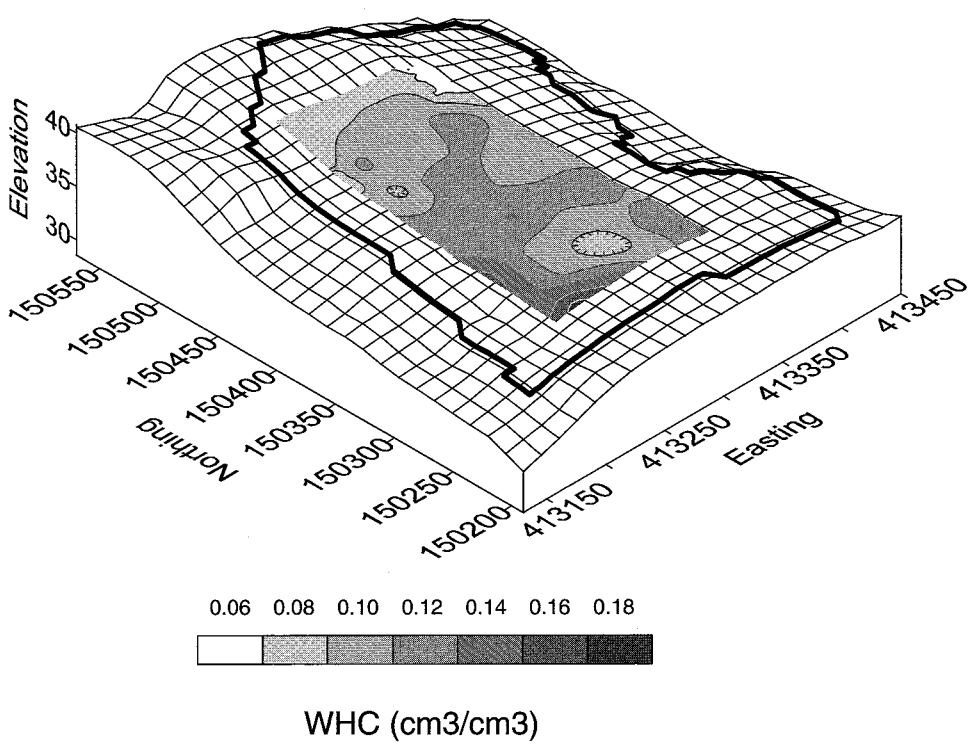

FIGURE 13.6 Elevation map of study site with contours of measured WHC. The field boundary is shown.

grids. The amount of error explained by α was about five times that explained by the regression. The proportion of the sums of squared differences explained by α was fairly constant among all the scales, about 0.041. The error explained by regression varied from 0.02 to 0.012 $cm^3\ cm^{-3}$. The total sums of squares, which is the variation about the mean of WHC, was constant for all scales because the same values of WHC were used in all the regressions; only the terrain variables differed among scales. The error sums of squares were also fairly constant, which is a reflection of the large component of spatial variability and also shows that, as the error explained by α decreases, the

TABLE 13.1
Regression Parameters for Slope and Tangential Curvature from Spatial Autoregression and Ordinary Least Squares Regression

	Spatial autoregression parameters							Ordinary least squares parameters		
Scale m × m	α	Intercept	P	Slope	P	Tangential curvature	p	Intercept	Slope	Tangential curvature
05 × 05	0.75	0.037	0.	–0.325	0.005	2.027	0.129	0.141	–1.137	2.330
10 × 10	0.76	0.034	0.	–0.294	0.004	1.030	0.499	0.136	–0.948	2.123
15 × 15	0.72	0.041	0.	–0.400	0.000	–1.140	0.651	0.141	–1.276	–3.196
20 × 20	0.64	0.052	0.	–0.505	0.000	10.635	0.057	0.143	–1.399	21.230
25 × 25	0.56	0.061	0.	–0.478	0.002	21.238	0.002	0.137	–1.108	44.495
30 × 30	0.61	0.055	0.	–0.512	0.003	16.236	0.050	0.141	–1.320	36.532
35 × 35	0.50	0.066	0.	–0.340	0.056	39.052	0.000	0.132	–0.647	77.906
40 × 40	0.63	0.052	0.	–0.439	0.045	14.196	0.127	0.136	–0.957	48.011
45 × 45	0.61	0.057	0.	–0.631	0.004	10.072	0.458	0.144	–1.505	31.250
50 × 50	0.59	0.059	0.	–0.667	0.002	13.280	0.298	0.143	–1.560	36.319
55 × 55	0.58	0.064	0.	–0.942	0.000	0.629	0.971	0.150	–2.149	8.394

Note: The slopes and tangential curvatures were calculated from elevations interpolated to the scales given below. The units for WHC are $cm^3\ cm^{-3}$, for the slope parameter are ($cm^3\ cm^{-3}$) per ($m\ m^{-1}$)and for the tangential curvature parameter are ($cm^3\ cm^{-3}$) per m^{-1}.

error explained by regression increases. The root mean squared error of the pure OLS model was somewhat more than the root mean square error of the spatial autoregressive model.

A large component of the error is explained by the spatial location and shows the importance of using local information to reduce error in regression relationships. A previous study[13] investigated relationships between yield and NDVI (normalized difference vegetation index) and reported that, for a 9- × 9-m grid, elevation and NDVI explained only 7% of the variance in yield, while the autocorrelation parameter, α, explained 86% of the variance. The spatial autocorrelation coefficient still explained a large portion of the error in spite of including elevation, a locational parameter in the regression.

The parameters for the regression models are given in Table 13.1. The parameter for slope was significant at all scales except the 35- × 35-m scale. Slope is a large scale variable and changes smoothly over distance in this field. The parameter for tangential curvature was not significant at the larger or smaller scales. The tangential curvature, however, is calculated as second differences of elevation and can be roughly interpreted as a change in slope. At the largest scales, the small scale changes in slope are largely smoothed and the effect of tangential curvature (changes in slope) is lost. At small scales, there is so much noise in the data that a meaningful relationship between tangential curvature and WHC cannot be determined. Note that, among the different scales, the largest weight was assigned to tangential curvature at the 35- × 35-m scale where the r^2 was highest. The parameter for slope for this scale, however, was not significant at $p = 0.05$. Also, the parameter for slope is smallest where the parameter for tangential curvature is highest. Some of this effect can be attributed to colinearity between the two variables, which is not unexpected because they are both calculated from elevation.

The best fit to the data in terms of r^2, error, and significance of parameters was provided by the terrain variables calculated at the 25- × 25-m scale. The terrain variables calculated at the 35- × 35-m scale had the lowest error but not all regression parameters were significant. Within the range of the 25- × 25-m and 35- × 35-m scales there is not a clear advantage in terms of r^2 and error to calculating terrain variables from any one scale of elevation data. This may be due to the selection

TABLE 13.2
Output Statistics for the Spatial Autoregression of Water Holding Capacity vs. Terrain Parameters Where Terrain Parameters Are Calculated from Different Scales of Interpolated Elevation Data

Scale of landscape parameters		Sums of squares ($cm^3\ cm^{-3})^2$				r^2			RMSE ($cm^3\ cm^{-3}$)	
m × m	α	α	Regression	Error	Total	α	Regression	Overall	Spatial	OLS
05 × 05	0.75	0.046	0.002	0.016	0.065	71.1	3.3	75.4	0.013	0.022
10 × 10	0.76	0.046	0.002	0.017	0.065	71.5	2.3	75.0	0.013	0.023
15 × 15	0.71	0.045	0.003	0.017	0.065	69.6	4.2	74.5	0.013	0.020
20 × 20	0.64	0.043	0.005	0.017	0.065	66.2	8.2	74.6	0.013	0.018
25 × 25	0.55	0.039	0.009	0.017	0.065	60.7	13.6	74.5	0.013	0.016
30 × 30	0.61	0.042	0.006	0.017	0.065	64.5	9.7	74.4	0.013	0.017
35 × 35	0.49	0.036	0.012	0.016	0.065	56.4	18.0	74.5	0.013	0.015
40 × 40	0.63	0.042	0.005	0.017	0.065	65.6	7.9	73.8	0.013	0.017
45 × 45	0.61	0.042	0.006	0.017	0.065	64.5	9.4	74.1	0.013	0.017
50 × 50	0.58	0.040	0.007	0.017	0.065	62.7	11.0	73.8	0.013	0.016
55 × 55	0.58	0.040	0.007	0.017	0.065	62.7	11.3	74.1	0.013	0.016

of the same distance for the grid dimensions on the east to west and north to south sides. Because of the anisotropy in the slope and curvature data, an anisotropic grid may have been a better choice.

The regression parameters from the pure OLS model were higher than from the spatial model (Table 13.1). This is because values of WHC predicted by the spatial autoregression have been adjusted by removing the effects of neighboring values. This reduces the magnitude of WHC in the regression relationship.

Only the effect of calculating slope and tangential curvature from different scales of elevation data has been investigated. The scale of the measured WHC is the measured scale and has not been changed. Thus the change in α over the different scales reflects the effects of calculating terrain variables from grids interpolated to varying levels of detail from a given set of measured elevations. The lowest value of α was associated with the scale with the highest regression r^2 (Table 13.2). This suggests that more spatial error can be explained by stronger relationships with the regression variables. Note that the RMSE for the OLS regression varied over the scales and was generally lower when terrain variables were calculated from the more coarse scales of elevation data (Table 13.2). The RMSE for the spatial autoregression was almost constant (the differences were only seen in the fourth significant digit after the zero). This indicates that the total information content in terrain parameters and WHC was roughly the same among all the scales when spatial relationships were taken into account.

D. Interpolation and Mapping of WHC at Different Scales

The value of α and the regression coefficients determined from the terrain variables calculated from the elevation grids were used to generate contour plots of WHC at 20- × 20-m, 25- × 25-m and 30- × 30-m resolutions (Figure 13.7). Smoothness and loss of detail increase as the size of the grid on which terrain variables were calculated increases. The coarsest scale of 30 × 30 m in Figure 13.7 shows larger, connected areas of uniform water availability. WHC in all the maps shows a general pattern of bands of increasing WHC from the upper section of the field to the lower section in the direction of decreasing slope. This is in accord with the measured WHC (Figure 13.6).

Overall, as the scale of the elevation data from which the terrain variables are calculated becomes more coarse, the level of spatial autocorrelation for WHC decreases and the amount of

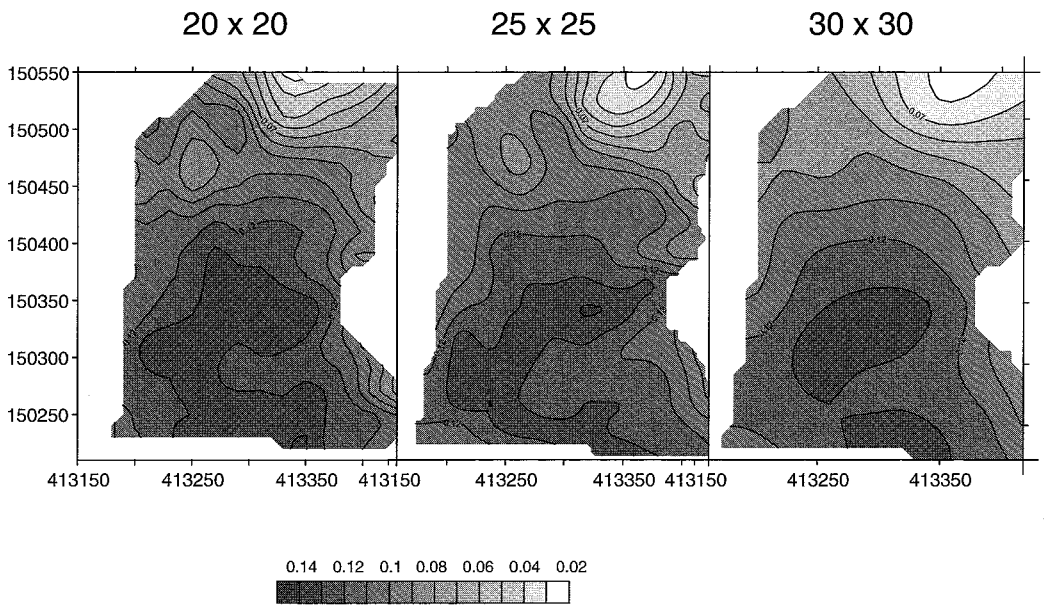

FIGURE 13.7 Contour maps of predicted and interpolated WHC at three scales.

error explained by the regression of WHC on slope and tangential curvature increases. This may be related, in part, to the aggregation of the elevations as terrain parameters are calculated from elevations averaged over larger distances, an effect termed the modifiable unit area problem.[13] The significance of the dependence of WHC on slope and tangential curvature also changes with scale. The dependence on tangential curvature is only significant at the midrange scales similar to that of the scale of the elevation data. Interpolation of tangential curvature to scales much finer than the scale at which elevations were measured probably increases the error so very little real information is related to WHC. At the coarsest scales, the effect of smoothing elevation results in a loss of detail in the tangential curvature, reducing its effect on WHC.

The dependence of WHC on slope is not greatly different among scales although its effect is least significant at the midscales, 35 × 35 m and 40 × 40 m. At these scales, the effect of tangential curvature is highly significant and the tangential curvature coefficients are largest. This suggests that tangential curvature is accounting for some of the effects of slope, especially at the 35- × 35-m scale. These results reflect the scale of the two predictors, slope and tangential curvature. At this site, slope is a large scale variable and varies slowly over large distances. Because tangential curvature is calculated as a derivative, it is sensitive to small changes in elevation and would have a smaller scale than elevation.

E. Measured Crop Yields and Interpolated WHC

The interpolated water holding capacities were compared to corn grain yields obtained from a yield monitor for the 1997 and 1998 growing seasons. Figure 13.8 shows the relationships between predicted WHC and yield. Second-order polynomial regressions were carried out for the WHC and yield relationships (Figure 13.8). These relationships predict a plateau at higher water holding capacities where increasing WHC no longer results in increasing yields. This relationship is supported by measured water holding capacities and yields at this site.[21]

Root mean square errors and r^2 values for these OLS polynomial regressions (Table 13.3) are presented for approximate comparison purposes only because they are likely to be inflated due to spatial correlation. The errors and r^2 values are largely similar among all the scales. Generally, all

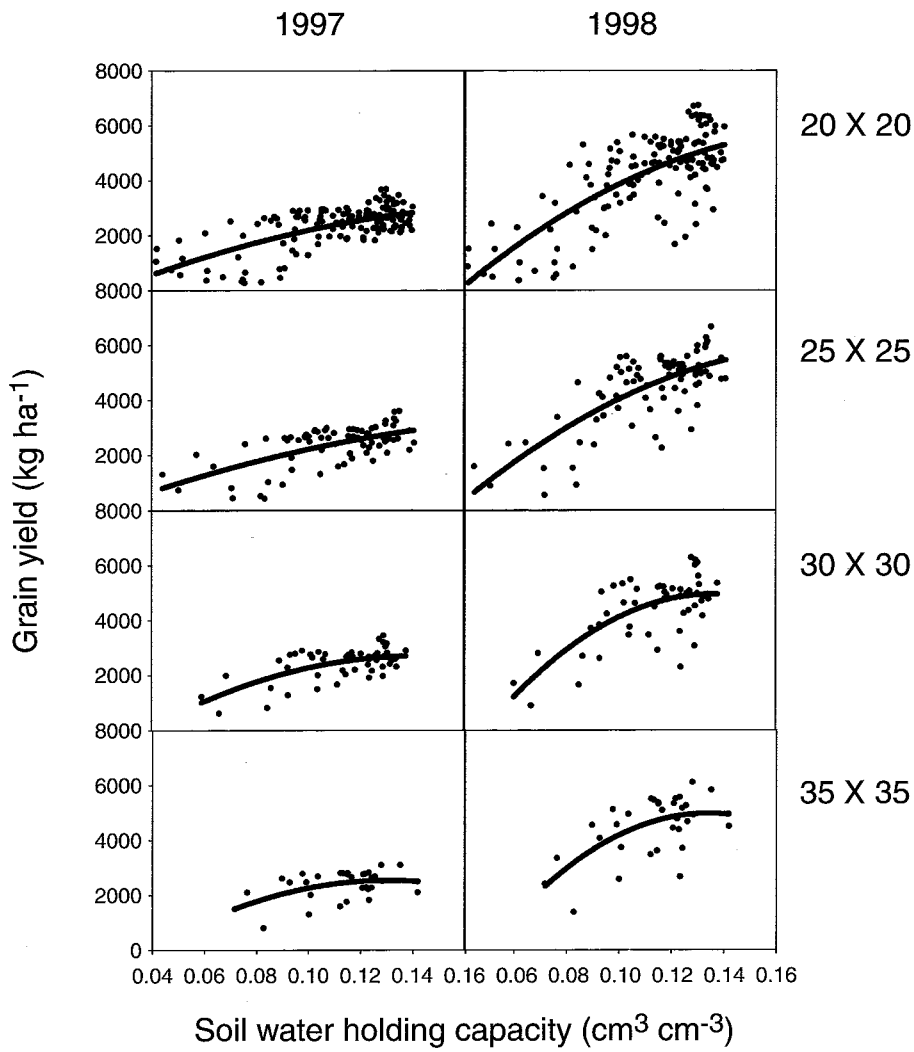

FIGURE 13.8 Relationships between predicted WHC and crop yields from 1997 and 1998 for several scales.

TABLE 13.3
Root Mean Square Errors (RMSE (cm^3cm^{-3})^2) and r^2 Values from the Second-Order Polynomial Regression between Interpolated Yield and Predicted Water Holding Capacity at Four Scales

	1997		1998	
Scale (m × m)	RMSE	r^2	RMSE	r^2
20 × 20	568.6	43.9	1037.3	54.6
25 × 25	543.9	40.3	905.0	55.1
30 × 30	474.2	39.5	900.2	46.6
35 × 35	472.8	24.4	898.9	37.5

the scales to 30 × 30 m seem to give similar results; only the 35- × 35-m scale seems very different from the others in terms of r^2 and RMSE. The minimum values and range of WHC vary by scale (Figure 13.8). The minimum values of WHC predicted at the 20- × 20- and 25- × 25-m grids were lower than in the input data. The 20- × 20- and 20- 25-m scales had the widest range of predicted WHC values and the measured WHC ranged from 0.068 to 0.19 cm^3 cm^{-3}. In spite of the extrapolation, however, the relationship of these lowest values with grain yield is consistent with the remainder of the yields at higher WHC. Others[22] have reported correlations on the order of 0.13 and 0.20 for corn and soybean yields, respectively, with soil water storage. The strength of the correlation varied by elevation.

The larger range of WHC values in the finer scales probably contributes to the higher r^2 values. Also, the r^2 value is proportional to the ratio between the amount of error explained by the model and the total error. As in the previous data, there is no clear advantage to choice of scale. When comparing the use of a single scale of yield data to scaling the yield data to the WHC scale, the latter method appears to result in lower RMSE over all the scales.

A water budget model has been used[21] to estimate soil water availability from yield map data collected after seasons with below average rainfall. The correspondence between grain yield and WHC found here is in agreement with the results of that study. This also suggests that the optimization method used[21] could be improved by including terrain variables and using some method of stochastic simulation to distribute the optimized water holding capacities spatially in the landscape.

IV. CONCLUDING REMARKS

Water holding capacities in the upper 10 cm of soil were sampled on transects and a grid on a 6-ha field. The objective of the study was to interpolate these sparsely sampled data to a more dense grid using terrain variables calculated from more readily available elevation data. Terrain variables (slope, tangential curvature and plan curvature) were calculated from elevations interpolated to scales from 5 × 5 m to 55 × 55 m. Spatial autoregression was used to predict WHC at the measured locations using calculated terrain parameters as predictors. Slope and tangential curvature were found to be significant predictors of surface WHC. The spatial autoregression parameter, α, explained 60 to 70% of the variance in the relationship between terrain variables and WHC. The 25- × 25-m scale of terrain variables gave the best fit. At finer scales, there was too much noise and at more coarse scales too much smoothing of the terrain attributes. The water holding capacities were predicted on regular grids of 20- × 20-, 25- × 25-, 30- × 30- and 35- × 35-m dimensions and compared with crop yields measured with a yield monitor and interpolated to the same grids using block kriging. There was a good correspondence between predicted WHC and measured corn grain yields.

Prediction of WHC to fill in "holes" using spatial autoregression is similar to forecasting time series analysis. Error is minimized by using neighboring values to better estimate WHC at a specific location.[15] In this sense, local information is taken into account. Estimated points that are further apart would have a higher prediction error than points closer together. Other geostatistical methods, such as co-kriging, have been used to relate two variables spatially, but this requires that the semivariograms for the variables be modeled as one semivariogram.[14] One advantage of spatial autoregression is that it is primarily a statistical tool as opposed to geostatistics where the principal goal is to produce a generalized map surface (Griffith and Layne,[15] page 469). Both methods, however, seek to quantify spatial autocorrelation and so have many similarities[15] describing linkages between autoregressive and semivariogram models (Griffith and Layne,[15] page 469). As shown in this chapter, spatial autoregression is a useful tool to develop prediction equations for spatial data.

Water holding capacities used in spatial models are often estimated from soil maps and soil texture.[23] Although textural components, as well as soil map units, are closely related to topography, the estimated water holding capacities may not be realistically spatially distributed. This spatial

distribution can be better accounted for by distributing the water holding capacities as a function of topography. The WHC predicted from terrain attributes using spatial autoregression can be used to generate a map of WHC. Such a map will be useful to develop management zones for this field or to use as input in crop models.

V. ACKNOWLEDGMENTS

The authors wish to express their gratitude to Dr. Kelley Pace and Dr. Jhonathan Ephrath, who provided very helpful suggestions for early drafts of the paper, and to the anonymous reviewers for their highly insightful comments and suggestions. These greatly improved this chapter.

REFERENCES

1. Yeh, P.J.F. and Eltahir, E.A.B., Stochastic analysis of the relationship between topography and the spatial distribution of soil moisture, *Water Resour. Res.,* 34, 1251, 1998.
2. Western, A.W., Grayson, R.B., Bloschl, G., Willgoose, G., and McMahon, T.A., Observed spatial organization of soil moisture and its relation to terrain indices, *Water Resour. Res.,* 35, 797, 1999.
3. Tomer, M.D. and Anderson, J.L., Variation of soil water storage across a sand plain hillslope, *Soil Sci. Soc. Am. J.,* 59, 1091, 1995.
4. Hanna, A.Y., Harlan, P.W., and Lewis, D.T., Soil available water as influenced by landscape position and aspect, *Agron. J.,* 74, 999, 1982.
5. Moulin, A.P., Anderson, D.W., and Mellinger, M., Spatial variability of wheat yield, soil properties and erosion in hummocky terrain, *Can. J. Soil Sci.,* 74, 219, 1994.
6. Sudduth, K.A., Drummond, S.T., Birrell, S.J., Kitchen, N.R., and Stafford, J.V., Spatial modeling of crop yield using soil and topographic data, in *Precision Agriculture '97. Volume I. Spatial Variability in Soil and Crop.* Papers presented at the First European Conference on Precision Agriculture, Warwick University, UK, 7-10 September, 1997, 439–447.
7. Pachepsky, Ya., Timlin, D.J., and Rawls, W.J., Soil water retention as related to topographic variables, *Soil Sci. Soc. of Am. J.,* 65, 1787, 2001.
8. Timlin, D.J., Pachepsky, Ya., Snyder, V., and Bryant, R.B., Spatial and temporal variability of corn yield on a hillslope, *Soil Sci. Soc. Am. J.,* 62, 764, 1998.
9. Sinai, G., Zaslavsky, D., and Golany, P., The effect of soil surface curvature on moisture and yield — Beer Sheba observation, *Soil Sci.,* 132, 367, 1981.
10. Moore, I., Gessler, P.E., Nielsen, G.A., and Peterson, G.A., Soil attribute prediction using terrain analysis, *Soil Soc. Am. J.,* 57, 443, 1993.
11. Bouma, J., Using soil survey data for quantitative land evaluation, *Adv. Soil Sci.,* 9, 177, 1989.
12. Pachepsky, Ya., Rawls, W.J., and Timlin, D.J., The current status of pedotransfer functions: their accuracy, reliability, and utility in field- and regional-scale modeling, in *Assessment of Non-point Source Pollution in the Vadose Zone,* Corwin, D.L., League, K., and Ellsworth, T.R., Eds., American Geophysical Union, Washington, D.C., Geophys. Monograph, 108: 223-234, 1999.
13. Long, D.S., Spatial autoregression modeling of site-specific wheat, *Geoderma,* 85, 181, 1998.
14. Goovaerts, P., Using elevation to aid the geostatistical mapping of rainfall erosivity, *Catena,* 34, 227, 1999.
15. Griffith, D.A. and Layne, L., *A Casebook for Spatial Statistical Data Analysis. Spatial Information Systems.* Oxford University Press, 1999, 506 pps.
16. Mitásova, H. and Hofierka, J., Interpolation by regularized spline with tension: II. Application to terrain modeling and surface geometry analysis, *Math. Geol.,* 25, 657, 1993.
17. Griffith, D.A., Estimating spatial autoregressive model parameters with commercial statistical software, *Geogr. Anal.,* 20, 176, 1988.
18. Griffith, D.A., A numerical simplification of estimating parameters of spatial autoregressive models, in *Spatial Statistics: Past, Present, and Future,* Griffith, D.A., Ed., Institute of Mathematical Geography. Ann Arbor, MI, 1990, 185–196.

19. Pace, K.R. and Barry, R., Quick computation of regressions with a spatially autoregressive dependent variable, *Geogr. Anal.*, 29, 232 ,1997.
20. Pace, K.R. and Gilley, O.W., Using the spatial configuration of the data to improve estimation, *J. Real Estate Finance Econ.,* 14, 333, 1997.
21. Timlin, D.J., Pachepsky, Ya., Walthall, C.L., and Loechel, S.E., The use of a water budget model and yield maps to characterize water availability in a landscape, *Soil Till. Res.,* 58, 219, 2001.
22. Logsdon, S., Pruger, J., Meek,, D., Colvin, T., Jaynes, D., and Milner, M., Crop yield variability as influenced by water in rain-fed agriculture, in *Proceedings of the Fourth International Conference on Precision Agriculture, St. Paul, Minnesota, USA, 19-22 July 1998. Part A and Part B.* Robert, P.C., Rust, R.H., and Larson, W.E., Eds., 1999, 453-465.
23. Turner, D.P., Dodson, R., and Marks, D., Comparison of alternative spatial resolutions in the application of spatially distributed biogeochemical model over complex terrain, *Ecol. Model,* 90, 53, 1996.

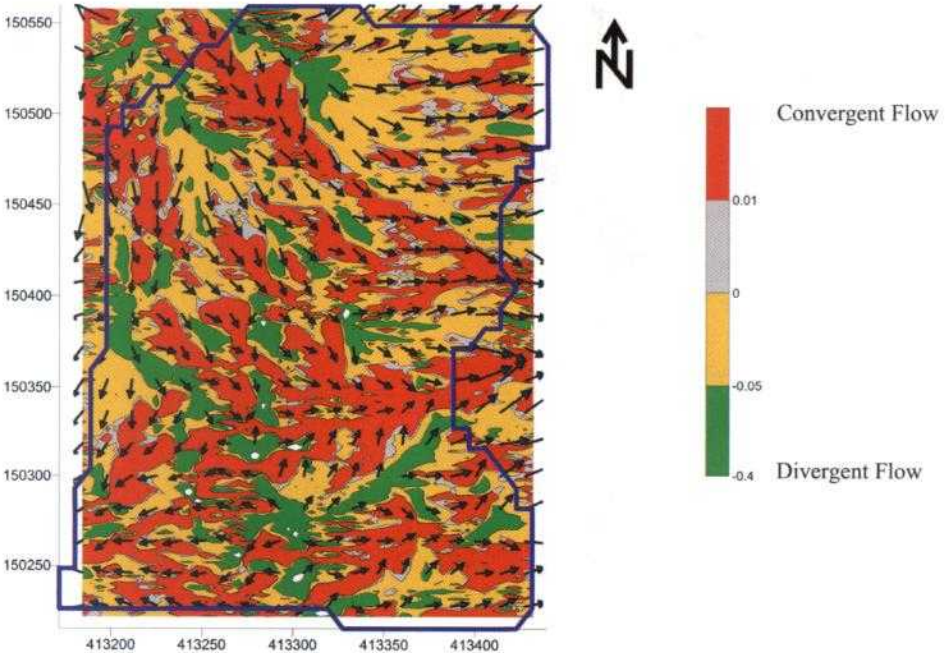

FIGURE 13.3 Tangential curvature and water flow vectors calculated from a 5- × 5-m DEM grid. The field boundary is shown in blue.