Terrain analysis in support of precision farming
by Damian Jeremiah Spangrud

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Earth Sciences
Montana State University

Abstract:
Precision farming refers to the practice of "Farming Soils Not Fields". This concept means that a farm field is not thought of as a homogenous unit but as series of separate management units based on soil attributes. The delineation of these management units requires precise knowledge of the soil across the field, to gain this knowledge GPS point sampling and soil interpretation have been used. This study has two aims: (1) to quantify the relationship between sampling methods and the resultant surface; and (2) to identify relationships between measured soil attributes and selected terrain and image attributes.

GPS point data were obtained for 6,284 locations across a 20 ha farm field in southwestern Montana. These data were sampled using stratified random areal and linear sampling techniques. Elevation surfaces were then interpreted and terrain attribute surfaces derived from these samples using ANUDEM and TAPES-G. These surfaces were then compared against each other graphically and statistically. A series of maps and tables show that the stratified random areal samples produced very good results (low RMSE and Moran's I) with much fewer sample points than the linear samples.

Three soil attributes (percent organic matter (OM), depth of mollic epipedon, and pH) were modeled across the same farm field using remotely sensed imagery and derived terrain surfaces. Two spectral band ratios explained 64% of the variation in OM. Three terrain attributes (Wetness index, slope gradient, and plan curvature) explained 48% of the same variation. A combination of the imagery and terrain data explained 70% of the variance in OM using two spectral band ratios, specific catchment area, and wetness index at the 66 sampling sites. Wetness index, slope gradient, and plan curvature combined to explain 48% of the variation in mollic epipedon thickness. Elevation and wetness index combined to explain just 13% of pH.

TERRAIN ANALYSIS IN SUPPORT OF PRECISION FARMING

by

Damian Jeremiah Spangrud

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Earth Sciences

Montana State University
Bozeman, Montana

December 1995

APPROVAL

of a thesis submitted by

Damian Jeremiah Spangrud

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency and is ready for submission to the College of Graduate Studies.

3/6/96
Date

John P. Wilson
Chairperson, Graduate Committee

Approved by Major Department

3-11-96
Date

Dave Cogs
Head, Major Department

Approved for the College of Graduate Studies

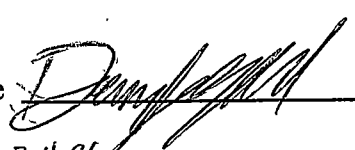
4/1/96
Date

R R Bowen
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under the rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Signature 

Date 3-4-96

TABLE OF CONTENTS

Page

1. INTRODUCTION 1

2. SENSITIVITY OF COMPUTED TERRAIN ATTRIBUTES TO THE NUMBER AND PATTERN OF GPS-DERIVED ELEVATION DATA 5

Introduction 5
Materials and Methods 9
Study Area 9
GPS Data Collection 12
GPS Point Sampling 12
Data Analysis Environment 18
Generation of Terrain Attributes 19
ANUDEM 20
TAPES-Grid 20
Data Analysis and Display 23
Results and Discussion 24
Stratified random areal Samples 26
Linear Samples 33
Topographic Attributes 36
Summary 40

3. INTERPOLATION OF SOIL ATTRIBUTE SURFACES USING SOIL-LANDFORM RELATIONSHIPS 42

Introduction 42
Literature Review 43
Materials and Methods 48
Soil Sampling 48
Generation of Terrain Attributes 49
Collection and Analysis of Remotely Sensed Data 53
Soil Survey Data 54

TABLE OF CONTENTS-Continued

Statistical Analysis and Data Display 55
Results and Discussion 56
Summary 62

4. CONCLUSIONS 64

REFERENCES CITED 67

LIST OF TABLES

Table		Page
1.	Primary topographic attributes that can be computed by terrain analysis from DEM data. (Adapted from Moore et al 1991a, 1993a)	10
2.	Stratified random areal sampling schemes and number of points sought and actually selected	16
3.	Spatial analysis of elevation differences from stratified random area sample points	30
4.	Spatial analysis of elevation differences from linear routes simulating data from a truck-mounted receiver	30
5.	Spatial analysis of topographic attribute differences from stratified random samples	37
6.	Arithmetic means for selected topographic attributes	37
7.	Spatial analysis of topographic attribute differences from linear samples	39
8.	Soil attribute values for the four map units and the entire field represented in the order 1 soil survey map	57
9.	Multiple regression results	58

LIST OF FIGURES

Figure — Page

1. Location map of study site, site is shown by the box in the shaded relief illustration. 11

2. Three-dimensional view of the study site from the south looking north. 13

3. GPS points collected by Tyler et al. The 6,284 points appear as lines at the scale of this figure.. 14

4. Example of stratified random areal sample 15

5. Map of GPS points used in this study. The 6,284 points appear as lines at the scale of this figure. 17

6. Contour map of study area 25

7. Random stratified samples of (a) 100 meter sampling unit with one point, (b) 100 meter sample with 2 points, and (c) a 100 meter sample with 4 points 27

8. Random stratified samples of (a) 50 meter sampling unit with one point, (b) 50 meter sampling unit with two points, and (c) 50 meter sampling unit with four points. 28

9. Random stratified samples of (a) 50 meter sampling unit with sixteen points, (b) 25 meter sampling unit with one points, and (c) 25 meter sampling unit with four points. 29

10. Linear samples of (a) every other north-south route, (b) every other west-east route, (c) combination of every other north-south route and every other west-east route. 34

LIST OF FIGURES-Continued

11. Linear samples of (a) every north-south route, (b) every west-east route, (c) combination of every north-south route and every west-east route. 35

12. Hypothetical soil variation (from Burrough et al. 1994). Variability (z) verses distance(x) 45

13. Map of 70 soil sample locations and mapped soils 50

14. Resultant maps from regression analysis for %OM: a) Inverse Distance Weighting surface (IDW); b)image data regression; c) terrain data regression; and d) terrain and image data regression. 61

Abstract

Precision farming refers to the practice of "Farming Soils Not Fields". This concept means that a farm field is not thought of as a homogenous unit but as series of separate management units based on soil attributes. The delineation of these management units requires precise knowledge of the soil across the field, to gain this knowledge GPS point sampling and soil interpretation have been used. This study has two aims: (1) to quantify the relationship between sampling methods and the resultant surface; and (2) to identify relationships between measured soil attributes and selected terrain and image attributes.

GPS point data were obtained for 6,284 locations across a 20 ha farm field in southwestern Montana. These data were sampled using stratified random areal and linear sampling techniques. Elevation surfaces were then interpreted and terrain attribute surfaces derived from these samples using ANUDEM and TAPES-G. These surfaces were then compared against each other graphically and statistically. A series of maps and tables show that the stratified random areal samples produced very good results (low RMSE and Moran's I) with much fewer sample points than the linear samples.

Three soil attributes (percent organic matter (OM), depth of mollic epipedon, and pH) were modeled across the same farm field using remotely sensed imagery and derived terrain surfaces. Two spectral band ratios explained 64% of the variation in OM. Three terrain attributes (Wetness index, slope gradient, and plan curvature) explained 48% of the same variation. A combination of the imagery and terrain data explained 70% of the variance in OM using two spectral band ratios, specific catchment area, and wetness index at the 66 sampling sites. Wetness index, slope gradient, and plan curvature combined to explain 48% of the variation in mollic epipedon thickness. Elevation and wetness index combined to explain just 13% of pH.

CHAPTER 1

INTRODUCTION

Precision farming refers to the practice of "Farming Soils Not Fields" (Carr et al., 1991). This concept means that a farm field is not thought of as a homogenous unit but as a series of separate management units based on soil attributes. Each mapped unit can then be managed independently of the others. Precision farming is an "old" idea which has become feasible again notwithstanding the large farm consolidations of the past fifty years. Prior to this recent period, farms were generally small enough for soil variability to be managed according to a farmer's "working knowledge" of how specific crops responded to specific soil environments across the field. Today's farms are too large to implement variable management based on "working knowledge" (Schueller et al., 1992). There can be large variations in the land resources within fields which result in crop yield variations across fields when fertilizers and pesticides are applied at a constant rate. Robert et al. (1990), Buchholz (1991), Carr et al. (1991), Wibawa (1991), and others have demonstrated the potential for using global positioning system technology (GPS) and variable rate technology (VRT) to manage for soil variability on the large farms that now dominate North American agriculture.

Soils and their variability are traditionally determined from existing USDA natural

Resource Conservation Service (NRCS) (formally the Soil Conservation Service (SCS)) soil maps; however, these maps are not produced with the locational accuracy required for variable rate applications (such as precision farming) and they do not show the full extent of the soil variability within a mapped soil (Mausbach et al., 1993). Soil scientists have proposed the use of pedo-transfer functions and geostatistics to develop soil attribute surfaces across and within mapped soil areas (e.g., Webster, 1977; Odeh et al. 1991; Burrough, 1993). Such systems use soil attribute information as the primary determinants of the resulting soil surfaces. Others have proposed using terrain attributes in conjunction with soil attributes to achieve the same end result (e.g., Klingebiel et al., 1987; Odeh et al., 1991; Moore et al., 1993a, 1993b). The terrain analysis approach requires digital elevation models (DEMs) from which terrain attributes are computed. The most widely available DEMs (1:250,000 and 1:24,000 scales) suffer from the same scale problems as current soil maps, although GPS receivers can provide rapid and relatively inexpensive data acquisition for the creation of larger scale DEMs (Tyler, 1993).

Precision farming uses VRT to alter application rates across the field, so that application rate matches the land resource base in that area of the field. Areas of the field which are traditionally over-applied with single application rates will receive decreased inputs which better match crop needs and minimize contamination of water resources. Areas which were under-applied with single application rates will have application rates increased to better address the needs of that area. VRT technology helps in improve the crop return per dollar spent on production. This matching of application rates and needs is the principal objective of precision farming.

The implementation of precision farming requires specific information about the soil and terrain attributes across the field and at the current location of the applicator. Derived terrain attributes can be used to help identify soil attribute variability beyond that of traditional SCS soil maps, and thus increase the level of accuracy of variable rate fertilizer and pesticide applications (Moore et al., 1993a, 1993b). Knowing the location of the applicator, by using a GPS receiver, takes much of the guesswork in the location of the applicator out of the "working knowledge" concept. When pre-existing field information is combined with locational data, the "working knowledge" concept is now quantified and repeatable.

Precision farming makes use of some of the leading technology of the day, but to what end? Macy Farms in Illinois have used VRT on 809.4 ha (2023 acres) since 1991 and reduced their total costs by $33.77 per hectare (Macy, 1993). Robert et al. (1990) and Carr et al. (1991) conducted multiple plot tests to evaluate the benefits of precision farming. Their results varied across the plots, but the overall profits and yields did not decrease with the new precision farming techniques. There may be additional benefits that should be considered, such as reduced water quality impacts and reduced environmental degradation, when the economic benefits do not match the additional costs of implementation of precision farming.

Precision farming faces many challenges in the future and there are a number of important questions that still have to be answered. The technology required to practice precision farming is evolving rapidly. VRT is advancing quickly as the agricultural community discovers the potential economic benefits of such a system and GPS technology

is advancing at an astonishing rate as receivers which used to be vehicle- mounted and very complex can now be carried in the palm of your hand and purchased cheaply from numerous vendors. One of the most important questions, yet to be answered, has to do with the division of fields into management units. Soils will help contribute to these decisions but in what manner and what other components are be to considered are some of the important research issues that still need to be addressed.

The following chapters address two research questions related to precision farming. As GPS data is now being used to create DEMs and delineate management units, it is important to evaluate the effects that the location and pattern of GPS data can have on the resultant terrain attributes. These attributes (elevation, specific catchment area, and slope) may represent important parameters for characterizing soil-landscapes. Soil scientists have proposed numerous methods for defining terrain and soil-landscape interactions ranging from geostatistical to classic statistical techniques. These interactions are very sensitive to the modeling assumptions and interpolation schemes used to create attribute data and further work is needed to characterize these relationships. This study, therefore, addresses two sets of issues: (1) the sensitivity of the terrain attributes computed from DEMs to the number and pattern of input GPS data, and (2) the contribution of these terrain attributes and remotely sensed imagery in the prediction of soil properties throughout the field. The effect which alterations to the number and pattern of input GPS data have on the resultant DEM are explored using several sampling strategies. The soil-landscape modeling is accomplished through the use of statistical and geographical analysis of predicted values and their associated errors.

CHAPTER 2

SENSITIVITY OF COMPUTED TERRAIN ATTRIBUTES TO NUMBER AND PATTERN OF GPS-DERIVED ELEVATION DATA

Introduction

The development of digital elevation models during the past decade has encouraged research into soil attribute prediction using many complex and quantifiable soil-landform relationships. Obtaining DEMs at the appropriate scale is difficult. DEMs are available for most of the continental United States at a scale of 1:24,000 (30 meter resolution) and the entire continental United States at a scale of 1:250,000 (3 arc second resolution). Soil maps are currently produced at scales of 1:12,000, 1:15,840, 1:20,000, and 1:24,000 with a minimum mapping unit of 0.8 to 4 ha (2 to 10 acres) (Mausbach et al., 1993). DEMs at soil map scales or larger are needed for soil-terrain modeling to help improve soil attribute prediction. The use of GPS technology in conjunction with surface interpolation algorithms has been proposed as a relatively inexpensive and quick method of obtaining larger scale (higher resolution) DEMs. The effect which the number and pattern of GPS points used in this interpolation has on the resultant DEM and landscape attributes is the focus of this chapter.

GPS technology was initially developed for the United States military in the late 1970s to assist with the locating and positioning of strategic forces. Currently there are 24 GPS satellites in orbit and their orbits are configured so at least three of the satellites are visible to a receiver at any given time. The process of obtaining locational data from a GPS relies on the triangulation of at least three points (satellites). The data received from each of these satellites is validated against the others to correct for any instrumental errors. A stationary receiver used in conjunction with a mobile receiver provides the most accurate measurements, as the fixed receiver provides an additional measurement of satellite data error. GPS satellites are U.S. Department of Defense owned and operated, and certain bands (P and Y) are scrambled to limit civilian accuracy to the tens of meters level for security reasons (Tyler, 1993). However, many manufacturers have developed ways around this "problem" and they continue to obtain sub-meter accuracy using a combination of the non-encrypted signals and real time error correction.

The GPS data can be interpolated from irregular spot heights to a regularly spaced elevation surface using several different interpolation algorithms. There are many grid interpolation algorithms in use today and these range from simple nearest-neighbor calculations to complex thin-plate spline algorithms. The most commonly used algorithms are those which attempt to minimize the variance across the surface and eliminate pits which may result from coarse surface interpolation (with the assumption pits are rare in nature and occur mainly in areas of recent glaciation or karst topography). Four major types of interpolation algorithms have been proposed: (1) local interpolation methods, (2) moving averages, (3) kriging and (4) partial thin-plate (or Laplacian) smoothing splines (Moore and

Hutchinson, 1991; Hutchinson and Gessler, 1994).

Local interpolation techniques apply simple fitted functions across small overlapping subsets of the entire dataset. The resultant values from this technique are quite sensitive to the spacing of the input data points, especially in cases of irregular data sets, and may result in the generation of numerous spurious pits in the data. Much of this sensitivity is due to the rigidness and local area focus of the functions. The functions can become quite complex and indeterminate when applied to more complex datasets and they cannot be easily adapted to the smoothing of noisy data (Hutchinson, 1989; Moore and Hutchinson, 1991).

Moving averages are often used for surface interpolation. The method is simplistic in nature in that the value at a given location is assumed to be a distance weighted average of surrounding data values. The distance weighting function uses a defined search radius to locate the values for this weighting. Data sets with sparse or irregular data used with this technique have the problem that the search radius may not include the required data for an accurate interpolation. The smoothing of data values becomes a concern in areas of sparse data when the search radius is defined as a very large region (Moore and Hutchinson, 1991).

Kriging uses a modification of the weighted-moving average, with the weights now computed to minimize the variance across the surface. This method assumes positive spatial autocorrelation between data. The variance used in kriging is a function of the semivariogram and the separation of the data points. Semi-variograms are plots of the total variance across the surface against distance from any given location on the surface. A function is fitted to this computed semivariogram that, in the best case scenario, defines the best fit curve to the semivariogram to minimize the interpolated variance. The elevation for

the interpolated locations is computed as a result of the moving average using the fitted semi-variogram function as the weighting factor (Isaaks and Srivastava, 1989; Cressie, 1991; Moore and Hutchinson, 1991). The selection and definition of a optimal semi-variogram function and associated parameters by the user may be difficult because real world data seldom conform to simple functions, and kriging's sensitivity to anomalous data may cause additional complications.

Partial thin plate (Laplacian) splines are formally related to kriging but do not require the computation or the imposition of semi-variograms. Smoothing parameters are used which determine the trade-off between data fidelity and smoothing. The smoothing parameters are computed for each location by minimizing the generalized cross validation (GCV). The GCV is a measure of the predictive error of the fitted function. It is calculated by systematically removing each point used in fitting the initial function and refitting the function and measuring the change in the predicted value. These parameters allow for the data to be interpolated while eliminating anomalous data. Partial thin plate splines have excellent functionality in their automatic computation of the smoothing parameters and the model fitting. Kriging and partial thin plate splines compare well when the appropriate semi-variograms are used in the kriging process (Cressie, 1991; Moore and Hutchinson, 1991). The main disadvantage of splines is their complexity and the need for elaborate computer programs to generate the functions.

The surfaces (DEMs) which result from these interpolation methods provide the basis for the calculation of topographically important attributes. Dikau (1989) summarized the landform attributes which can be computed from DEMs, and proposed a method for dividing

landscapes into relief units based on slope, plan curvature, and profile curvature attributes. Odeh et al. (1991) added upslope distance and upslope area to Dikau's list and found that these accounted for the largest amount of soil variability within their study area. Moore et al. (1991) defined these primary terrain attributes and their significance (Table 1) and Moore et al. (1993a, 1993b) related soil variability to these primary attributes as well as several secondary terrain attributes (wetness index, stream power index, and transport capacity index).

Materials and Methods

Study Area

The study area is located on a farm in the Gallatin Valley of southwest Montana. Data were collected over a 20.23 ha (40 acre) portion of a field owned by Bill Wright and located near the community of Springhill, Montana (T1N R6E Sec 8; Figure 1). The Wright farm is located at the base of the Bridger Mountains (2,700-2,950 m) which run to the north and east of the farm. The site itself has a general southerly aspect, moderate relief (43 m) and an average elevation of approximately 1,509 m. A small intermittent stream runs through the field in a south-south-westerly direction (Figure 2). The soils in the field are complexes of fine-loamy, mixed mineralogy Pachic and Udic Haploborolls and Argiborolls. The field has been farmed for about 50 years with a grain-fallow rotation.

Table 1. Primary topographic attributes that can be computed by terrain analysis from DEM data.(adapted from Moore et al., 1991, 1993d)

Attribute	Definition	Significance
Altitude	Elevation	-Climate, vegetation, potential energy
Upslope height	Mean height of upslope area	-Potential energy
Aspect	Slope azimuth	-Solar insolation, evapotranspiration, flora and fauna distribution and abundance
Slope	Gradient	-Overland and subsurface flow velocity and runoff rate, precipitation, vegetation, geomorphology, soil water content, land capability class
Upslope slope	Mean slope of upslope area	-Runoff velocity
Dispersal slope	Mean slope of dispersal area	-Rate of soil drainage
Catchment slope	Average slope over catchment	-Time of concentration
Upslope area	Catchment area above a short length of contour	-Runoff volume, steady state runoff rate
Dispersal area	Area downslope from a short length of contour	-Soil drainage rate
Catchment area	Area draining to a outlet	-Runoff volume
Specific catchment area	Upslope area per unit width of contour	-Runoff volume, steady-state runoff rate, soil characteristics, soil water content, geomorphology
Flow path length	Maximum distance of water flow to a point in the catchment	-Erosion rates, sediment yield, time of concentration
Upslope length	Mean length of flow paths to a point in the catchment	-Flow acceleration, erosion rates
Dispersal length	Distance from a point in the catchment to the outlet	-Impedance of soil drainage
Catchment length	Distance from the highest point to the outlet	-Overland flow attenuation
Profile curvature	Slope profile curvature	-Flow acceleration, erosion/ and deposition rate, geomorphology
Plan Curvature	Contour curvature	-Converging/diverging flow, soil water content, soil characteristics

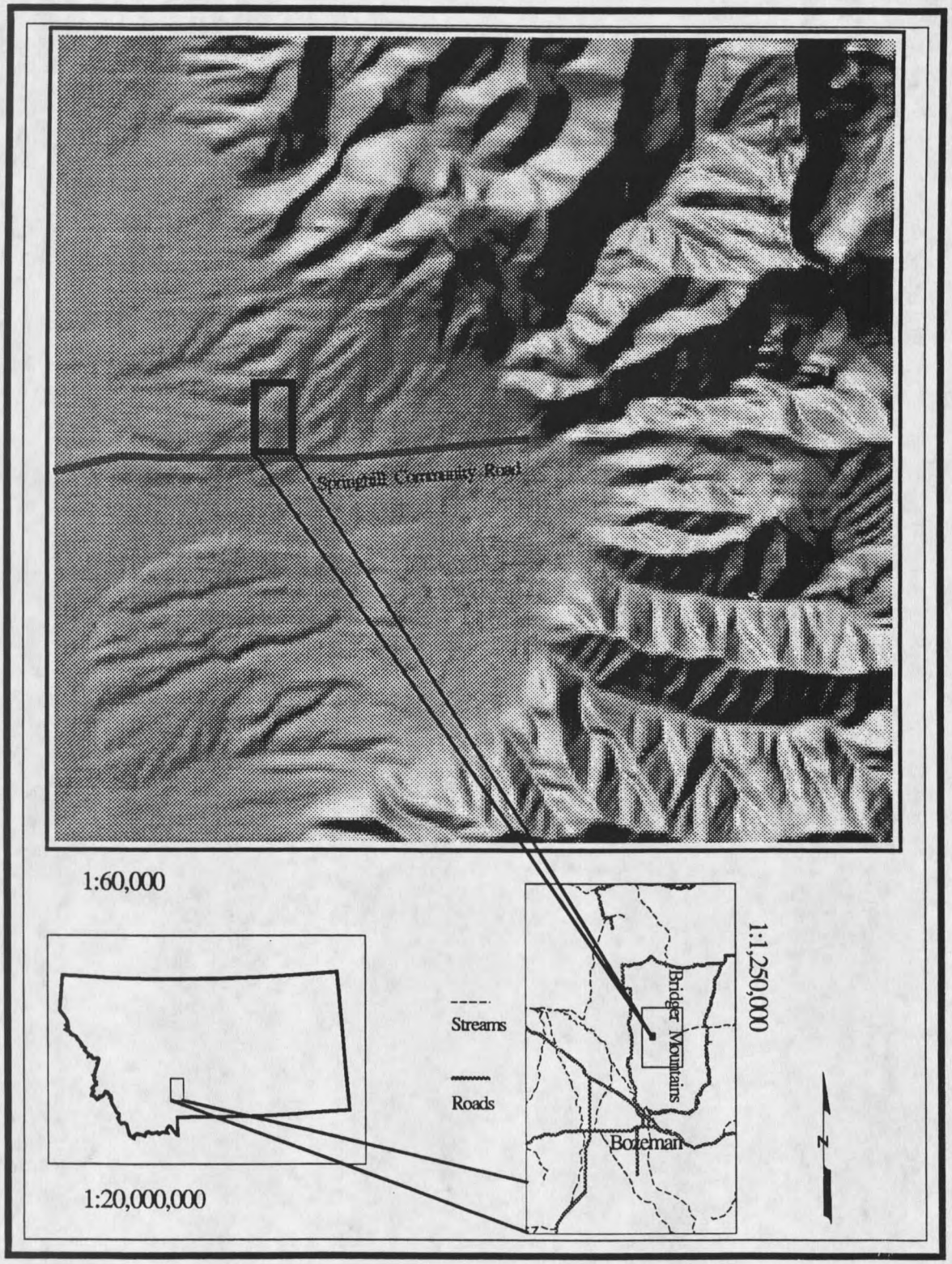


Figure 1. Location map of study site, site is shown by the box in the shaded relief illustration.

GPS Data Collection

Positional and elevation data were collected at 6,284 unique coordinate locations in the spring of 1992 using an Ashtech P-12 GPS receiver in kinematic mode as the base and an Ashtech Sensor receiver as the mobile unit (points appear as lines dues to scale of the map in Figure 3) (D.A. Tyler, 1992, pers. comm.). The data collected included longitude and latitude and elevation in meters above mean sea level. Positional and vertical accuracy for the data were estimated at less than 1 m. The data were collected as part of a larger study examining the potential nitrate contamination of ground water from agricultural applications.

GPS Point Sampling

Sampling schemes were devised so that the collected data could be used to explore the advantages and disadvantages of varying the number and pattern of GPS data. Two sampling schemes were explored, one method is a variation of the stratified random areal sample described by Berry and Baker (1968) and the other strategy selected GPS sample points along linear paths. The latter sample typifies how data would be collected by agricultural managers using vehicle-mounted GPS receivers while the former approach (at low sampling rates) can be thought of as point sampling with a hand-held GPS receiver.

The stratified random areal sampling method places a fishnet mesh over the entire data collection area and a user-defined number of points are randomly selected within each mesh area (sampling unit). Figure 4a shows the selection of random points using the standard stratified random areal sample at a rate of one point per mesh unit. For this study

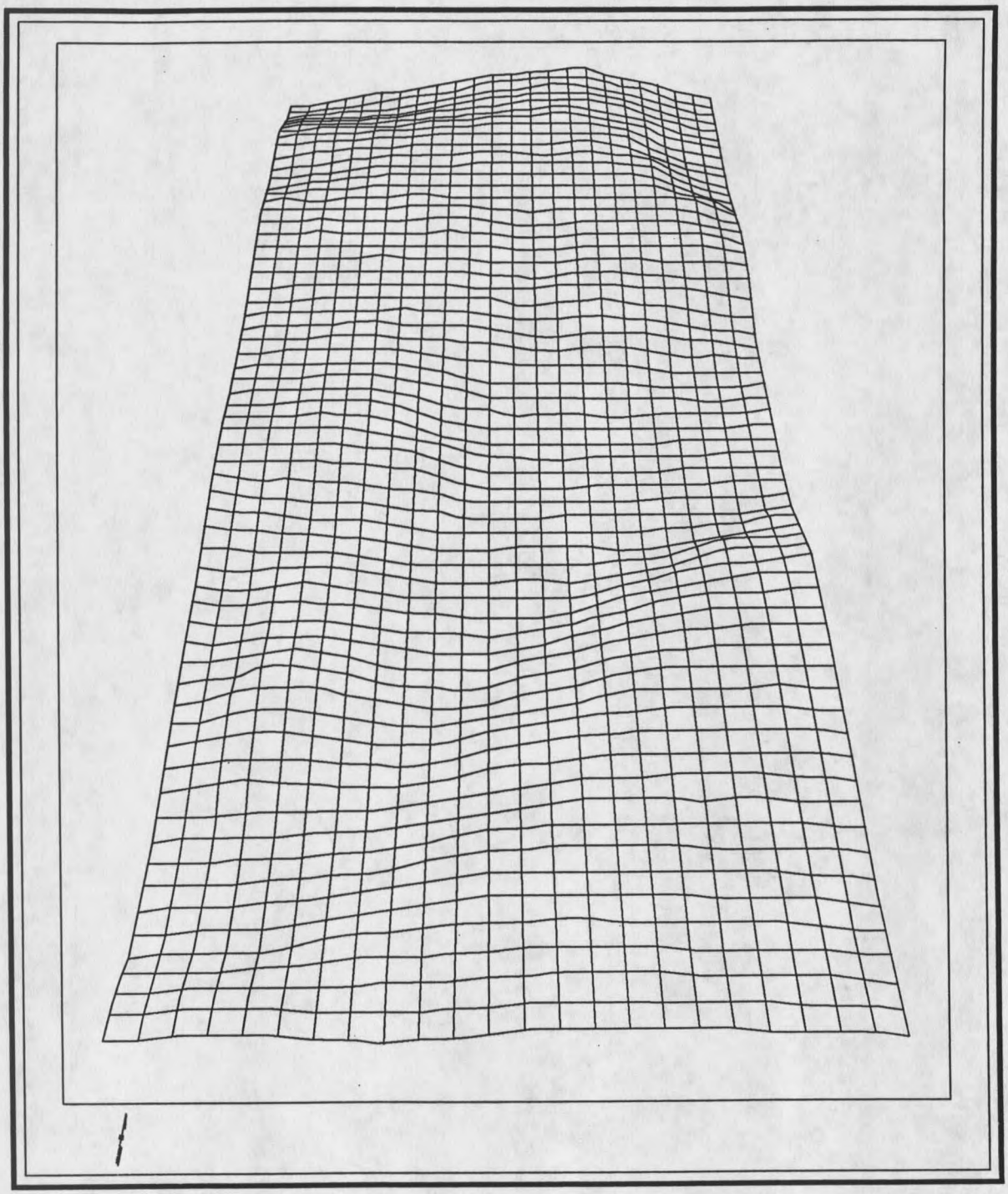


Figure 2. Three-dimensional view of the study site from the south looking north with a vertical exaggeration of two.

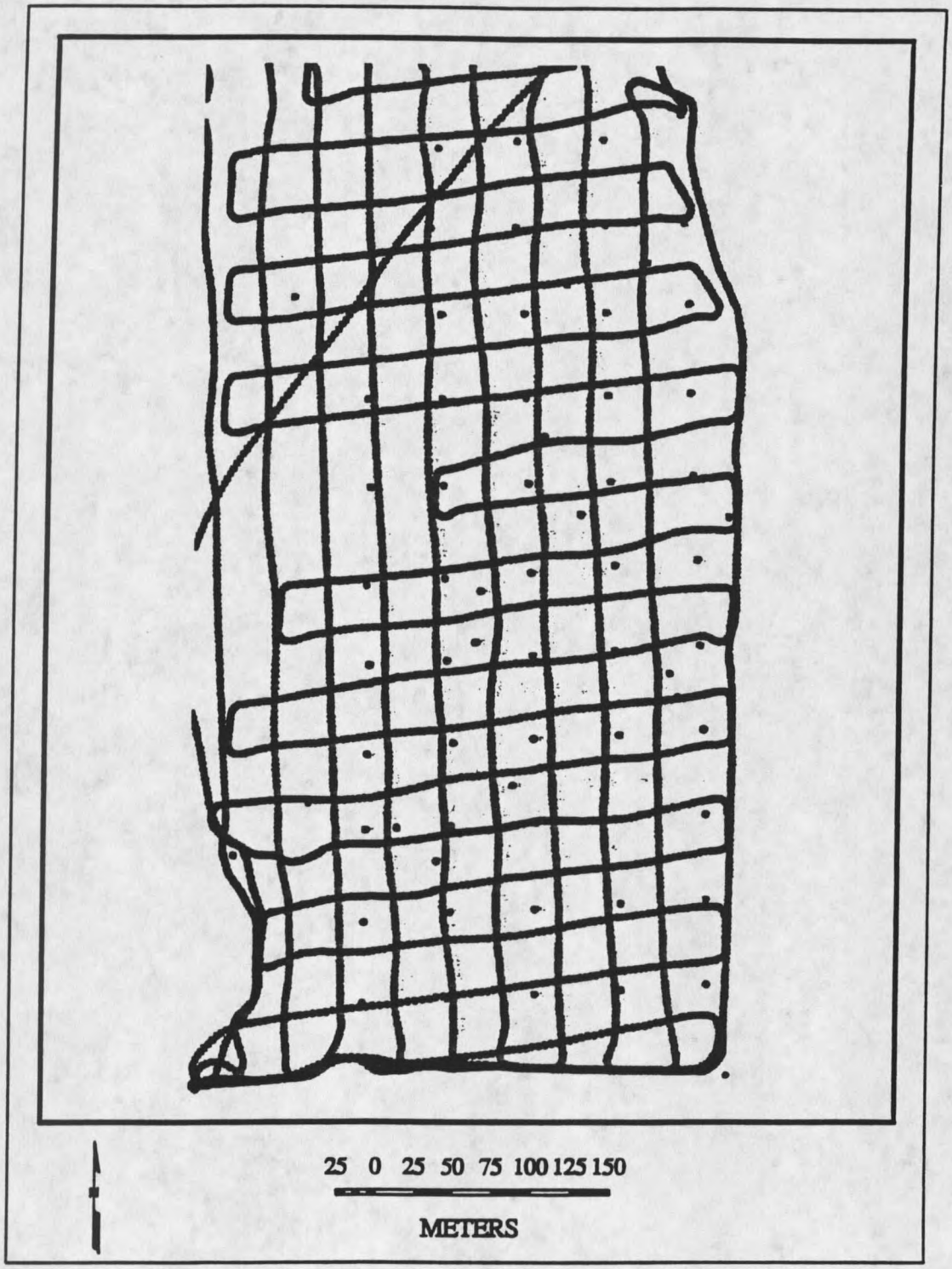


Figure 3. GPS points collected by Tyler et al. for study site. The 6,284 points appear as lines at this map scale with isolated points visible as separate points.

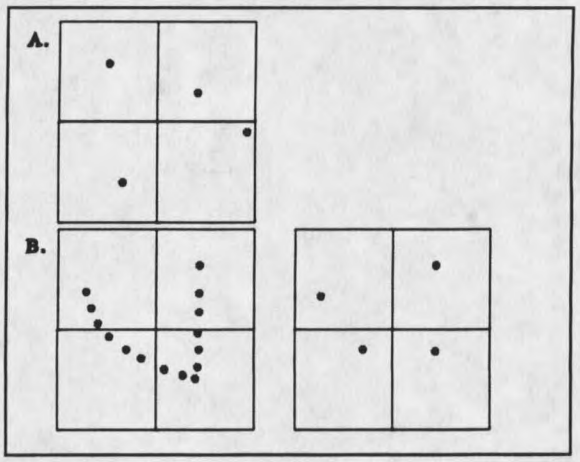


Figure 4. Schematic showing stratified random areal sample method.

the selected points are restricted to a random subset of the pre-existing GPS points. Figure 4b shows how this sampling strategy was used with the existing data. The left diagram shows the location of sample points and the sampling mesh, and the right diagram shows the points selected via a stratified random areal sample with a sampling rate of one point per mesh box. Using the available data as the selection restrictions reduced the "randomness" of the selections, particularly when the maximum number of points to be selected approached the total number of points within each mesh area. This situation would occur if three or more points were selected per sampling unit in Figure 4b. This sampling routine was coded in Arc Macro Language (AML) for repeated use in sampling the GPS points stored in the ARC/INFO Geographic Information System (Environmental Systems Research Institute, Inc., Redlands, CA).

Table 2 shows the number of points selected with each stratified random areal sample. The maximum number of points desired indicates how many points would have

Table 2. Stratified random areal sampling schemes and number of points sought and actually selected.

Gridsize (m)	No. of points selected per mesh box	Max. desired no. of points	No. of points actually selected
100 m	1	32	29
100 m	2	64	58
100 m	4	128	116
50 m	1	120	111
50 m	2	240	221
50 m	4	480	441
50 m	16	1920	1747
25 m	1	435	391
25 m	4	1740	1535

been selected had there been sufficient GPS points available within each sampling unit. The size of the sampling units and number of selections per unit were chosen to produce similar sample sizes for the two methods and examine how the spread versus the number of data affected the interpolations.

The linear sample is similar to a vehicle-mounted GPS receiver sampling the field. Data were collected with a series of row and column passes (Figure 3). The points scattered outside of the rough grid pattern (see Figure 3) were eliminated for the linear samples. The study area is only 20.23 hectares which would be a small part of most large farms, so the vehicle turns at the ends of sampling routes were eliminated from the linear samples as well (Figure 5).

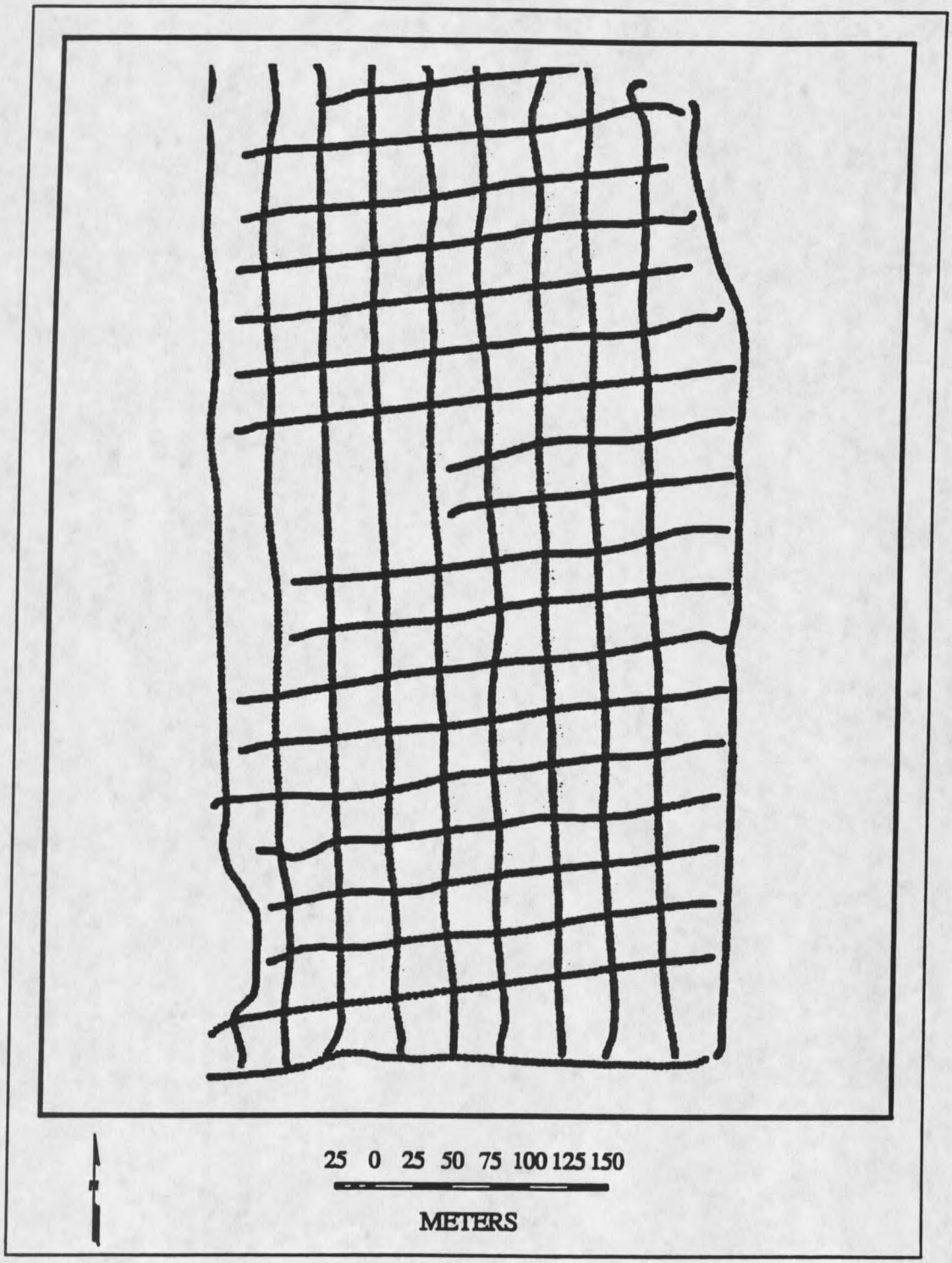


Figure 5. GPS points used in sampling methods.

Data Analysis Environment

The storage and manipulation of geographic data were accomplished with the use of a geographic information system (GIS). GIS technology allows for large amounts of spatial data to be analyzed, manipulated, displayed and stored. ARC/INFO is a modular geographic information system (GIS) with many procedures for analysis and display. Raster, vector, tabular, and TIN data structures are all supported within ARC/INFO. Programs can be written in Arc Macro Language (AML) and used within ARC/INFO to facilitate user input and decrease processing time. The initial GPS point data were stored, projected, and tagged with attributes in ARC/INFO. A program was coded in AML to automate the stratified random areal sampling. This AML was used to select the sample data. The linear samples were prepared by manually editing the GPS data set.

The samples were exported to ASCII text files consisting of x and y coordinates and elevation (z) for input into ANUDEM (Hutchinson, 1989) and TAPES-G (Moore, 1992) which are discussed in the following section. The ARC/INFO Grid module was used to evaluate and display the grids resulting from TAPES-G. The ARC/INFO Grid module supports grid combinations, grid algebra, statistical analysis and data display within one module. The TAPES-G results were brought into ARC/INFO (through an in-house preprocessing program (TAPESOUT) which generates an ARC/INFO compatible file from a TAPES-G file) and evaluated against the "true surface" in ARC/INFO. The "true surface" here is defined as the surface created using all 6,284 GPS points.

Generation of Terrain Attributes

The programs used for the generation of the terrain attributes in this study were ANUDEM and TAPES-G. ANUDEM creates a regularly spaced lattice from irregularly spaced data points. TAPES-G uses this lattice to compute primary and secondary topographic attributes.

Elevation data can be digitally represented in three ways, as linear models (contours), triangulated irregular networks (TINs) and geographic arrays (lattices). Contours are a series of linear features showing lines of equal elevation. Contours do not record the behavior of the surface between the interval lines. The manipulation and comparison of contours is often difficult due to the linear nature of the features. TINs are triangular landscape facets created from irregular points. They are often used to portray surfaces because of their ability to incorporate break lines representing ridges and streams. Problems may arise when comparing TINs computed from different input data because the differences in source inputs may lead to the creation of different triangle facets. The variation in facets may be difficult to quantify due to non-exact overlays. TINs also suffer from areal changes in the facets themselves between computed TINs. Lattices are regularly spaced rectangular arrays of elevation spot heights. Lattices are often preferred because comparisons between data sets are simple and the data storage structure is efficient and easy to manipulate. DEMs are lattices which are formed into a standard structure. Lattices are used by the TAPES-G programs for terrain attribute calculations. They are also the data structure most frequently supported by GIS systems, hence the lattice data structure was chosen for this study.

ANUDEM. ANUDEM interpolates a lattice from irregularly spaced point data and contours. The program uses an iterative finite difference interpolation algorithm which aims to achieve computational efficiency without sacrificing global trends and has the ability of conforming all data to derived or imputed drainage patterns. This method is a hybrid of local, kriging, and spline methods. ANUDEM computes successively finer resolution grids to a user-supplied tolerance. The program uses all values in grids of coarser resolution as potential weights for the resultant grids in contrast to local interpolation methods. The values for the iterative grids are determined from a derived function using a user-supplied roughness penalty and user-supplied elevation tolerances as function parameters. This roughness penalty can be thought of as the fitted taughtness of the resultant grid, similar to the smoothing parameter in splines. Elevation tolerances are used as break points for drainage enforcement. They reflect the validity of the input data and therefore help to determine how far the input data can be altered (Hutchinson, 1989, 1991; Hutchinson and Gessler, 1994; Jersey, 1993). Each successive grid resolution can be fitted to derived or user-specified drainage paths to ensure that streams do not flow to pits or along the crest of hills in the derived surface. This drainage enforcement option is a significant improvement over many other methods which produce adequate elevation surfaces but poor data for hydrologic analysis and modeling. ANUDEM has been used to create continent wide DEMs for Australia and Africa.

TAPES-G. TAPES-G is one of a suite of programs used to compute grid-based terrain attributes from square-grid inputs. The attributes computed with TAPES-G are slope,

aspect, specific catchment area, profile, plan and tangential curvatures, and flow path length for each cell (Moore et al., 1993a, 1993b). TAPES-G allows for different attribute prediction algorithms to be used for topographically important variables such as slope and specific catchment area.

Slope can be calculated in TAPES-G using two methods: the finite difference approach (FD) and an approximate approach (D8). The FD approach uses the arctan of the square-root of the sum of the squares of the first derivatives in the x and y directions to compute slope. The D8 approach calculates the gradient as the steepest slope in one of the eight cardinal directions. The results obtained from each of these slope calculations are similar, even though the FD approach generates slightly higher values, and is generally more accepted then the D8 approach (Moore et al., 1993c). The FD approach was used in this study.

Specific catchment areas can be estimated using either the D8, Rho8, or FRho8 algorithms in TAPES-G. The D8 algorithm (O'Callaghan and Mark, 1984) allows drainage from a cell to one of its eight neighbors based on the path of steepest descent. The D8 method is the most widely used of the methods, even though it tends to produce parallel flow lines along preferred directions (multiples of 45°) (Moore et al., 1993c). Rho8 and FRho8 allow flow to be routed to any of the eight neighbors on a slope-weighted basis. The Rho8 algorithm produces more realistic flow networks than the D8 algorithm by using a quasi-random direction algorithm similar to that first proposed by Fairfield and Leymarie (1991). The FRho8 algorithm (Moore et al., 1991) allows flow to be distributed to multiple cells in upland areas above the channel head. This is accomplished using a multiple direction slope-

weighted method similar to those proposed by Freeman (1991) and Quinn et al. (1991). TAPES-G reverts to either the D8 or Rho8 algorithm when the critical contributing area defining stream channels (called cross-grading area in TAPES-G itself) is reached (Moore et al., 1993c). The FRho8 algorithm is especially important in convex areas which experience radial drainage (Moore et al., 1991). The FRho8 algorithm was used in this study because most of the study site consists of upland areas and there is no permanent stream channel system.

A compound topographic index or steady-state wetness index was also calculated from two of the TAPES-G primary attributes. This index has been used to describe the effects of topography on the location and size of the saturated source areas of runoff generation in complex landscapes (Moore et al., 1993a; Jersey, 1993), the spatial distribution of soil water content (Burt and Butcher, 1985; Moore et al. 1988; Moore et al, 1993c), and the spatial variability of soil properties (Moore et al. 1993a). The index is computed as:

$$W_i = \ln (A_i / \tan \beta) \qquad (1)$$

where W_i indicates the wetness index, A_i is the specific catchment area (m^2m^{-1}), and β is the slope (degrees) on a cell-by-cell basis. The steady state wetness index relies on several explicit assumptions; notably, that the subsurface flow is parallel to the gradient of the surface topography and the system has achieved steady-state equilibrium (Moore et al, 1993c). The values given by the wetness index increase with increasing specific catchment area and decreasing slope. Hence, the values are higher in areas of water concentration (valleys) and lower in source areas and rapid drainage areas (hilltops and steep side-slopes) (Moore et al, 1993c).

Data Analysis and Display

Several statistical measures were used to analyze the differences between derived surfaces. The mean absolute difference (MAD) between the sampled surface and true surface, root mean squared error (RMSE) of the difference surfaces, and the Moran's Index of Spatial Autocorrelation for the difference surfaces were calculated. The analysis is similar to that preformed by Lee (1991) in his pioneering work comparing different elevation models. The mean absolute difference between surfaces helps define the magnitude of the error and is defined as:

$$MAD = (\sum | X_i - X_x |) / n \qquad (2)$$

where X_i is the value of the "true" terrain attribute at a cell location, X_x is the value at the same location using a DEM created with one of the sampling techniques, and n is the total number of cells. Values approaching zero indicate very similar surfaces. The RMSE is commonly used as a measure of differences between values across grids and is expressed as:

$$RMSE = (\sum (Z_i - Z_m)^2 / n)^{-2} \qquad (3)$$

where Z_i is the difference between a cell value in the sampled surface and the true surface, Z_m is the mean value of the difference surface, and n represents the total number of cells. Moran's Index (I) coefficient is a measure of autocorrelation (clustering) of data within a grid and in this study Moran's I is used to compute the clustering of errors in derived grids. Moran's I values tend to range from -1 to +1 with negative values representing negative autocorrelation (similar values further apart) and positive values indicating positive autocorrelation (similar values closer together). Values near zero represent random scattered

values with no significant autocorrelation (Barber, 1988; Ebdon, 1988; Griffith et al., 1991; Lee, 1991). Griffith et al. (1991) provide an excellent discussion of the equation used to estimate the Moran Index.

The values returned by these descriptive statistics indicate patterns within the data. Large MAD and RMSE values indicate large discrepancies between the sample grids and the grid constructed with all of the GPS data. Moran's Index values closer to positive one indicate a high clustering of errors while values nearer to zero indicate no significant spatial patterns in the errors and values nearer to -1 indicate negative autocorrelation. All these techniques assume that the closer the sampled surface is to the surface created with all of the GPS points the better the surface reflects reality (Lee, 1991; Spangrud et al. 1995).

The data and accompanying errors were visualized with a series of diagrams. Each of the sampling strategies is represented with maps of the sample locations, derived contours, and a classified difference grid showing the magnitude and pattern of errors.

Results and Discussion

Figure 6 shows the contour map that was produced using all 6,284 GPS data points. Three major topographic features are evident in this map: (1) the hilltop marking the northern boundary of the study area which is flanked by gentle side slopes to the west and south and steep side slopes to the east; (2) the shallow drainage (valley) that traverses the study area in a south-south-westerly direction in the bottom (southern) half of the study area;

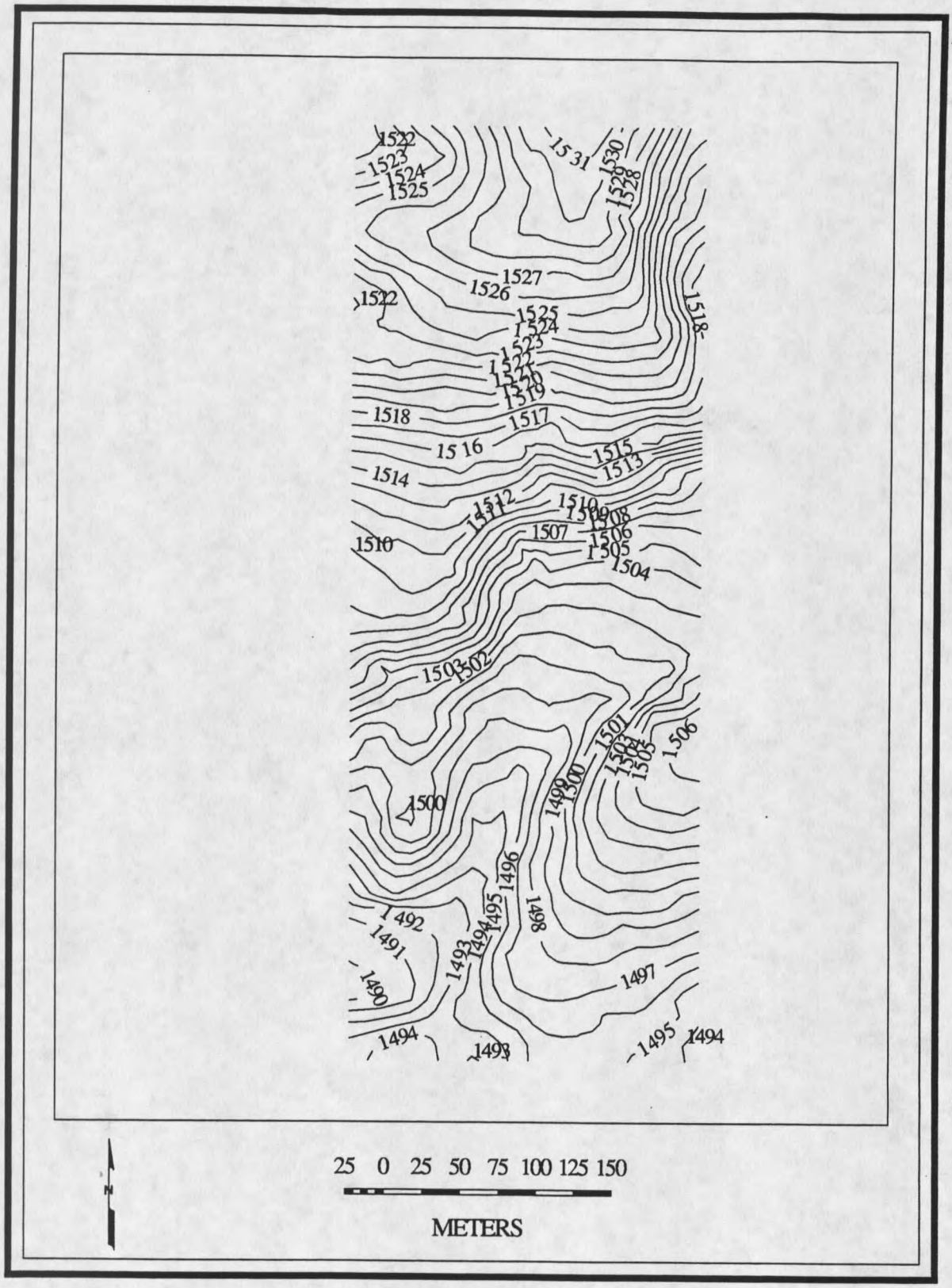


Figure 6. Contour map of study area (one meter contour interval).

and (3) the hilltop located in the southeast quadrant whose northern slopes mark the southern margin of the valley identified as (2). These topographic features were also observed in the field and they are used along with visual and statistical analysis of the differences in computed topographic attributes to assess the effectiveness of the different sample data sets.

Stratified Random Areal Samples

The stratified random areal samples were designed to mimic the samples that might have been obtained using a hand-held GPS receiver. Figures 7 through 9 show the GPS points, contour maps, and elevation difference maps. The contour maps were clipped to a portion of the GPS data to minimize any edge effects associated with surface interpolation. The final surfaces for all sampling schemes were 63 rows and 24 columns in size with a cellsize of 10 by 10 m. Tables 3 and 4 show the statistical differences for the stratified random areal sample surfaces as summarized by the mean absolute difference (MAD), the root mean square elevation difference (RMSE), and Moran's Index of spatial autocorrelation.

Figure 7 shows the effect of sampling one or more points from a 100 m grid. One point was selected for each 100 m sampling unit in the first diagram (Figure 7a). This sampling scheme did not delineate the western side-slopes for the northern hill (feature 1) or the side-slopes that mark the southern boundary of the stream channel (feature 2). This scheme produced a large spread of data and a surprisingly good mean elevation result (MAD equals 1.39); however, the RMSE is large and Moran's Index approaches +1 (indicating positive autocorrelation (or clustering) of the elevation surface errors). The elevation surface exceeded 1 m of absolute difference in 41 percent of the total area.

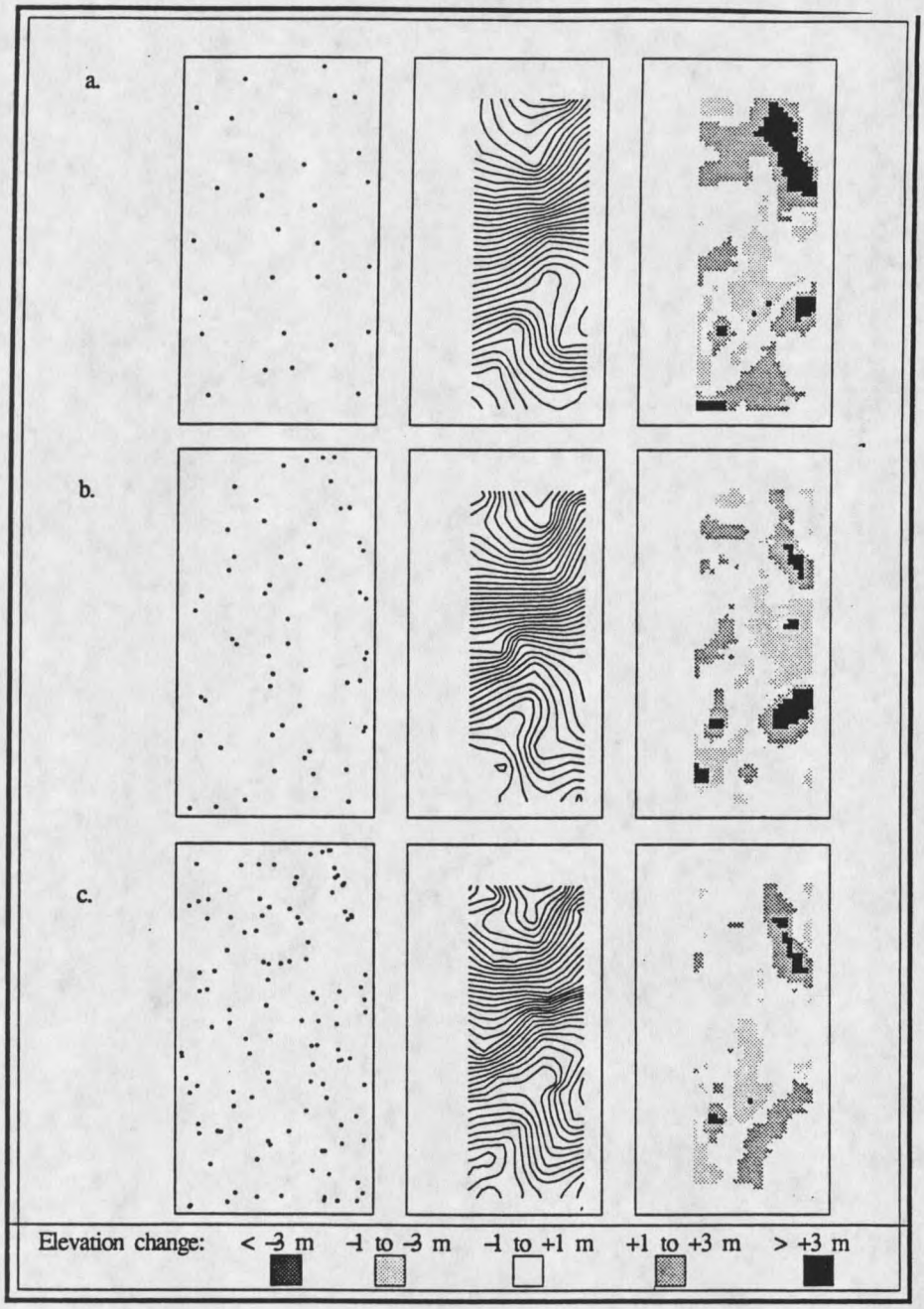


Figure 7. Random stratified samples of (a) 100 meter sampling unit with one point, (b) 100 meter sampling unit with two points, and (c) 100 meter sampling unit with four points. The left column of grids shows the sample points, the middle column shows the contour lines (one meter contour interval) from the derived elevation surface, and the right-hand column shows the elevation difference maps.

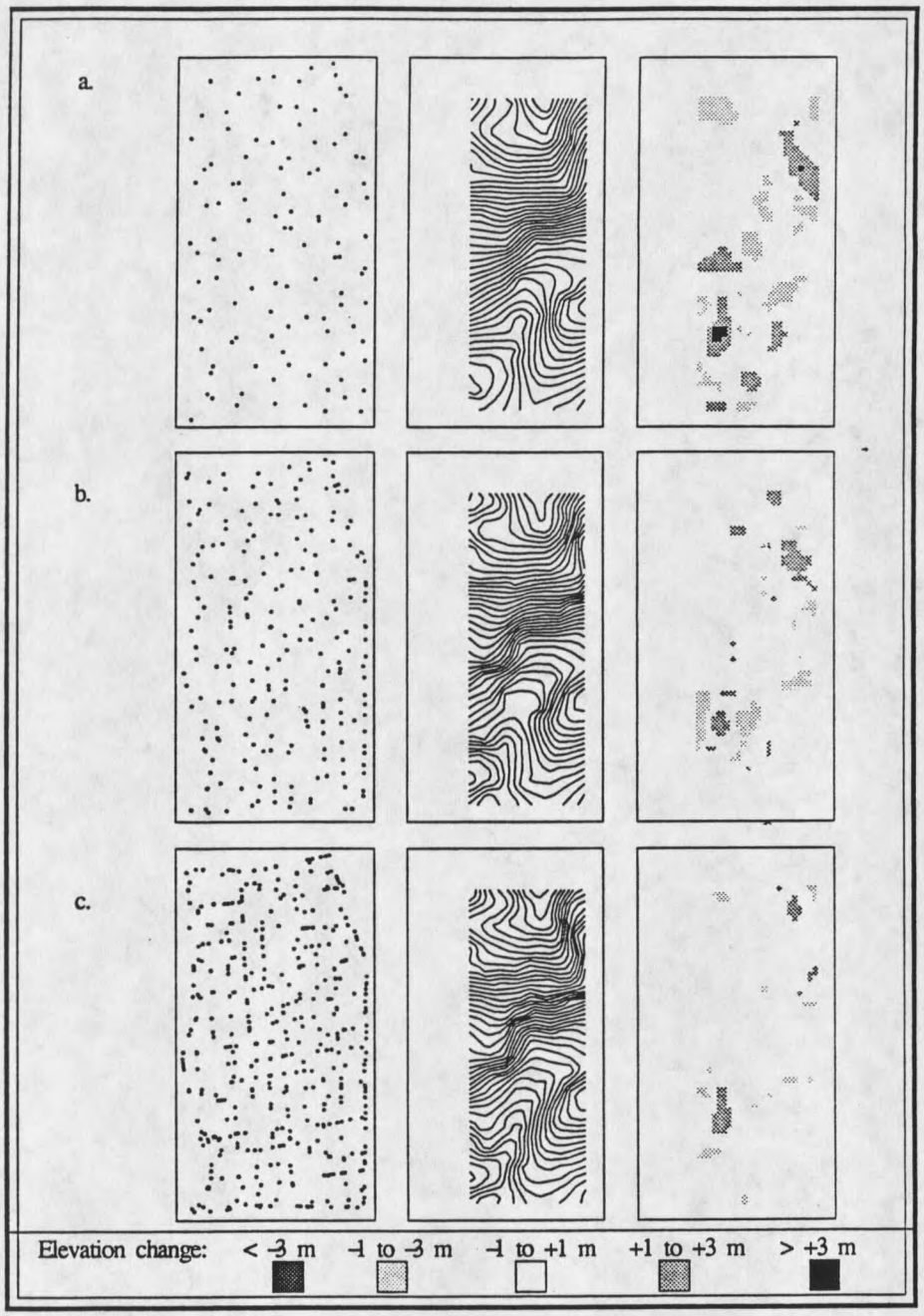


Figure 8. Random stratified samples of (a) 50 meter sampling unit with one point, (b) 50 meter sampling unit with two points, and (c) 50 meter sampling unit with four points. The left column of grids shows the sample points, the middle column shows the contour lines (one meter contour interval) from the derived elevation surface, and the right-hand column shows the elevation difference maps.

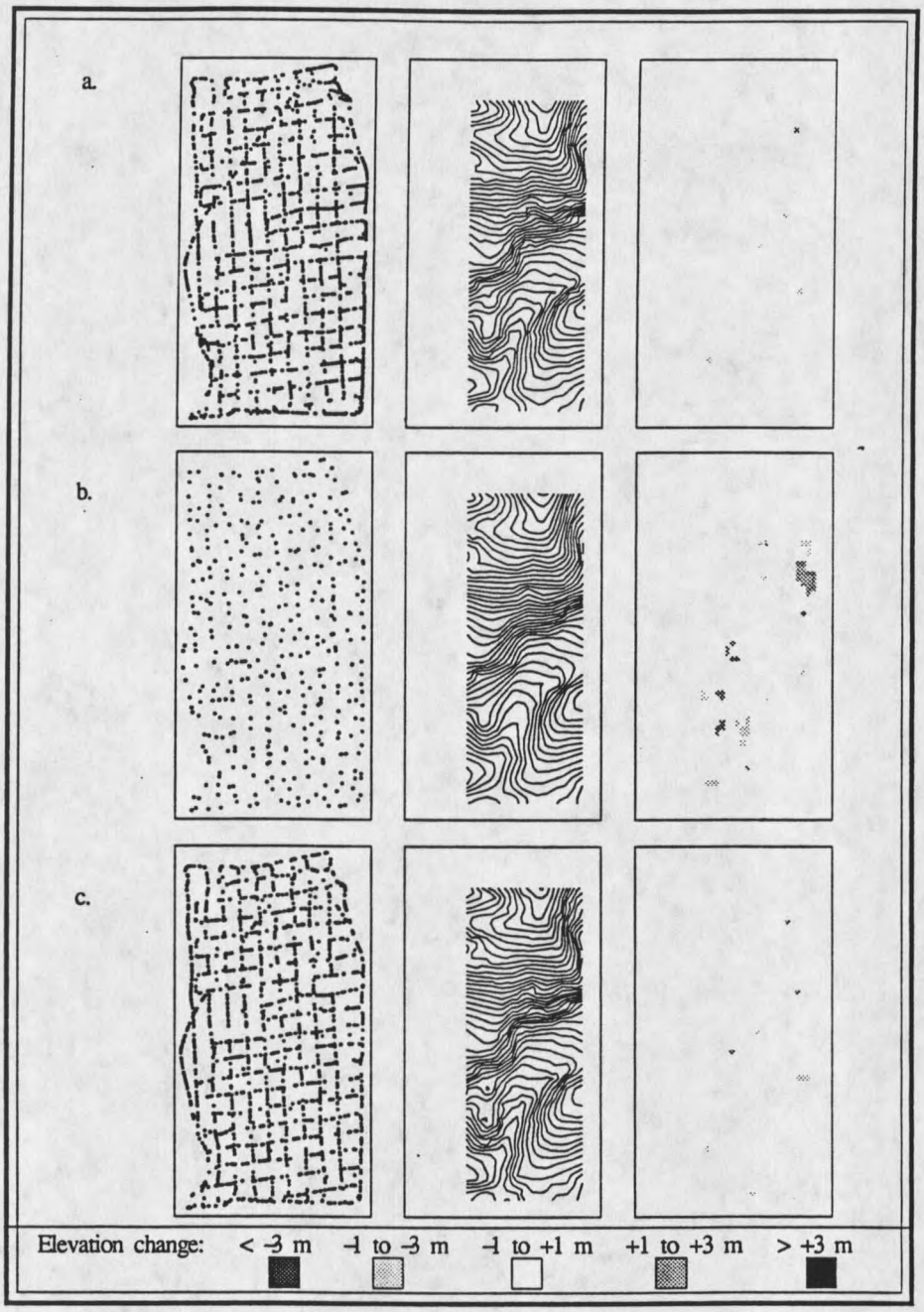


Figure 9. Random stratified samples of (a) 50 meter sampling unit with sixteen points, (b) 25 meter sampling unit with one points, and (c) 25 meter sampling unit with four points. The left column of grids shows the sample points, the middle column shows the contour lines (one meter contour interval) from the derived elevation surface, and the right-hand column shows the elevation difference maps.

Table 3. Spatial analysis of elevation differences from stratified random area sample points.

Grid size and # of points / grid cell	No. points selected	Mean absolute elev. diff. (m)	RMSE (m)	Moran's Index	Number of cells exceeding the 1 m error level
100/1	29	1.39	1.71	0.89	624
100/2	58	1.04	1.43	0.86	369
100/4	116	0.78	1.16	0.81	296
50/1	111	0.60	0.95	0.74	180
50/2	221	0.43	0.73	0.58	89
50/4	441	0.31	0.60	0.46	53
50/16	1747	0.16	0.40	0.19	2
25/1	391	0.27	0.54	0.39	36
25/4	1535	0.16	0.40	0.19	4

Table 4. Spatial analysis of elevation differences from linear routes simulating data from a truck-mounted receiver.

Linear routes	No. points selected	MAD (m)	RMSE (m)	Moran's Index	Number of cells with differences exceeding ±1 m
0-1-0 NS	1363	0.35	1.20	0.81	178
All NS	2682	0.21	0.51	0.42	45
0-1-0 WE	1120	0.62	0.96	0.72	140
All WE	2135	0.22	0.48	0.34	7
0-1-0 NSWE	2483	0.35	0.66	0.60	66
All NSWE	4817	0.05	0.23	0.11	1

The next sample (Figure 7b) used two points per 100 meter sampling unit. While this scheme is better then the previous sample it has two major problems, a pit has now formed in the lower part of the stream channel (feature 2) and the southern hill (feature 3) exhibits a concave curvature along the north-west margins. These results contributed to the poor statistical performance as evidenced by the MAD and RMSE values. The Moran Index was again very high (.86) indicating that the errors in the surface are still highly clustered (Table 3).

The third 100 m sample used 4 sample points per sampling unit (Figure 7c). This sample included 116 sample points and the resulting contour map seems to mimic the contour map produced with all the GPS data fairly well. The stream channel (feature 2) and southern hill (feature 3) are clearly delineated, but the steep northern hill (feature 1) lacks some definition along the eastern side-slopes and the associated ridge line. The visual comparisons are a little deceiving because the statistical differences are still great. While the MAD and RMSE both are close to one, indicating a much better fit than previously attained, the Moran Index is still high indicting that the errors are still highly clustered (Table 3). The number of cells with elevation differences greater than 1 m is 50% less than the number generated when 1 point was chosen per 100 m sampling unit.

The next set of maps (Figures 8 and 9a) show the results of using a 50 m sampling unit. Figure 8a shows the sample locations, contours, and difference map obtained with 1 point per 50 m cell. This sample size (111 points) is very similar to that generated with 4 points per 100 m grid (116 points). The contoured result is also very similar to that obtained with 4 points per 100 m cell but the elevation differences and statistical results are superior with

the 50 meter sampling scheme because of the increased distribution of points (spread) across the area. The elevation differences are reduced in the 50 m sample, especially in the northern hill area (feature 1), the mean absolute elevation difference is only .60 m, and the RMSE and Moran index values are slightly lower than those computed for the equivalent 100 m sample. These improvements continued when 2 and 4 points were selected per 50 m sampling unit (Figures 8b and 8c). Selecting four points per sampling unit (Figure 8c) delineated the steep east slopes of the northern hill (feature 1). The elevation difference map shows how the number of cells with over 1 m of elevation change (53) is lower than in the previous samples. Moran's Index has now dropped below .50 indicating that the errors are more scattered and more randomly distributed across the landscape (Table 3). With sixteen points per 50 m sampling unit (Figure 9a) the map appears nearly identical to that produced with all of the GPS data points. The MAD value is now within the range of possible data collection error and the errors exhibit little autocorrelation, as indicated by a Moran's index value of 0.19 (Table 3). The error difference map now shows only 2 cells with greater than 1 m absolute elevation differences.

The final stratified random areal sample used a sampling unit of 25 m. This sampling scheme reflects the original data pattern because the number of points collected is always large. The two samples (Figures 9b and 9c) are very similar to some of the samples collected with the 50 m sampling unit because the sample sizes based on selecting one and four points per 25 m sample unit matched those collected at four and sixteen points per 50 m unit, respectively. Both show the major topographic features and the higher sampling rate (4 points) mimics the complete data set very well. The statistical comparisons obtained with

the 1 point per 25 m sample are slightly better than those obtained with four points per 50 m unit even though the sample sizes were similar. Selecting 4 points per 25 m sampling unit selected 212 less points than the comparable 50 meter sample, but produced almost identical results (Table 3). The pattern of elevation differences produced with these samples is almost identical (Figures 9b and 9c).

Linear Samples

The samples collected as linear patterns (collection routes) tell a very interesting but different story relative to the stratified random areal samples in that increasing the total number of sample points does not guarantee a better surface. Figures 10 and 11 show the sample points and contours derived from these linear samples. The data points appear as lines due to the density of sampling. Figures 10a and 11a show samples based on north-south routes. Notice the alignment of the errors in a north-south direction in the error difference maps. The east-west sampling routes are shown in Figures 10b and 11b. The smaller sample produced errors aligned in a west-east direction (Figure 10b). The last two samples (Figures 10c and 11c) reflect combinations of west-east and north-south routes. These samples captured data across the entire field without a directional bias and provided samples with better spread. These samples produced surfaces very similar to those created with the entire data set (compare contour maps and error maps in Figures 10c and 11c to the contour map in Figure 6).

The linear sampling results are much more a function of the orientation of landscape features relative to the sampling direction than those obtained with the stratified random

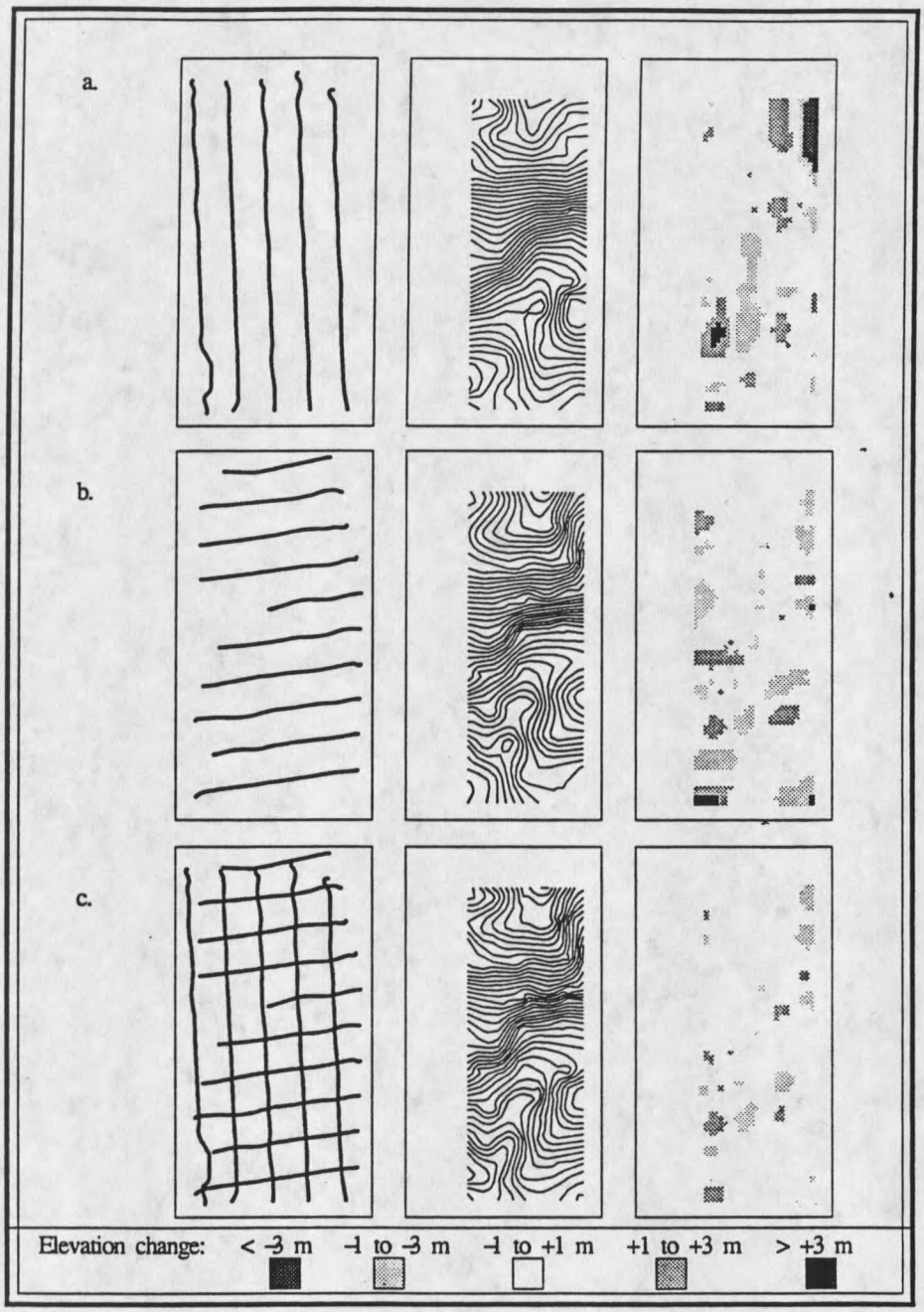


Figure 10. Linear samples of (a) every other north-south route, (b) every other west-east route, (c) combination of every other north-south route and every other west-east route. The left column of grids shows the sample points, the middle column shows the contour lines (one meter contour interval) from the derived elevation surface, and the right-hand column shows the elevation difference maps.

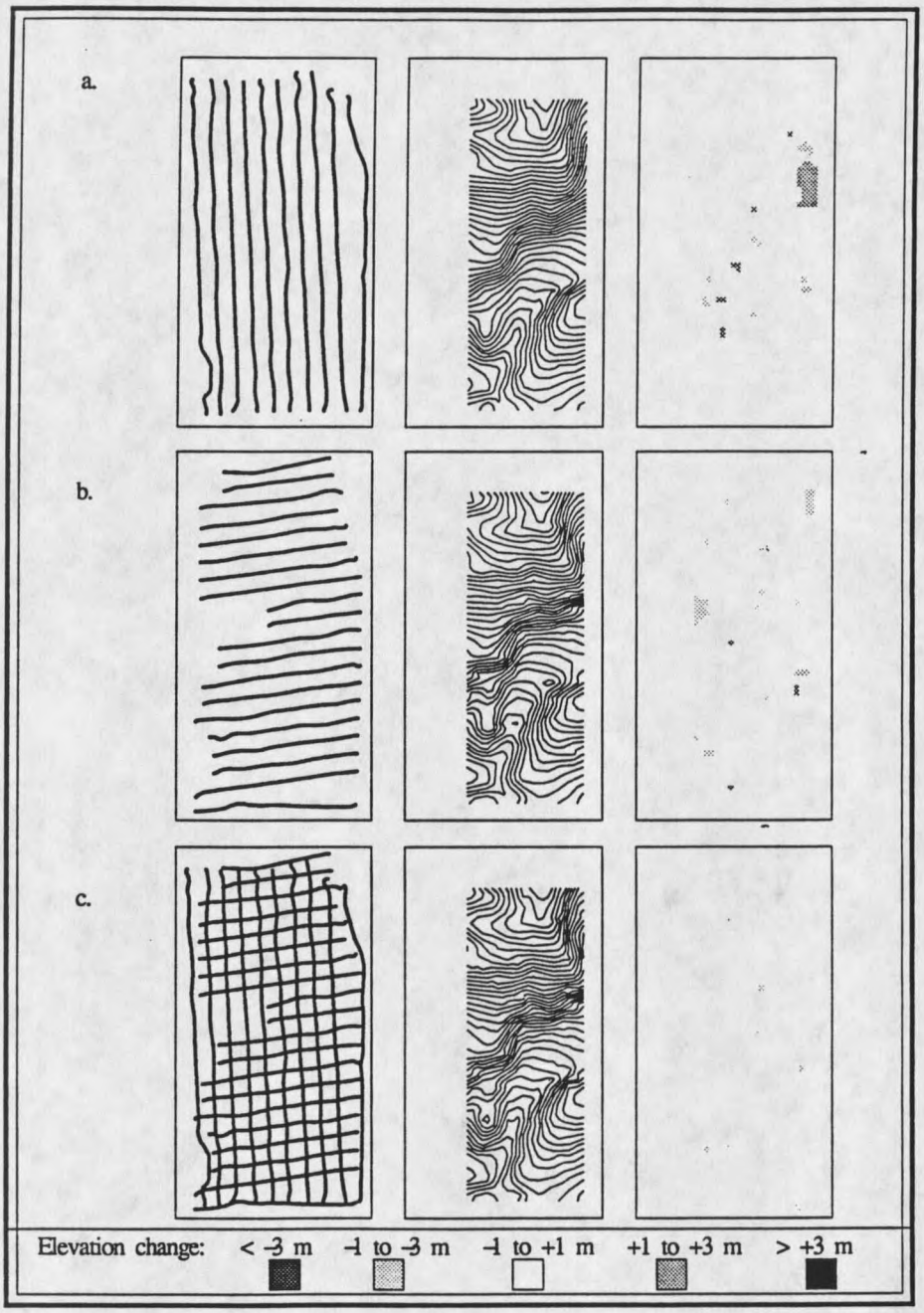


Figure 11. Linear samples of (a) every north-south route, (b) every west-east route, (c) combination of every north-south route and every west-east route. The left column of grids shows the sample points, the middle column shows the contour lines (one meter contour interval) from the derived elevation surface, and the right-hand column shows the elevation difference maps.

samples. The west-east sample routes produced pits in the lower half of the drainage channel (feature 2). A pit occurred in only one of the stratified random samples. The northeast-southwest path taken by the stream channel across this field meant that some of the linear samples did not pick up enough of this downslope gradient to resolve (eliminate) the pits. Notice also how the north-south samples exhibit linear patterns of error, most distinct in the contour and error maps generated with the alternating north-south sampling routes (Figure 10a).

Table 4 shows the statistical differences between the linear samples and the complete data set. The general trend of increased sample points leading to better results is repeated; hence, the sample routes which chose every other path (Figures 10a and 10b) produced consistently higher errors, as evidenced by larger values in the absolute difference means, RMSE and Moran Index values. The sampling bias implicit in the linear paths resulted in much larger and more clustered errors for these samples compared to the stratified random areal samples (see Tables 3 and 4).

Topographic Attributes

The differences in elevation exhibited by the different sampling schemes tell only part of the story. Minor changes in elevation can dramatically alter other topographic attributes. Slope, specific catchment area, and a steady state wetness index are components of the landscape which may help to define management areas for precision farming. Table 5 shows the means of the computed attributes for each of the sampling schemes. Notice how minor

Table 5. Arithmetic means for selected topographic attributes.

Sample	Elevation (m)	Slope gradient (degrees)	Specific catchment. area (m^2 m^{-1})	Wetness Index
100/1	1508.4	7.1	306.9	7.53
100/2	1508.9	7.6	263.5	7.33
100/4	1508.9	8.3	297.8	7.12
50/1	1509.0	8.6	238.8	7.09
50/2	1509.0	9.4	234.8	6.93
50/4	1509.0	9.9	243.5	6.78
50/16	1509.1	11.1	242.6	6.63
25.1	1509.1	9.9	238.8	6.79
25/4	1509.1	11.1	243.7	6.62
0-1-0 NS	1509.1	8.3	353.3	7.20
All NS	1509.0	10.0	240.2	6.77
0-1-0 WE	1509.2	9.1	281.2	6.95
All WE	1509.1	10.5	271.3	6.70
0-1-0 NSWE	1509.1	9.8	282.6	6.81
All NSWE	1509.1	10.8	255.9	6.67
All data	1509.1	9.5	222.5	8.45

Table 6. Spatial analysis of topographic attribute differences from stratified random samples.

Sample	Slope gradient			Specific catchment area			Wetness index		
	MAD	RMSE	Moran	MAD	RMSE	Moran	MAD	RMSE	Moran
100/1	3.48	4.14	0.78	333.6	1381.5	0.05	1.20	1.41	0.62
100/2	3.18	3.85	0.76	294.3	1262.2	0.05	1.06	1.37	0.60
100/4	2.68	3.54	0.72	299.5	1399.8	0.07	0.90	1.27	0.54
50/1	2.37	3.06	0.65	227.2	1041.9	0.01	0.76	1.10	0.48
50/2	2.03	2.72	0.57	234.6	1109.7	-0.09	0.65	1.05	0.34
50/4	1.97	2.65	0.53	131.2	738.6	-0.11	0.50	0.82	0.31
50/16	2.02	2.46	0.36	103.7	747.1	0.01	0.35	0.61	0.08
25/1	1.82	2.47	0.49	172.3	909.3	-0.12	0.50	0.84	0.21
25/4	2.06	2.42	0.35	98.6	694.9	0.07	0.37	0.62	0.16

elevation changes are translated into large changes in the other terrain attributes. While elevation change was less than 0.05 percent, the other terrain attributes changed by 10.8, 27.5 and 36.0 percent (wetness, specific catchment area and slope gradient, respectively) relative to the surfaces produced with all 6,284 GPS points. Slope is a function of the surrounding surface, and errors are magnified when the elevation in one adjacent cell is increased and another surrounding cell's elevation is decreased. The specific catchment area also reflects these minor elevation changes, because as water drains across a surface a small change in the downslope elevation may result in a change in the direction of maximum descent which forces water to either flow around areas of increased elevation or toward areas of decreased elevation. The steady state wetness index combines these terms (slope and specific catchment area) and any errors associated with them.

Table 6 summarizes the topographic attribute differences for the stratified random sampling schemes. The slope gradient and steady state wetness index exhibited a trend similar to elevation given that the magnitude and clustering of errors decreased as the total number of sample points increased and the sampling unit decreased in size. The Moran's Index and RMSE values clearly show this trend. The specific catchment area and steady state wetness index statistics summarized in Table 6 show different patterns. The magnitude of the differences in specific catchment area decreased with increased sample size and there was no error clustering at any sample size.

The effects of the spread and pattern of the linear samples on topographic attributes are summarized in Table 7. The north-south samples repeat earlier trends in that the more

widely dispersed north-south sample (0-1-0 NS) produced higher errors across all attributes when compared to the denser north-south sample (All NS). Both north-south samples produced larger errors than those generated by the west-east samples. The Moran's index computed for slope is higher in the alternate north-south sample than in the comparative west-east sample, indicating that the slope errors were more clustered in the north-south sample than in the west-east sample. The relationships between the north-south and the west-east sample statistics are similar. The sample using both north-south and west-east samples improved results due to the better spread of data across the field.

Table 7. Spatial analysis of topographic attribute differences from linear samples.

Sample	Slope gradient			Specific catchment area			Wetness index		
	MAD	RMSE	Moran	MAD	RMSE	Moran	MAD	RMSE	Moran
0-1-0 NS	2.69	3.79	0.76	347.5	1443.8	0.14	0.92	1.34	0.57
All NS	1.67	2.31	0.47	163.4	941.0	-0.03	0.45	0.80	0.21
0-1-0 WE	2.48	3.28	0.68	260	1248	0.08	0.67	1.09	0.39
All WE	1.79	2.27	0.45	143.8	891.2	0.10	0.40	0.72	0.21
010-NSWE	2.01	2.78	0.57	243	1228	0.15	0.55	1.01	0.33
All NSWE	1.62	1.97	0.32	102	789.3	0.1	0.29	0.60	0.13

The specific catchment area and steady state wetness index reflect the same trends shown by the slope attribute. While the alternating north-south and west-east samples produced a similar number of samples, the orientation of these samples accounts for the differences in the mean absolute difference (MAD). The alternating north-south sample did not delineate the channel system because of its orientation along the major drainage direction. The equivalent west-east route had a smaller MAD and more visible channel

system (similar to the real field). The increase in the number of sampling points with the complete NS, EW, and NSEW samples increases the accuracy of the resultant specific catchment area and compound topographic index surfaces similar to the results from the slope attribute. The steady state wetness index reflects the errors of each attribute because it is a function of slope and catchment area.

Comparing the results from the stratified random samples to those generated with the linear sampling routes, the stratified random samples achieve results as good as those of the every other route linear samples with slightly less than 33% of the input data points. Notice also that the Moran's Index is consistently higher for the linear samples (with the exception of the NSEW sample, which incorporates almost the entire data set) than the random stratified samples (Tables 3 and 4) signifying that the linear sample errors were more clustered as well.

Summary

GPS technology is increasingly advocated as a cost-effective and accurate method for collecting input positional and elevation data for use in the creation of digital elevation models. These models can be used to compute primary and secondary terrain attributes in support of precision farming. Surfaces were created using 17 different sampling strategies and these surfaces were compared to a surface created using all 6,284 GPS data points. Comparisons were made using three primary terrain attributes (elevation, slope and specific catchment area) and a secondary terrain attribute (steady state wetness index). Differences

between the surfaces were shown graphically and with statistical summaries. Mean absolute attribute differences, root mean squared errors of the differences between grids and Moran's Index of spatial autocorrelation were used to quantify the extent and pattern of the error.

These results show that: (1) the magnitude and clustering of the spatial distribution of errors in the stratified random areal samples diminish rapidly as sample size increases; (2) the magnitude and clustering of the spatial distribution of errors in the linear samples are much greater and do not diminish as quickly with the addition of more points (indicating the linear samples do not perform as well as the stratified random areal samples holding sample size constant); (3) for a given total number of sample points, the magnitude and clustering of the spatial distribution of errors diminish as the sampling unit size decreases indicating that the spread of data (controlled by the sampling unit size) is very important; and (4) seemingly small variations in elevation may result in large differences in computed primary and secondary terrain attributes.

The last observation has important implications for those who want to use these spatially-variable attributes to predict other environmental variables (e.g., Dikau, 1989; Moore et al 1993a, 1993b) and those who want to use these attributes as inputs to environmental models (e.g., Panuska et al., 1991). Researchers, farmers and consultants wanting to use GPS technology to build DEMs and perform terrain analysis should be aware of these differences and should plan their data collection efforts accordingly.

CHAPTER 3

INTERPOLATION OF SOIL ATTRIBUTE SURFACES USING SOIL-LANDFORM RELATIONSHIPS

Introduction

The over-application of nitrates on farm fields has been shown to cause increased contamination of ground water in many agricultural areas. Precision farming may reduce this nitrate leaching. Precision farming divides fields into management units based on attributes such as yield or leaching potential and strives to manage each unit separately. The creation of these management units requires information on key soil properties (Carr et al., 1991). Direct measurement is not economically feasible for many attributes whereas the generation of terrain attributes from an elevation surface is a fairly straightforward procedure (as demonstrated in the previous chapter). The linkage between these terrain attributes and selected soil properties is the current focus of a large body of research (e.g., Moore et al., 1993a, 1993b). This chapter uses terrain attributes, remotely sensed imagery, and soil survey data in conjunction with soil attribute point data (pH, percent organic matter, and depth of the mollic epipedon) in an attempt to interpolate soil attribute surfaces across a field.

Literature Review

Soil classifications are typically prepared by human experts (soil surveyors) and are defined in terms of "experience" or "judgement". The general soil patterns between surveyors will be similar, although the exact classification of locations of soil map unit boundaries and other features will change due to differences in "experience" and "judgement" (Hewitt, 1993; Burrough et al., 1994). These soil surveys usually specify ranges in soil attribute values and these ranges may be too large to distinguish spatial attribute variation in soils or crops within map units (Hewitt, 1993; Burrough et al., 1994). These large ranges in attribute values occur because only one set of soil mapping units is defined. Separate soil maps are not produced for each soil attribute. Not surprisingly the development of predictive models in soil science is directed toward quantifying human expert systems and soil map units into reproducible classifications of soil attributes.

Soil attribute prediction is hindered by the variable nature of soils across a landscape, which has long been recognized but not quantified (Arnold and Wilding, 1991). Variability of soils associated with geological, pedological and human-induced processes is well documented (e.g., Wilding et al., 1983; Buol et al., 1988; Bouma and Finke, 1993). Bouma and Finke (1993) argued that soil variability over time can be expressed in three ways: (1) static variability, (2) dynamic variability and (3) management-induced variability. Static variability refers to the soil properties which express little spatial change across a mapped soil unit, such as organic matter and soil texture. Dynamic variability refers to the variability across short distances within mapped soil units, which results from varied "..water, solute, air, and temperature regimes... (which) are considered to be of practical significance"

(Bouma and Finke 1993). Dynamic variability affects soil nutrient levels and other agriculturally important variables. Human or management-induced variability results from human interactions with the soil landscape, ranging from tillage and leveling of land to the unequal application of fertilizers (equal-rate broadcast fertilizer spreading often results in a 10 to 15 percent variability in application rates across a field) (Finke et al., 1992; Bouma and Finke, 1993). This human-induced variability affects most soil attributes, such as pH and percent organic matter, and can alter the inherent static and dynamic variability. The human aspect is perhaps the most difficult to quantify.

Burrough et al. (1994) used a series of diagrams to show numerous hypothetical variations in soil attributes across fields (Figure 12). Figure 12a shows the ideal static variation of soil units where all soil changes are assumed to occur at the mapped soil boundaries. Figure 12b shows a continuous variation model with no expressed uncertainty. Homogeneous geologically dominant features can result in variations within soil units that are smaller than those occurring across map unit boundaries as shown in Figure 12c. Continuous surfaces can also be represented with small variations following a general continuum, such as ground water levels with local noise (Figure 12d). Most adequately mapped soil attributes are thought to follow the non-stationary variation shown in Figure 12e, where part (1) shows the general trend variation, part (2) shows abrupt variations, and part (3) shows the larger than average within-unit variation. Figure 12f shows the results when the signal from the soil surface is swamped by short-range noise variations that may be attributed to local human actions (Burrough et al., 1994).

The diagram which applies to any particular situation will depend on the particular soil attribute of interest. While one soil attribute may have excessive within-unit variation (e.g.,

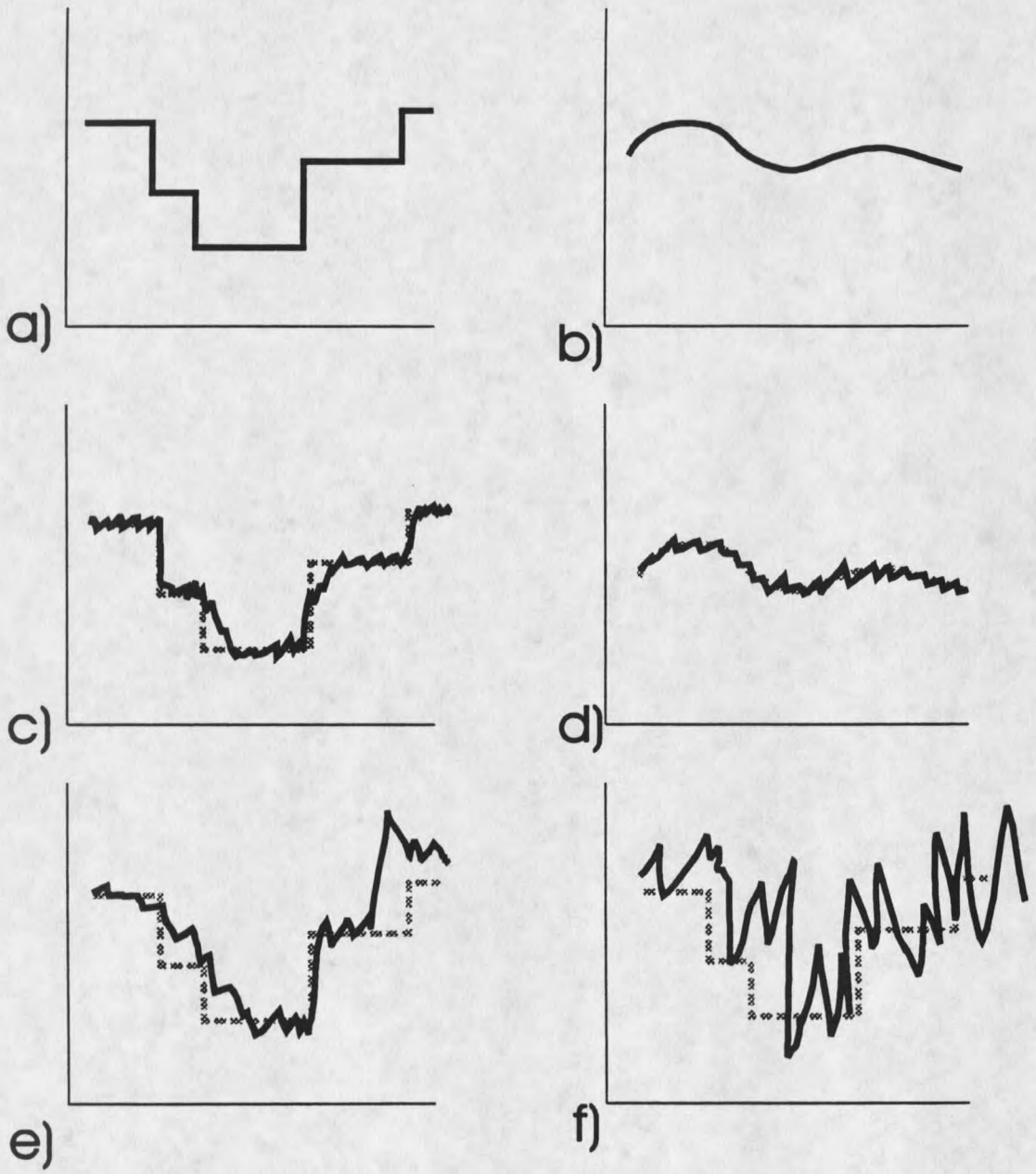


Figure 12. Hypothetical soil variation (from Burrough et al. 1994, p. 315). Variability (y) verses distance (x).

part 2 of Figure 12e) another may exhibit little variation within mapped units (Figure 12d). Such within unit variation is clearly shown graphically and may be captured statistically using numerous statistical tests including ANOVA and the Krustal-Wallis H statistic, a non-parametric alternative to the classical F ratio for analysis of variance. The sample size, normality, and the skewness of data all affect which test is used.

There are two types of approaches that have been proposed for modeling soil variability: target-property models and soil-landscape models. Target property models (also referred to as pedotransfer functions) predict a particular target soil property from other measured soil properties. The calculation of base saturation using soil pH and an appropriate base saturation curve exemplify this modeling approach (Hewitt, 1993). Such target property models are usually applied to mapped soils or specific soil pedons. Soil-landscape models relate soil attributes to landscape attributes that can be consistently recognized. Such landscape models are used at various scales, ranging from broad landscape views using large topographic features to detailed site-specific views using higher resolution terrain attributes. These may include qualitative and quantitative descriptions of physical landforms, terrain and image attributes, erosional action, geology, and vegetation (Hewitt, 1993). Soil scientists use landscape features which affect or demonstrate soil variability in the field to prepare soil surveys. But the quantification of this "landscape" is difficult given that the soil scientist operates as an expert-system and internally incorporates more that just physically observable features in any soil-landscape judgment (Arnold and Wilding, 1991).

The elucidation of soil variations based on landform properties has been demonstrated by many researchers (e.g., Aandahl, 1948; Troeh, 1964; Webster, 1977; Anderson and Burt, 1978; O'Loughlin, 1981; Burt and Butcher, 1985; Odeh et al., 1991; Moore et al., 1993a,

1993b). Troeh (1964) correlated slope gradient, profile curvature, and plan curvature with selected soil properties, based on qualitative mapping units rather than quantifiable terrain attributes. Topographic attributes have been linked quantitatively to A-horizon depth, clay content and depth, soil mottling, pH, water capacity, organic carbon, and carbonates by many individuals during the past two decades (e.g., Vreeken, 1975; Kreznor et al., 1989; Moore et al., 1993a, 1993b). These studies fall under the relatively new soil-science area of pedometrics given their use of statistical and probability-based systems for soil classification and soil surface generation.

Moore et al. (1993a, 1993b) explored the classical statistical relationships of soil and landscape attributes in their work in Northern Colorado. Using a well documented site (231 soil observations on a 15.25 m grid), they attempted to use terrain attributes calculated from a DEM of the area (15.25 m cellsize) to explain soil-landscape relationships. A-horizon thickness, extractable phosphorus, organic matter, pH, and sand and silt percentages (in the top 0.1 m of soil) were correlated to terrain attributes using stepwise linear regression. Slope and wetness index were significant at the 99% level for all attributes, except organic matter where the stream power index was the significant variable. The terrain variables explained a substantial portion of the spatial variability in these soil properties (r^2 values ranging from .40 for pH to .63 for silt).

Gessler et al. (1993) investigated soil-landscape models in a similar manner to Moore et al. (1993a, 1993b) for use in the development of soil surfaces. This study used generalized linear models to quantify soil-landscape relationships and found wetness index (referred to as the Compound Topographic Index (CTI) by Gessler et al.) and plan curvature were significant (0.05 significance level) in predicting A horizon depth and solum depth (r^2

of .63 and .68, respectively). Soil attribute surfaces were constructed using the coefficients obtained from the regression analysis. Gessler at al. (1993) argued that this approach offers three advantages over traditional soil surveys: (1) evidence is not confused with interpretation; (2) all data are consistent and repeatable; and therefore (3) models can be communicated in an objective way.

An alternative approach utilizes areal photography and satellite imagery to predict soil properties. Some soil properties have different spectral reflectance properties. Frazier and Cheng (1989) used Landsat TM imagery to predict the organic matter levels in the Palouse region of Eastern Washington. Using image band combinations to produce ratioed digital numbers (DN) they were able to correlate the image's DN value with percent organic matter.

Modeling the soil-landform relationships along with an increased understanding of the within-unit soil variability may not be just an academic exercise. If the soil attributes and the variation of these attributes within a field can be adequately mapped/derived then we may acquire sufficient information to variably apply fertilizer and/or pesticide where it is needed and thereby eliminate excess applications (Burrough, 1993).

Materials and Methods

Soil Sampling

Soil samples were collected and described at 70 sites in May 1992 across the same field discussed in Chapter 2 (Figure 13). At each site a 5.08 cm diameter PVC neutron

access tube was installed to 200 cm below the soil surface using a truck mounted Giddings hydraulic probe. Numerous soil characteristics were recorded or samples were collected for laboratory analysis at each site in May 1992. Field descriptions of the mollic, Bt, Bk, and 2C horizons including their color, texture class, and percent clay were obtained from the soil cores removed when the neutron access tubes were installed. The cores were divided into 15 cm and 30 cm increments down to depths of 30 cm and 210 cm, respectively for laboratory analysis. Soil samples were air-dried, ground, and screened through a 2 mm sieve prior to analysis. Soil pH, EC (1:2 mmhos/cm), OM (%), Olsen P (mg/kg), and K (mg/kg) were determined in the top 30 cm and NO^3-N (mg/kg) was determined for each increment to 210 cm. Only the thickness of the mollic epipedon, OM, and pH data were used for this part of the study.

Generation of Terrain Attributes

The GPS data were converted to a regular 10 m grid with ANUDEM (Hutchinson, 1989) for subsequent analysis and display. This program takes irregular point or contour data and creates grid-based DEMs. The data source and processing steps were discussed in Chapter 2.

Primary terrain attributes were computed directly from the interpolated 10 m grid using TAPES-G, a grid-based program for terrain analysis that calculates slope, aspect, specific catchment area, profile, plan, and tangential curvature, and flow path length for each cell in a square grid DEM (Moore, 1992). The maximum slope gradient, β (in degrees), was computed with a finite difference algorithm from the directional derivatives by:

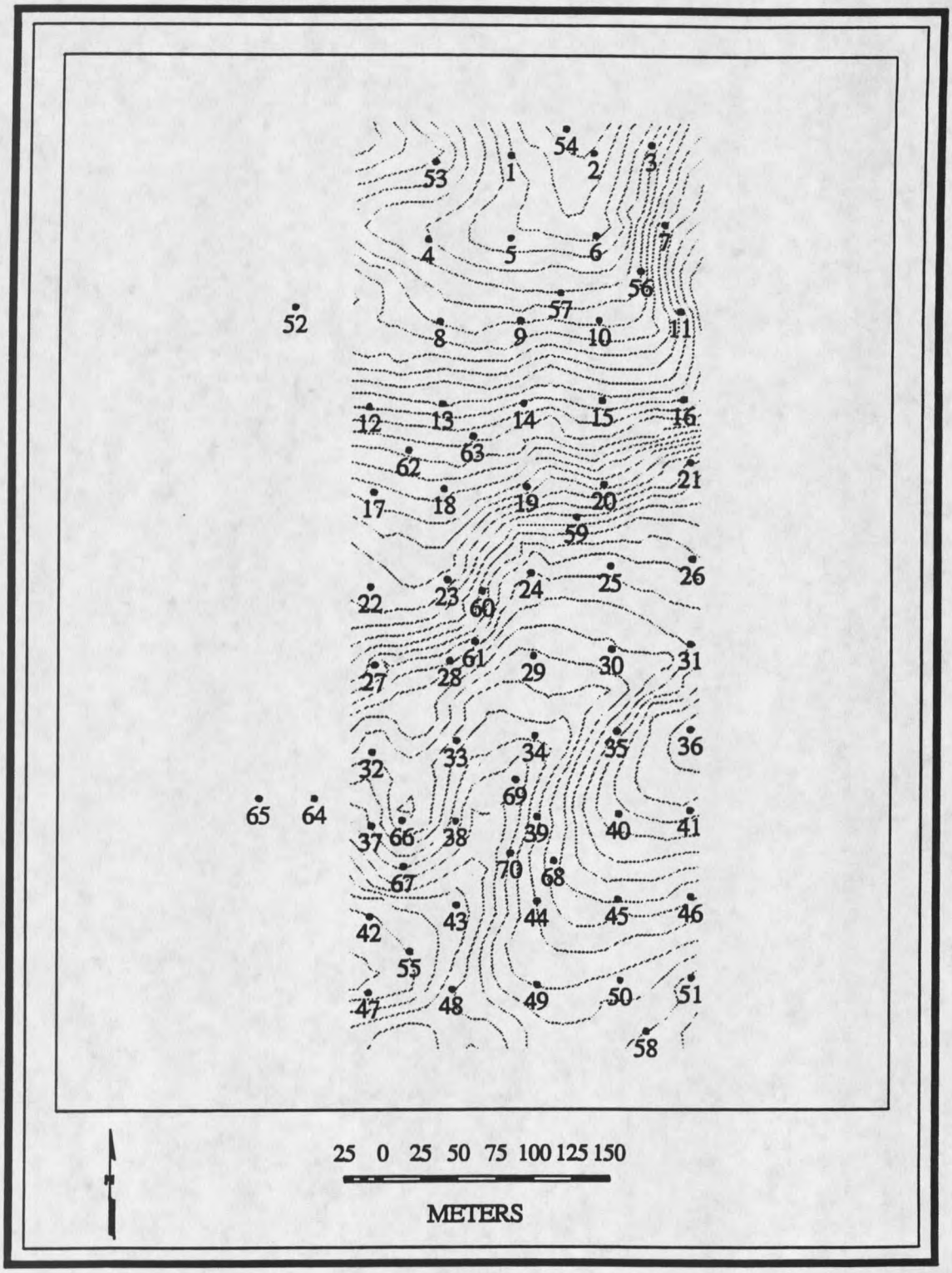


Figure 13. Map of 70 soil sample locations

$$\beta = \arctan [(f_x^2 + f_y^2)^{0.5}] \quad (4)$$

where β represents the slope gradient in degrees and f_x and f_y are the directional derivatives. Specific catchment area (A_s) and flow path length were calculated with the FRho8 algorithm in upland areas above defined channels and the Rho8 algorithm (which allows drainage from a node to only one of eight nearest neighbors on a slope weighted basis) in channel areas. The Rho8 algorithm produces more realistic flow networks than the D8 algorithm and the FRho8 algorithm permits modeling of flow dispersion in upland areas, which is important in areas with complex topography (Moore et al., 1991; Moore, 1992; Moore et al., 1993b). A minimum drainage area of 2 ha (approximately 10% of the study area) was used to initiate channel flow. The proportion of flow or upslope contributing area assigned to multiple downslope nearest neighbors above these channels was determined on a slope-weighted basis using methods similar to those proposed by Freeman (1991) and Quinn et al. (1991), so that the fraction of catchment area passed to a neighbor i is given by:

$$F_i = \text{Max}(0, \text{Slope}_i^{1.1}) / \Sigma [\text{Max}(0, \text{Slope}_j^{1.1})] \quad (5)$$

where F_i is the catchment area, Slope is the slope from the current node to the nearest neighbor, and i and j represent the cell and its 8 neighbors.

Four additional terrain attributes that may help in predicting the spatial distribution of soil attributes are the wetness index, w, the stream power index, Ω, a sediment transport capacity index, τ, and a landform curvature ratio, LCR. These compound indices are computed from two or more primary attributes and, in their simplest forms, can be expressed as:

$$w = \ln (A_s / \tan \beta) \quad (6)$$

$$\Omega = A_s \tan \beta \quad (7)$$

$$\tau = (A_s / 22.13)^{0.6} (\sin \beta / 0.0896)^{1.3} \quad (8)$$

$$LCR = \varphi / \bar{\omega} \quad (9)$$

where A_s is the specific catchment area ($m^2 m^{-1}$), β is the slope angle (in degrees), φ is the profile curvature or curvature in the direction of maximum slope (m), and $\bar{\omega}$ is the plan curvature (curvature transverse to this slope) (m). The first three equations all assume that A_s is proportional to the discharge per unit width (q) and that steady-state conditions apply (Moore et al. 1993a, 1993d; Moore and Wilson, 1994).

The compound topographic wetness index has been used to: 1) characterize the spatial distribution of zones of surface saturation toposequence in Colorado (Moore et al., 1988; O'Loughlin, 1981;, 2) map forest soils (Skidmore et al., 1991); and 3) delineate the spatial variability of soil properties in a toposequence in Colorado (Moore et al., 1993a, 1993b). The value of this index increases with increasing specific catchment area and decreasing slope gradient, resulting in moderate values on hilltops (flat areas with low specific catchment area), high values in valleys (high specific catchment areas and low slopes) where water concentrates, and low values on steep hillslopes (high slopes) where water drains more freely (Moore et al., 1993c). The stream power index is directly proportional to stream power, which is the time rate of energy expenditure and so is a measure of the erosive power of overland flow (Moore et al., 1993b, 1993d). The sediment transport index characterizes erosion and deposition processes and, in particular, the effects of topography on soil loss (Moore and Wilson, 1992, 1994). This index is applicable to three-dimensional landscapes and is analogous to the length-slope factor in the Revised Universal Soil Loss Equation (Renard et al., 1991). Profile curvature is a measure of the rate of change of the potential gradient and is therefore important for water flow and sediment

transport, whereas plan curvature is a measure of the convergence or divergence and hence the concentration of water in the landscape (Moore et al., 1993b). Dikau (1989), for example, used slope, plan, and profile curvature to delineate geomorphological relief units.

Any interpolation method may suffer from poor results along the edges due to limitations in data and algorithms. These edge problems are referred to as "edge effects". These edge effects were minimized in this study by clipping the DEM to an area approximately within 5 m inside the soil sampling outer boundary after the derivation of the terrain attributes. Three of the sites (52, 64, and 65) were dropped because they fell outside the clipped rectangle.

Collection and Analysis of Remotely-Sensed Data

Spectral data were captured with the ADAR System 5000 (Positive Systems, Inc., Kalispell, Montana) on June 3, 1992. The field was free of vegetation and stubble. Four image bands were acquired: band 1 was blue (450/80 nm center wavelength/bandwidth), band 2 was red (650/80 nm center wavelength/bandwidth), band 3 was red/near infrared (700/40 nm center wavelength/bandwidth), and band 4 was near infrared (850/80 nm center wavelength/bandwidth). Differing amounts of translation and rotation were removed and a 3 by 3 window (one pixel = 2.08 m^2) was averaged in an effort to reduce sample location error prior to analysis. Three bands were then ratioed to produce a 2 band image of 1/4 and 2/4 (which correspond to ADAR's blue/NIR and red/NIR bands, respectively) by:

$$\text{ratioed band DN} = [DN_1 / (DN_2 + 0.5)] * 127 \quad (10)$$

where DN_1 and DN_2 are the first and second user-specified bands. This image was created

to gauge the potential of the ADAR scanner for identifying organic matter. Frazier and Cheng (1989) have successfully used Landsat TM 1/4, 3/4, and 5/4 band ratios to map soil organic carbon in the Palouse region of eastern Washington state; however, there is little correspondence between the Landsat TM and ADAR bands. The bandwidth of ADAR's blue and red filters are much wider than the corresponding Landsat TM bands, and Landsat TM's IR bands are beyond the spectral response of ADAR. The two ADAR band ratios used more or less mimic the Landsat TM 1/4 and 3/4 band ratios used by Frazier and Cheng (1989).

Soil Survey Data

Two SCS (Soil Conservation Service) soil surveys were also available for the site: (1) an unpublished order 2 soil survey at a scale of 1:24,000, and (2) a specially requested order 1 soil survey produced by the local SCS staff at a scale of 1:7,920. The order two soil survey delineated one primary soil area across the study site with a secondary soil shown in the south-eastern quadrant. The order one survey delineated four major soil map units in four slope classes using a 0.25 ha minimum delineation (equivalent to 25 10 m by 10 m cells) from aerial photographs and field reconnaissance. Soil attributes were estimated from soil pits dug in the field and published series descriptions.

Statistical Analysis and Data Display

The soils, terrain, and soil survey variables were exported to SAS to facilitate correlation and regression analysis. The Krustal-Wallis H statistic, which is a non-parametric alternative to the F ratio for classical analysis of variances, was used to compare the variation in soil attributes between Order 1 soil survey map units. The Moran Index was used to measure the spatial structure or autocorrelation for the three soil attributes. The Moran statistic varies between -1 (negative spatial autocorrelation) and +1 (positive spatial autocorrelation) and measures the relationship among values of a single variable that is caused by the geographic arrangement of areal units or points on a map (Griffith et al., 1991). The most commonly used methods for computing spatial autocorrelation capture locational information in binary configuration tables. Sites are either adjacent to each other or not. Thickness of mollic epipedon, OM, and pH site values were interpolated to a regular 50 m grid using ARC/INFO's Inverse Distance Weighting (IDW) function and values from the three nearest sampling sites. Moran Index values were computed with the method described by Griffith et al. (1991). Stepwise multiple regression was used to identify significant relationships between the independent variables (soil survey and terrain attributes) and the dependent variables (thickness of the mollic epipedon, OM, and pH values). The F level for entry or deletion of an independent variable was set to 0.05. Statistical significance of the overall equation was determined by an F test. The t test was used to test the significance of each independent variable, and the models were evaluated for over- or under- specification using Mallow's Cp value. The grid algebra tools in the ARC/INFO software's GRID module were used to prepare the final soil maps

Results and Discussion

Table 8 summarizes soil attributes values for the four map units delineated by the order 1 soil survey and as a whole. The values from the Krustal-Wallis test statistic exceeded the critical value at the 0.05 level of significance (7.81 with 3 degrees of freedom) for all three soil properties indicating that the null hypothesis (that the means were equal) is rejected and there is evidence that the soil property mean for at least one soil map unit is not equal to the others. However, the soil map units may not help to delineate spatial variability because the range of values within each unit varies greatly for each soil property, indicating substantial variation in the three critical soil attributes (thickness of the mollic epipedon, OM, and pH values) within the map units.

Moran Index values of 0.40 - 0.46 indicate moderate clustering of similar values for the three soil attributes (Table 8). This result suggests that kriging may be a more appropriate interpolation method than a simple inverse distance weighted (IDW) interpolation for estimating the thickness of the mollic epipedon, OM and pH values. The three closest sampling sites and a 50 m grid were chosen to approximate the spacing between the initial soil sampling sites and minimize any inflation in Moran Index values caused by the conversion of the original site data to a regular 50 m grid with the ARC/INFO software's IDW function.

The multiple regression results treating OM as the dependent variable show that ADAR band ratios 2/4 and 1/4 explained 64% of the variation in OM across the field (Table

Table 8. Soil attribute values for the four map units and the entire field represented in order 1 soil survey map.

	No. of samples	Min.	Max.	Mean	S.D.	Median	Moran's I
A. Thickness of the mollic epipedon (cm)							
Alpha silt loam	29	15	80	36.4	18.0	34.0	-
Beta silt loam	7	19	66	43.1	19.4	49.0	-
Gamma Loam	9	24	126	73.3	36.9	64.0	-
Omega-Beta-Sigma complex	22	10	118	44.4	29.0	41.0	-
Entire field	67	10	126	44.7	27.3	41.0	0.45
B. OM (%)							
Alpha silt loam	29	2.3	5.0	3.6	0.7	3.7	-
Beta silt loam	7	2.1	4.0	3.2	0.7	3.1	-
Gamma Loam	9	1.1	5.6	3.8	1.3	4.2	-
Omega-Beta-Sigma complex	22	1.7	5.6	3.3	1.0	3.4	-
Entire field	67	1.1*	5.6	3.5	0.9	3.5	0.46
C. pH							
Alpha silt loam	29	5.6	6.6	6.0	0.2	6.0	-
Beta silt loam	7	5.7	7.1	6.2	0.4	6.2	-
Gamma Loam	9	5.9	7.1	6.3	0.4	6.2	-
Omega-Beta-Sigma complex	22	5.8	8.3	6.7	0.9	6.4	-
Entire field	67	5.6	8.3	6.3	0.6	6.2	0.40

* This area included an anomalous value, 1.1%, obtained at location #31. This location occurs near an inflow culvert. The OM was less then 50% of the next lowest (2.6% at location #54) and this sample site was dropped from the regression analysis.

9). The Palouse results obtained by Frazier and Cheng (1989) and Wilcox et al. (1994) using Landsat TM 3/4 and 5/4 band ratios suggest an even higher R^2 might have been obtained if the ADAR scanner had been able to better replicate the Landsat TM's IR band widths. This shortcoming may negate the timing (ADAR imagery is flown from small aircraft and can be obtained at specified dates and times) and pixel resolution (1 m^2 pixel resolution) advantages of the ADAR system relative to the Landsat TM data.

The multiple regression results treating OM, the thickness of the mollic epipedon, and pH values as dependent variables and numerous terrain attributes as independent

Table 9. Multiple regression results.

Step	Variable	Parameter Estimate	Partial R^2	Model R^2	Prob > F
	Dependent Variable: OM (%)				
	Independent Variables:				
1	BAND2/BAND4	-0.0685	0.550	0.550	0.0001
2	BAND1/BAND4	0.0533	0.093	0.643	0.0001
	Dependent Variable: OM (%)				
	Independent Variables:				
	Y-Intercept	24.1831			
1	Wetness Index	0.3883	0.332	0.332	0.0001
2	STI	-0.0455	0.068	0.400	0.0099
3	Elevation	-0.0153	0.050	0.449	0.0210
4	SLOPE2	0.0328	0.035	0.484	0.0469
	Dependent Variable: THICKNESS OF MOLLIC EPIPEDON				
	Independent Variables:				
1	Wetness Index	15.9640	0.389	0.389	0.0001
2	Slope	2.0395	0.045	0.434	0.0274
3	Plan Curvature	-1.0132	0.041	0.475	0.0300
	Dependent Variable: pH				
	Independent Variables:				
1	Elevation	-0.0177	0.069	0.069	0.0318
2	Wetness Index	-0.0903	0.059	0.128	0.0407
	Dependent Variable: OM (%)				
	Independent Variables:				
1	BAND2/BAND4	-0.5604	0.550	0.550	0.0001
2	BAND1/BAND4	0.0408	0.093	0.643	0.0001
3	Specific catchment area	-0.0002	0.032	0.675	0.0128
4	Wetness Index	0.2347	0.029	0.704	

variables are summarized in Table 9. Four terrain attributes combined to explain almost as much of the variation in OM (48.4%) as the image data (64.3%). The SLOPE2 term disappears from the response function for south-facing slopes (because the ASPECT

indicator variable was set to 0 in these cells). The exact meaning of this indicator variable can be demonstrated by examining the response functions for north- and south-facing slopes separately. The response function for the north-facing slopes can be written as:

$$E(Y) = 241831 + 0.3883\ WI - 0.0455\ STI - 0.0153\ ELEV + 0.328\ SLOPE2 \quad (11)$$

where Y is predicted OM, WI is the wetness index, STI is the sediment transport index, ELEV is the elevation in meters (above mean sea-level), and SLOPE2 is the slope gradient in degrees. This equation shows how OM increases positively with wetness index and slope, and decreases with elevation and sediment transport index on north-facing slopes. The SLOPE2 term disappears from the response function for south-facing slopes, so that:

$$E(Y) = 241831 + 0.3883\ WI - 0.0455\ STI - 0.0153\ ELEV \quad (12)$$

where Y, WI, STI, and ELEV are the same as in equation 11. This response function indicates how predicted OM increases with wetness index and decreases with elevation and sediment transport index increase on south-facing slopes. These results were similar to those of Moore et al. (1993a) who explained 48% of the variation in OM in a field in Colorado using the steady state wetness index, stream power index, and aspect as independent variables.

Three terrain attributes (wetness index, slope gradient, and plan curvature) explained 48% of the variability in the thickness of the mollic epipedon and two terrain terms (elevation and wetness index) combined to explain only 13% of the variability in pH (Table 9). The thickness of the mollic epipedon (A horizon depth) results are similar to those of Moore et al. (1993a) who predicted 50% of the variation in the mollic depth using slope and wetness index and Bell at al. (1994) who explained 51% of variation in the mollic thickness (A horizon depth) using wetness and drainage proximity terms. However, Moore at al.

(1993a) was also able to predict 41% of the variation in pH using slope gradient and plan curvature in their study area.

The final regression model (Table 9) used the image and terrain variables to predict OM. The two band ratios and two terrain variables (specific catchment area and wetness index) combined to explain 70% of the OM variation at the soil sites. This model shows: (1) the image attributes performed better than the terrain attributes in predicting OM, and (2) the terrain variables were able to explain approximately 6% of the variability that was unexplained by the image variables. The substitution of a higher resolution DEM may have produced better correlations since the grids and images used for this regression analysis varied in size and placement. Approximately 25 ADAR pixels were required to match the DEM used in the study. However Moore et al. (1993a, 1993b) argue that it is unrealistic to expect better then a 70% success rate because of inherent variations due to other soil processes (erosion, deposition, hydrology, etc.) which occur at different scales than those modeled. Moore et al. (1993b) also point out that the optimum scales for studying and characterizing landscape processes affecting the development of the soil catena are unknown and represent a major research effort. No attempt was made to generate a higher resolution DEM in this study.

The final maps show OM delineated in 1% intervals across the field. The first map (Figure 14a) was generated using the IDW function in ARC/INFO with the three nearest values. This map shows less variation than the others because 50m x 50m grid cells were used. The second map (Figure 14b) was produced with the regression equation incorporating the two ADAR band ratios and shows the most detail with a 2.08 m pixel resolution. The final two maps contain 10 m by 10 m cells (Figures 14c and 14d) and were

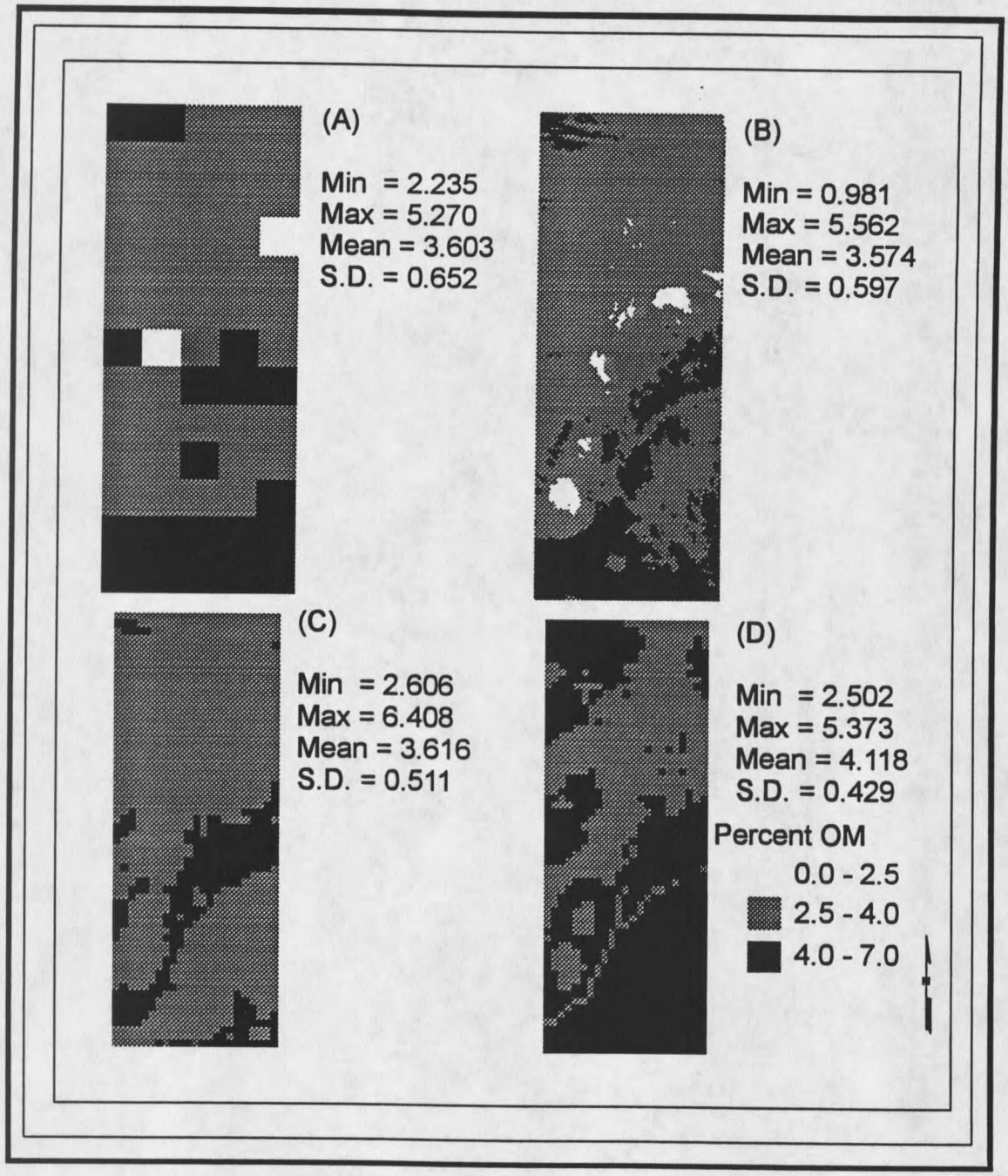


Figure 14. Organic Matter maps produced using (a) inverse distance weighting n ARC/INFO, (b) regression model using image data as independent variables, (c) regression model using terrain data as independent variables, and (d) regression model using image ratios and terrain data as independent variables.

produced using the regression equations incorporating terrain and terrain/image attributes as independent variables (see Table 9 for regression equations). As expected, the low and high OM values in Figure 14 follow the ridge and channel features; however this pattern was modified when the image and terrain regression equations were used (Figure 14d). This change can be attributed to the differences in resolution between the image and terrain grids. Overall, the four maps show how different data sources and analytical methods may lead to different estimates of economically and environmentally significant soil properties, such as percent organic matter.

Summary

Soil attribute information is needed to define farm management units, although traditional soil surveys may not be detailed enough to support precision farm management. Terrain and image data can be used to augment traditional soil surveys and interpolate finer scale soil attributes thereby reducing the manual labor required for an intensive soil survey (e.g., an order 1 soil survey). A combination of image data and derived terrain attributes was correlated against soil attribute information in a south-western Montana field. The terrain and image data produced a very high correlation (0.70 r^2) with the percent organic matter and lesser correlations with the other attributes. This high degree of correlation is the first step in using terrain and image data to help plan and model the spatially varying soil attributes for precision farm management.

This spatially varying soil attribute information is needed to divide farm fields into management units. These management units can be used with automated navigation

(guidance) systems and variable rate spreaders to apply only those quantities of seed, fertilizers, and pesticides that match the land resource potential and/or the farmers yield goals. The difficulty, expense, and uncertainty currently involved in delineating management units will limit the immediate adoption of these technological aids. More work is needed to identify significant soil properties and to compare and contrast alternative data sources and methods for estimating them and dividing farm fields into management units.

CHAPTER 4

CONCLUSIONS

The concept of "farming soils not fields" (Carr et al., 1991) encompasses much more than the traditional agricultural and soil sciences. Precision farming requires that traditional research fields must be supplemented by the latest technology and a multi-disciplinary approach to achieve success. In this study two supporting aspects of precision farming were explored by examining: (1) the effects of varying the number and pattern of GPS data on computed terrain attribute surfaces and (2) the use of terrain attribute surfaces in combination with image data to predict selected soil attributes across a field.

As research and implementation of precision farming continues, the use of GPS derived surfaces for modeling and planning precision farming increases. This requires a better understanding of the effects GPS collection methods have on the surfaces created. GPS point data (6,284 points with x, y, and elevation attributes) were collected in a "loose" grid fashion across a 20 ha farm field in south-western Montana using a truck-mounted GPS receiver. The database collected was large enough that it was sub-sampled to represent different field sampling schemes. A series of DEMs were developed using ANUDEM (Hutchinson, 1989) and terrain attributes were derived using TAPES-G (Moore et al., 1991) and ARC/INFO software's grid analysis tools. The different surfaces were compared with

each other and the surfaces created using the complete GPS dataset. The results indicate that primary terrain attributes (elevation, slope and aspect) differed little depending on the sampling scheme, however the secondary terrain attributes which reflect sediment transport and water flow varied more depending on sampling method. These data show that samples collected by linear methods produced poorer resultant surfaces than those generated from stratified random areal samples. The random samples achieved similar results using fewer data points to generate the surfaces. The linear samples tended to distort drainage patterns and landform features to favor the sampling scheme's linear direction.

Many researchers have investigated soil-landform interactions and have discovered relationships between terrain attributes, image reflectance and soil attributes. These relationships have yet to be well defined across landscapes because of the variable nature of landscapes and because researchers rarely used similar techniques and methods. Following methods used by Frazier and Cheng (1989), Moore et al. (1993a) and Gessler et al. (1994), in their image-soil and soil-landscape modeling, the attribute surfaces created using all the GPS points sampled across the field were combined with remotely sensed imagery and evaluated against measured soil attributes. Soil attributes were collected in a grid fashion across the site (70 sites, 66 used). Numerous soil attributes were collected at these sample sites and this study examined percent organic matter, pH, and thickness of the mollic epipedon. Regression models were developed for each soil attribute based on the derived terrain attributes and image values. These models resulted in r^2 values of 12% for pH, 47% for depth of the mollic epipedon, and 70% for percent organic matter. These equations were then implemented in ARC/INFO's Grid module to create soil attributes maps. The results build on previous work done by Frazier and Cheng (1989), Moore et al. (1993a and 1993b),

Wilcox et al. (1994) and Gessler et al. (1994).

References Cited

Aandahl, A.R., 1948. The Characterization of slope positions and their influence on the total nitrogen content of a few virgin soils of western Iowa. *Proceedings of the Soil Science Society of America* 13:449-454.

Anderson, M.G. and T.P. Burt. 1978. The role of topography in controlling through-flow generation. *Earth Surface Proceedings*, 3: 331-334.

Arnold, R.W. and L.P. Wilding. 1991. The need to quantify spatial variability. pp 1-8 in M.J. Mausbach and L.P. Wilding (eds.), *Spatial Variabilities of Soils and Landforms*, SSSA Special Publication Number 28. Soil Science Society of America, Madison, WI.

Barber, G.M. 1988. *Elementary statistics for geographers*. New York, NY: The Guilford Press.

Bell, J.C., J.A. Thompson, C.A. Butler and K. McSweeney. 1994. Modeling soil genesis from a landscape perspective. pp. 179-195 in *Transactions of the 15th World Conference on Soil Science*. International Society of Soil Science, Madison, WI.

Berry B. and A. Baker. 1968. Geographic Sampling. pp 91-100 in B. Berry and D. Marble (eds.) *Spatial Analysis: A Reader in Statistical Geography*. Englewood Cliffs, NJ: Prentice Hall.

Bouma, J. and P.A. Finke. 1993. Origin and nature of soil resource variability. pp. 3-13 in Robert, R.C., Rust, R.H. and W.E. Larson (eds), *Soil Specific Crop Management: A Workshop on Research and Development Issues.* Madison, WI: American Society of Agronomy.

Buol, S.W., F.D. Hole, and R.J. McCracken. 1988. *Soil genesis and classification.* (3rd edition). Ames, IA: Iowa State University Press.

Buchholz, D.D. 1991. Missouri grid sampling project. pp. 6-13 in *Proc. 21st North Central Extension-Industry Soil Fertility Conference, St. Louis, MO*. Potash & Phosphate Institute, Manhattan, KS.

Burrough, P.A. 1993. Soil Variability: a late 20th century view. *Soils and Fertilizers,* 56(5): 529-562.

Burrough, P.A., J. Bouma, and S.R. Yates. 1994. The state of the art in pedometrics. *Geoderma*, 62: 311-326.

Burt, T.P. and D.P. Butcher. 1985. Topographic control of soil moisture distributions. *Journal of Soil Science,* 36: 469-486.

Carr, P.M., G.R. Carlson, J.S. Jacobsen, G.A. Nielsen, and E.O. Skogley. 1991. Farming Soils, Not Fields: A strategy for increasing fertilizer profitability. *Journal of Production Agriculture,* 4(1): 57-61.

Cressie, N. 1991. *Statisics for spatial data.* New York, NY: John Wiley & Sons.

Dikau, R. 1989. The application of digital relief models to landform analysis in geomorphology. pp. 51-77 in Raper, J. (ed), *Three dimensional applications in geographic information systems.* New York, NY: Taylor and Francis.

Ebdon, D. 1988. *Statistics in geography.* New York, NY: Basil Blackwell.

Fairfield, J. and P. Leymarie. 1991. Drainage networks from grid digital elevation models. *Water Resources Research,* 27(5): 709-717.

Finke, P.A., J. Bouma, and A. Stein. 1992. Measuring field variability of disturbed soils for simulation purposes. *Soil Science Society of America Journal,* 56 (1):187-192.

Frazier, B.E., and Y. Cheng. 1989. Remote sensing of soils in the eastern Palouse region with Landsat Thematic Mapper. *Remote Sensing of the Environment,* 28:317-325.

Freeman, G.T. 1991. Calculating catchment area with divergent floe based on a regular grid. *Computers and Geosciences,* 17(3): 413-422.

Gessler, P.E., I.D. Moore, N.J. McKenzie, and P.J. Ryan. 1993. Soil-landscape modeling in southeastern Australia. in *Proceedings of the 2nd International Workshop on Integrating Geographic Information Systems and Environmental Modeling,* Breckenridge, Colorado (in press).

Griffith, D.A., C.G. Amrhein and J.R. Desloges. 1991. *Statistical analysis for geographers.* Englewood Cliffs, NJ: Prentice Hall.

Hewitt, A.E. 1993. Predictive modeling in soil survey. *Soils and Fertilizers,* 56(3): 305-314.

Hutchinson M.F. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology,* 106:211-232.

Hutchinson, M.F. 1991. ANUDEM user guide. Canberra; Australian National University, Centre for Resource and Environmental Studies. 6 p.

Hutchinson, M.F. and P.E. Gessler. 1994. Spline: more than just a smooth interpolator. *Geoderma.* 62: 45-67.

Isaaks, E.H. and R.M. Srivastava. 1989. *An introduction to applied geostatistics.* New York, NY: Oxford University Press.

Jersey, J.K. 1993. Assessing vegetation patterns and hydrologic characteristics of a semi-arid environment using a geographic information system and terrain based models. Department of Plant, Soil, and Environmental Science. Bozeman, MT: Montana State University, Unpubl. Masters Thesis.

Klingebiel, A.A., E.H. Horvath, D.G. Moore, and W.U. Reybold. 1987. Use of Slope, Aspect, and Elevation Maps Derived from Digital Elevation Models Data in Making Soil Surveys. in Soil Survey Techniques, SSSA special pub 20. Madison, WI: Soil Science Society of America.

Kreznor, W.R., K.R. Olson, W.L. Banwart, and D.L. Johnson. 1989. Soil, landscape, and erosion relationships in a northwest Illinois watershed. *Soil Science Society of America Journal*, 53: 1763-1771.

Lee, J. 1991. Comparison of existing methods for building triangular irregular network models of terrain from grid digital elevation models. *International Journal of Geographical Information Systems*, 5(3):267-285.

Macy, T.S. 1993. Macy Farms - site specific experiences. pp. 75 in Robert, R.C., Rust, R.H. and W.E. Larson (eds), *Soil Specific Crop Management: A Workshop on Research and Development Issues.* Madison, WI: American Society of Agronomy.

Mausbach, M.J, D.J. Lytle, L.D. Spivey. 1993. Application of soil survey Information to Soil Specific Farming. pp. 57 in Robert, R.C., Rust, R.H. and W.E. Larson (eds), *Soil Specific Crop Management: A Workshop on Research and Development Issues.* Madison, WI: American Society of Agronomy.

Moore, I.D. 1992. Terrain analysis programs for the environmental sciences: TAPES. *Agricultural Systems Information Technology,* vol 4 (2):37-39.

Moore, I.D. and M.F. Hutchinson. 1991. Spatial Extension of Hydrologic process modelling. *International Hydrology & Water Resources Symposium*: Perth, Australia.

Moore, I.D. and J.P. Wilson. 1992. Length-slope factors in the revised universal soil loss equation: Simplified method of estimation. *Journal of Soil and Water Conservation*, 47(5):423-428.

Moore, I.D. and J.P. Wilson. 1994. Reply to Comment on Length-slope factors for the Revised Universal Soil Loss Equation: Simplified method of estimation. *Journal of Soil and Water Conservation*, 49(2):174-180.

Moore, I.D., G.J. Burch and D.H. McKenzie. 1988. Topographic effects on the distribution of surface water and the location of ephermeral gullies. *Transcripts of the American Society of Agricultural Engineering,* 31:1383-1395.

Moore, I.D., R.B. Grayson, and A.R. Ladson. 1991. Digital Terrain Modelling: A review of Hydrological, Geomorphological, and Biological Applications. *Hydrological Processes,* 5:3-30.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993a. Terrain analysis for soil-specific crop management. pp. 27-56 in Robert, R.C., Rust, R.H. and W.E. Larson (eds), *Soil Specific Crop Management: A Workshop on Research and Development Issues.* Madison, WI: American Society of Agronomy.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993b. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal,* 57:443-452.

Moore, I.D., A. Lewis and J.C. Gallant. 1993c. Terrain attributes: Estimation methods and scale effects. chap. 8 in A.J. Jakeman, M.B. Beck and McAleer (eds), *Modelling Change in Environmental Systems.* New York, NY: John Wiley & Sons.

Moore, I.D., A.K. Turner, J.P. Wilson, S.K. Jenson, and L.E. Band. 1993d. GIS and Land Surface-Subsurface Process Modelling. pp. 196-230 in Goodchild, M.F., Parks, B.O., and L.T. Steyaert (eds), *Environmental Modeling with GIS.* New York, NY: Oxford University Press.

Odeh, I.O.A., D.J. Chittleborough, and A.B. McBratney. 1991. Elucidation of soil-landform inter-relationships by canonical ordination analysis. *Geoderma,* 49:1-32.

O'Callaghan, J.F. and D.M. Mark. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics and Image Processing,* 28: 323-344.

O'Loughlin, E.M. 1981. Saturation regions in catchments and their relationships to soil and topographic properties. *Journal of Hydrology,* 53: 229-246.

Panuska, J.C., I.D. Moore, and L.A. Kramer. 1991. Terrain analysis: Integration into the Agricultural Non-Point Source (AGNPS) pollution model. *Journal of Soil and Water Conservation,* 46(1): 1-32.

Quinn, P., K. Beven, P. Chevallier and O. Planchon. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes,* 5(1): 59-79.

Renard, K.G., G.R. Foster, G.A.Wieesies, and J.A. Porter. 1991. RUSLE: Revised Universal Soil Loss Equation. *J. Soil and Water Conserv.* 46(1):30-33.

Robert, P., S. Smith, W. Thompson, W. Nelson, D. Fuchs, and D. Fairchild. 1990. Soil Specific Management. pp. 54-59 in *Minnesota Agric. Exp. Stn. Misc. Publ.* 62-1990.

Schueller, J., M. Mailander, S.W. Searcy, and D.S. Motz. 1992. Working Group Report: Engineering Technology. pp. 181 in Robert, R.C., Rust, R.H. and W.E. Larson (eds), *Soil Specific Crop Management: A Workshop on Research and Development Issues.* Madison, WI: American Society of Agronomy.

Skidmore, A.K., P.J. Ryan, W. Dawes, D. Short, and E. O'Loughlin. 1991. Use of an expert system to map forest soils from a geographical information system. *International Journal of Geographic Information Systems,* 5(4):431-445.

Spangrud, D.J, J.P. Wilson, G.A. Nielsen, J.S. Jacobsen, and D.A. Tyler. 1995. Sensitivity of computed terrain attributes to the number and pattern of GPS-derived elevation data. in P.C. Robert (ed) *Soil Specific Crop Management.* Madison, WI: American Society of Agronomy (in press).

Troeh, F.R., 1964. Landform parameters correlated to soil drainage. *Proceedings of the Soil Science Society of America,* 28: 808-812.

Tyler, D.A. 1993. Positioning Technology (GPS). pp 159-165 in Robert, R.C., Rust, R.H. and W.E. Larson (eds), *Soil Specific Crop Management: A Workshop on Research and Development Issues.* Madison, WI: American Society of Agronomy.

Vreeken, W.J. 1975. Variability of depth to carbonates in fingertop loess watersheds in Iowa. *Catena,* 2: 321-336.

Webster, R. 1977. Canonical correlation in Pedology: How useful? *Journal of Soil Science,* 28: 196-221.

Webster, R. 1994. The development of pedometrics. *Geoderma,* 62: 1-15.

Wibawa, W.D. 1991. Variable rate of fertilization based on yield goal, soil fertility, and soil series. North Dakota State University Library, Fargo, ND, Unpubl.M.S. thesis.

Wilcox, C.H, B.E. Frazier, and S.T. Bell. 1994. Relationships between soil organic carbon and Ladsat TM data in eastern Washington. *Photogrametric Engineering and Remote Sensing,* 60(6): pp. 777-781.

Wilding, L.P., N.E. Smeck, and G.F. Hall (ed.) 1983. Pedogenesis and soil taxonomy. *Developments in Soil Science. vol: 11.* New York, NY: Elsevier.

Wilson, J.P, D.J. Spangrud, M.A. Landon, J.S. Jacobsen, and G.A. Nielsen. 1994. Mapping soil attributes for site-specific management of a Montana field. pp. 324. in Meyer, G.E., and J.A. DeShazer (eds), *Optics in Agriculture, Forestry, and Biological Processing.* Boston, MA: International Society for Optical Engineering.