2010

Spatio-temporal patterns of field-scale soil moisture and their implications for in situ soil moisture network design

Lingyuan Yang
Iowa State University

Spatio-temporal patterns of field-scale soil moisture and their implications for *in situ* soil moisture network design

by

Lingyuan Yang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-majors: Environmental Science; Agricultural Engineering

Program of Study Committee:
Amy L. Kaleita, Major Professor
Petrutza C. Caragea
Matthew J. Helmers
Robert Horton
Brian L. Steward

Iowa State University

Ames, Iowa

2010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation describes efforts to overcome the challenges in designing *in situ* soil moisture observation network. The surface soil moisture data collected at two spatial scales in a working field in Iowa throughout the growing seasons in 2004 to 2008 were used to describe the spatio-temporal characteristics of soil moisture at field scale. The rank stability analysis was used to identify the locations on the ground to represent the mean soil moisture at the field scale across five different growing seasons. Optimal sampling locations (OSLs), giving accurate estimates of field mean soil moisture for each season, were selected using the rank stability analysis. The results indicated that there were OSLs in the field for each growing season and for the compiled five-season, and these locations were different from each sampling season to the next, which suggested that it is not sufficient to use only one year or a few years' data to identify the soil moisture rank stable behavior using rank stability analysis. The spatial patterns of soil moisture exhibited certain consistency across multiple seasons. The OSLs all tended to be located at those locations with higher elevation. Therefore, multiple linear regression was used to predict recurring soil moisture patterns with topographic indices at optimal resolutions. A genetic algorithm was developed to select the input independent variables over a range of resolutions for multiple linear regression models. Using this approach, not only were the primary influential topographic indices to soil moisture patterns uncovered, but the most appropriate resolutions for each influential index was identified. The recurring patterns at field scale were well predicted by the combination of static topographic indices at optimal resolutions. Although the studies included in the dissertation contributed knowledge to in situ soil moisture network design, more work is

required to obtain a complete scheme for implementing a ground-based observation network effectively.

CHAPTER 1. GENERAL INTRODUCTION

1.1. Introduction

Soil moisture or soil water content is the amount of fresh water stored in the unsaturated zone above the water table. Although the water contained in soil is a small portion in comparison to other water reserves in the hydrological cycle (such as groundwater, lakes, rivers, glaciers, etc.), soil water plays an important role in many land surface processes, like the hydrological, bio-/geo-chemical, and geomorphologic processes. The surface soil moisture is an important medium for both water and energy exchanges between the atmosphere and the land surface. It controls the processes of partitioning solar energy into latent and sensible heat [*Western et al.,* 2003] and the processes of turning precipitation into surface runoff, infiltration, and evapotranspiration [*Krajewski et al.,* 2006]. Soil moisture modulates plant growth and hence primary production in terrestrial ecosystems and has an important influence on a variety of soil processes including erosion, soil chemical processes and solute transport, and ultimately pedoenesis [*Western et al.,* 2003]. Surface soil moisture can also be used to retrieve soil moisture at root zone depth and soil profile depth and to estimate soil hydraulic properties (such as hydraulic conductivity) with modeling [*Vereecken et al.,* 2008]. Accurate and reliable estimates of soil moisture are important for characterizing and modeling the above processes.

Three major approaches are used to estimate soil moisture contents in a region. These are remote sensing, *in-situ* or ground-based measurement, and land surface modeling. They all have different advantages and limitations. Remote sensing has been used to capture the temporal and spatial characteristics of surface soil moisture in various study areas [*Mohanty* *et al.*, 2001; *Bosch et al.*, 2006; *Cosh et al.*, 2006; *Choi et al.*, 2008]. Remote sensing can cover a large spatial extent and collect soil moisture data in all weather conditions [*Jackson,* 1993]. However, the large spatial coverage (known as a footprint) is generally combined with low spatial resolution, thus, remotely sensed data is unable to capture the heterogeneity of soil moisture within the remote sensing footprint. Also, current microwave sensing technologies respond to only a shallow depth (the first few inches of the surface). Another limitation is that remote sensors do not measure soil moisture directly, and the remote sensing signals have to be converted to soil water content values using retrieval algorithms. These algorithms need to be validated with accurate *in situ* soil moisture measurements. The interpretation of the remotely sensed signals is often difficult. Specifically, there are a number of confounding factors such as vegetation characteristics and soil texture that may affect the remotely sensed signal much more strongly than does the actual soil moisture. Ground-based techniques can provide precise, point-scale soil moisture measurements. At the same time, obtaining accurate and representative data of an entire region can be time- and resource-consuming. Another popular way to obtain soil moisture information in a region is using land surface models, which is hampered by limited measurements of physical parameters [*Mohr et al.*, 2000; *Whitfield et al.*, 2006] and input data errors [*Reichle and Koster,* 2004; *Reichle et al.,* 2004]. Land surface models also require *in situ* measurements for comparison to the simulated soil moisture for the purpose of quality control and evaluation [*Robock et al.,* 2003].

These three approaches are complementary in many ways. Utilizing soil moisture data collected *in situ* as well as remotely sensed data and land surface model output provides a unique opportunity to obtain accurate and reliable soil moisture information. However, there is mismatch in scales between remote sensing footprints and ground samples [*Western and Blöschl*, 1999], as well as between the input grid size of land surface models and ground samples. It is not practical to do exhaustive ground sampling. Guidance or prior knowledge is critical for choosing a limited number of optimal sites on the ground to use as long-term and permanent *in situ* soil moisture network locations. The challenging questions for designing an in situ soil moisture network include:

- Where the soil moisture sampling locations should be;
- How many of sampling locations are needed to accurately upscale point observations to areal values of an area of interest (in other words, to mitigate the up-scaling error); and
- How readily available data, for example, topography, can be used to predict soil moisture patterns *a priori* in the design of an *in situ* soil moisture network.

This dissertation addresses these challenges associated with *in situ* soil moisture network design.

Vachaud et al. [1985] found that some locations consistently represented the mean and extreme values respectively of field water content, introducing the concept of rank stability in soil moisture monitoring. The locations consistently representing the mean an extreme values are considered rank stable. Some researchers have used rank stable analysis provide by *Vachaud et al.* [1985], discussed in details in Chapter 3, to upscale point soil water contents on the ground to areal moisture within remote sensing footprints [*Cosh et al.*, 2004; *Cosh et al.*, 2006; *Vivoni et al.*, 2008], in order to overcome the mismatch of scales between different measurement approaches when using them together. Rank stable locations will provide an important basis for *in situ* soil moisture network deployment and operation with regards to remote sensing validation [*Cosh and Jackson, 2009*].

To make the rank stability analysis valuable for long-term soil moisture monitoring, more studies are needed to determine if the rank stable locations are consistent across the multiple years. Only a few studies, such as *Martinez-Fernandez and Ceballos* [2003] and *Schneider et al.* [2008], have analyzed rank stability over multiple seasons or years. These studies have suggested that the rank stable locations might lose their temporal rank persistence from year to year. *Schneider et al.* [2008] remain open to debate whether more accurate rank-stable locations can be obtained with a time series longer than two years. Therefore, further studies are needed to confirm the applicability of the rank stability concept for study periods longer than two years.

Rank stability analysis has been tested on the long-term data at intermediate scale in this dissertation. Moreover, research is still needed to determine how to rapidly identify these rank stable locations *a priori* given limited data [*Robinson et al.*, 2008].

Numerous studies have related the soil moisture patterns to various factors/variables for different study regions in different seasons. [*Western et al.,* 1999; *Gómez-Plaza et al.,* 2001*; De Lannoy et al.,* 2006], and they have only qualitatively explained how each factor or variable influenced soil moisture patterns. Some studies have investigated the soil moisture variability as a quantitative function of those influential factors affecting the soil moisture, including empirical, physical and statistical models. These models require detailed information concerning characteristics of soil, vegetation, rainfall, etc. For example, *Isham et al.* [2005], *Rodriguez-Iturbe et al.* [2006] and *Manfreda and Rodriguez-Iturbe* [2006] have used an analytically derived covariance function to characterize spatial pattern of soil moisture, which accounted in great detail for soil characteristics, vegetation patterns, and rainfall dynamic in a region but neglected the topography. But it is still a key challenge to find better ways to characterize the soil and vegetation [*Wilson et al.,* 2004].

Topography is the only readily available information [*Wilson et al,* 2004] that is, relatively inexpensive to obtain, only needs to be obtained once, and can be determined at very high resolution. Digital elevation models and the technology for handling such data (for example, ARCGIS) are available and mature. It is critical to look for a way to use this readily available information and technology to predict the recurring static soil moisture pattern [*Starr*, 2005; *Kaleita et al.*, 2007; *Guber et al.*, 2008] *a priori*.

Fewer studies have used topographic indices to predict soil moisture patterns. *Western et al.* [1999] used multiple linear regression to model soil moisture pattern with terrain indices (wetness index, tangent curvature and potential radiation index) at Tarrawarra catchment in Australia, which captured 61% of the soil moisture spatial variation under wet conditions and up to 22% under dry conditions.

The resolution of topographic factors to be used in predicting recurring soil moisture patterns has never been considered. It is important to know to which resolutionthe available detailed topographic information to be included, since the high resolution data may not be necessary or each influential variable may have different resolution requirement. In this dissertation, the influential factors are considered, as well as the optimal combination of resolutions for these factors. It is difficult to identify the best factors at the best resolutions for the functions using traditional multiple linear regression, because there is huge number of possible combinations of the factors and their resolutions and it is not practical to test every possible combination. Therefore, an effective optimization method, genetic algorithm (GA), has been used.

Genetic algorithms (GAs) are adaptive heuristic search algorithms premised on the evolutionary ideas of natural selection and genetics. The basic concept of GAs is designed to simulate the Darwinian concept of "survival of the fittest." GAs are computationally simple yet powerful to provide robust search for difficult combinational search problems in complex spaces, without being stuck in local extremes [*Goldberg,* 1989a; *Tang, et al.,* 1999; *Steward, et al.,* 2005]. Therefore, GAs are a powerful alternative tools to traditional optimization methods [*Goldberg,* 1989a]. Using this methodology, it is possible to identify the most influential factors on soil moisture patterns, as well as the optimal resolutions of these factors.

The overall goal of this dissertation is to contribute efforts to overcome the existing challenges in designing *in situ* soil moisture network. The research objectives were to:

1. identify the optimal sampling locations (OSLs), which give accurate estimates of field mean soil moisture to bridge the gap between point-scale and areal scale.
2. predict recurring soil moisture patterns using readily available topographic data.

1.2. Dissertation Overview

This dissertation is presented as a compilation of three articles, including one article in the review process for refereed publication, one article about to be sent out for refereed publication, and one article that compiles important findings and adds context to the data presented in the other two papers. Chapter Two, "Spatio-temporal characteristics of soil moisture across two scales in an agricultural field", describes the temporal and spatial characteristics of the soil moisture data included in the other two chapters which are both framed as journal articles. Chapter Three, “Optimal sampling locations of soil moisture for validating remote sensing at field scale across multiple seasons”, describes using rank stability analysis to identify the optimal sampling locations and uses them to upscale the point-scale soil moisture content to an areal value, bridging the gap between different measurement approaches of soil moisture. This paper has been submitted to Journal of Hydrology for review. Chapter Four, “Genetic algorithm for parameter selection to predict soil moisture pattern using topographic indices”, describes the development of genetic algorithm for selection of the most influential topographic factors on underlying soil moisture patterns, as well as identification of the optimal resolutions of these factors. This paper is to be submitted to Transactions of ASABE.

1.3. References

Bosch, D.D., V. Lakshmi, T.J. Jackson, M. Choi, and J.M. Jacobs, 2006. Large scale measurements of soil moisture for validation of remotely sensed data: Georgia soil moisture experiment of 2003. *J. Hydrol.*, 323, 120-137.

Choi, M., J. M. Jacobs, and D. D. Bosch, 2008. Remote sensing observatory validation of surface soil moisture using Advanced Microwave Scanning Radiometer E, Common Land Model, and ground based data: Case study in SMEX03 Little River Region, Georgia, U.S., *Water Resour. Res.*, 44, W08421, doi:10.1029/2006WR005578.

Cosh, M.H., T.J.Jackson, R. Bindlish, and J.H. Prueger, 2004. Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sens. Environ.*, 92, 427-435.

Cosh, M.H., T.J. Jackson, P. Starks, and G. Heathman, 2006. Temporal stability of surface soil moisture in the Little Washita River watershed and its appplicaions in satellite soil moisture product validation. *J. Hydrol.*, 323, 168-177.

Cosh, M.H. and T.J. Jackson, 2009. In situ validation of soil water content from satellite remote sensing. *ASA-CSSA-SSSA Annual Meeting*, Nov. 1-5.

De Lannoy, G. J. M., N. E. C.Verhoest, P. R. Houser, T. J. Gish, and M. Meirvenne, 2006. Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE3), *J. Hydrol.*, 331, 719–730.

Goldberg, D. E., 1989a. Genetic Algorithms in Search, Optimization and Machine Learning. Reading, Mass.: Addison-Wesley Publishing Co., Inc.

Gómez-Plaza, A., M. Martınez-Mena, J. Albaladejo, and V. M. Castillo, 2001. Factors regulating spatial distribution of soil water content in small semiarid catchments. *J. Hydrol.*, 253, 211–226.

Guber, A. K., T. J. Gish, Y. A. Pachepsky, M. T. van Genuchten, C. S. T. Daughtry, T. J. Nicholson, and R.E. Cady, 2008. Temporal stability in soil water content patterns across agricultural fields, *Catena*, 73, 125–133.

Isham, V., D. R. Cox, I. Rodríguez-Iturbe, A. Porporato, and S. Manfreda, 2005. Mathematical characterization of the space-time variability of soil moisture, *P. R. Soc. A*, 461(2064), 4035 – 4055, doi:10.1098/rspa.2005.1568.

Jackson, T. J., 1993. Measuring surface soil moisture using passive microwave remote sensing, *Hydrol. Process.*, 7, 139–152.

Kaleita A. L., M. C. Hirschi, and L. F. Tian, 2007. Field-scale surface soil moisture patterns and their relationship to topographic indices. *Trans. ASABE*, 50(2), 1-8.

Krajewski, W.F., M.C. Anderson, W.E. Eichinger, D. Entekhgabi, B.K. Hornbuckle, P.R. Houser, G.G. Katul, W.P. Kustas, J.M. Norman, C. Peters-Lidard, and E.F. Wood, 2006. *Water Resour. Res.*, 42, W07301, doi:10.1029/2005WR004435.

Manfreda, S., and I. Rodríguez-Iturbe, 2006. On the spatial and temporal sampling of soil moisture fields, *Water Resour. Res.*, 42, W05409, doi:10.1029/2005WR004548.

Mohanty, B.P., and T.H. Skaggs, 2001. Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Adv Water Resour.*, 24, 1051-1067.

Mohr, K. I., J. S. Famiglietti, A. Boone, and P. J. Starks, 2000. Modeling soil moisture and surface flux variability with an untuned land surface scheme: A case study from the Southern Great Plains 1997 Hydrology Experiment, *J. Hydrometeorol.*, 1: 154–169.

Reichle, R., and R. Koster, 2004. Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, 31, L19501, doi:10.1029/2004GL020938.

Reichle, R., R. Koster, J. Dong, and A. Berg, 2004. Global soil moisture from satellite observations, land surface models, and ground data: Implication for data assimilation, *J. Hydrometeorol.*, 5, 430–442.

Robinson, D. A., C. S. Campbell, J. W. Hopmans, B. K. Hornbuckle, S. B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth, 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7 (1), 358-389.

Robock, A., et al., 2003. Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season, *J. Geophys. Res.*, 108(D22), 8846, doi:10.1029/2002JD003245.

Rodríguez-Iturbe, I., V. Isham, D. R. Cox, S. Manfreda, and A. Porporato, 2006. An analytical framework for the space-time variability of soil moisture fields with heterogeneous vegetation, *Water Resour. Res.*, 42, W06D05, doi:10.1029/2005WR004497.

Schneider, K., J.A. Huisman, L. Breuer, Y. Zhao, and H.-G. Frede, 2008. Temporal stability of soil moisture in various semi-arid steppe ecosystems and its application in remote sensing. *J. Hydrol.*, 359 (1-2), 16-29.

Starr, G. C., 2005. Assessing temporal stability and spatial variability of soil water patterns with implications for precision water management, *Agric. Water Manag.* 72, 223–244.

Steward, B.L., A.L. Kaleita, R.P. Ewing, and D.A. Ashlock, 2005. Genetic algorithms for hyperspectral range operator selection, *ASABE paper No. 053063.* St. Joseph, Mich.: ASABE.

Tang, L., L. F. Tian, and B. L. Steward, 2000, Color image segmentation with genetic algorithm for in-field weed sensing*, Trans. ASAE,* 43, 1019-1027.

Vachaud, G., A. Passerat de Silans, P. Balabanis, and M. Vauclin, 1985. Temporal stability of spatially measured soil water probability density function. *Soil Sci. Soc. Am. J.*, 49, 822-828.

Vereecken, H., J. A. Huisman, H. Bogena, J. Vanderborght, J. A. Vrugt, and J. W. Hopmans, 2008. On the value of soil moisture measurements in vadose zone hydrology: A review, *Water Resour. Res.*, 44, W00D06, doi:10.1029/2008WR006829.

Vivoni, E.R., M. Gebremichael, C.J. Watts, R. Bindlish, and T.J. Jackson, 2008. Comparison of ground-based and remotely-sensed surface soil moisture estimates over complex terrain during SMEX04. *Remote Sens. Environ.*, 112, 314-325.

Western, A. W., and G. Blöschl, 1999. On the spatial scaling of soil moisture. *J. Hydrol.*, 217, 203–224.

Western, A. W., R. B. Grayson, G. Blöschl, G. R. Willgoose, and T. A. McMahon, 1999. Observed spatial organization of soil moisture and its relation to terrain indices, *Water Resour. Res.,* 35(3), 797–810.

Western, A.W., R.B. Grayson, G. Blöschl, and D.J. Wilson, 2003. Spatial variability of soil moisture and its implications for scaling. In: Perchgepsky, Y., Selim, M., Radcliffe, A. (Eds.), Scaling Methods in Soil Physics. CRC Press, Boca Raton, pp. 119-142.

Whitfield, B., J. M. Jacobs, and J. Judge, 2006. Intercomparison study of the Land Surface Process model and the Common Land Model for a prairie wetland in Florida, *J. Hydrometeorol.*, 7(6), 1247–1258.

Wilson, D. J., A. W. Western, and R. B. Grayson, 2004. Identifying and quantifying sources of variability in temporal and spatial soil moisture observations, *Water Resour. Res.*, 40, W02507, doi:10.1029/2003WR002306.

CHAPTER 2. SPATIO-TEMPORAL CHARACTERISTICS OF SURFACE SOIL MOISTURE ACROSS TWO SCALES IN AN AGRICULTURAL FIELD

2.1. Introduction

Soil moisture is a key variable in hydrologic processes at the land surface. It has a major influence on a wide spectrum of hydrological processes including flooding, erosion, solute transport, and land-atmosphere interactions, as well as agricultural and ecologic applications. Numerous studies have focused on capturing the spatial variability of soil moisture across the areas of study [*Crave and Gascuel-Odoux*, 1997; *Grayson and Western*, 1998; *Famiglietti et al.*, 1999]. These studies were conducted at different spatial scales (1 m^2 to a few km^2), at different temporal scales (few days to few years), in a variety of hydrologic and climatic conditions [*Hupet and Vanclooster*, 2002], and with different measurement techniques (e.g. *in situ* techniques and remote sensing). Due to the complexity of the natural environment and the factors governing soil moisture, it highly varies in both time and space which makes accurately capturing its variability difficult. Therefore, knowledge of the characteristics of the temporal and spatial variability is important for guiding a soil moisture campaign, and for understanding and predicting/modeling the above processes.

2.2. Materials and Methods

2.2.1. Study Area and Data

Distributed near-surface soil moisture data was collected across an agricultural field, Brooks Field, southwest of Ames, Iowa, during the growing seasons in 2004-2008. Brooks is a 10-ha corn-soybean rotation field with moderate topographic variation.

There were a total of 78 sampling locations in the entire field for 2004, 2005, 2007, and 2008 (Figure 2.1), which include 42 regular grid nodes with a 50-m interval and two transects with a 5-m interval. Transect 1 and Transect 2 were located in different landscape positions. Six of those locations labeled differently in Figure 2.1, were not sampled in 2006. Also, two transects were not included in 2006, either. The sampling time window during the day was consistent within each season and slightly different across the seasons, but in all cases was limited to a maximum of two hours in order to minimize soil moisture differences due to drying.

At each sampling location, three readings for 0-6 cm depth were taken using a Theta probe moisture meter (Delta-T Devices, Cambridge UK, marketed in the United States by Dynamax, Inc., Houston, Tex.). The Theta probe output voltage values were converted to volumetric soil moisture using the calibrated relationship for Des Moines Lobe soils provided by *Kaleita et al.* [2005]; the field component of that study was also conducted at the Brooks field. The average value of the three soil moisture contents at each location was used as the representative soil moisture at that location on that date for the analysis in this study.

In each season, data collection was begun only after planting (usually May); data collection in 2008 was delayed because heavy spring rains delayed planting. In every growing season, there were missing data points on some sampling days due to the presence of standing water on the site; missing data was generally only in the two "potholes" in the field, where water tended to pond. For the days when data was missing for more than one sampling location, the sampling day was not included in further analysis for the sake of consistency. In this study, one day was excluded for this reason in 2004, four days in 2005, and four days in 2006. This may bias the results slightly, since generally the days excluded from analysis were very wet. When these data are excluded, there remain a total of twenty four days' data (from mid-May to early July) in 2004, twenty six days' data (from early-May to Mid-August) in 2005, twenty six days' data (from mid-May to mid-July) of data in 2006, twenty nine days' data (late May to Mid-August) in 2007, and twenty days' data in 2008 (late June to late August) available for analysis. On average, soil moisture data were collected twice a week in 2005, 2007 and 2008, and three times a week in 2004 and 2006. Data collection overall captured days with wet and with dry soil moisture conditions, as well as the dry out process.

According to NRCS web soil survey (http://websoilsurvey.nrcs.usda.gov/app/), Brooks has five soil types (Figure 2.1): Nicollet loam, Harps loam, Webster clay loam, Clarion loam, and Canisteo clay loam. The description of characteristics of each soil type is included in Table 2.1.

Elevation data was obtained by LiDAR (**Li**ght **D**etection **a**nd **R**anging) survey at a 2-m horizontal resolution at Brooks Field (Figure 2.1). LiDAR is an optical remote sensing technology that measures properties of scattered light to find range and/or other information of a distant target. The prevalent method to determine distance to an object or surface is to use laser pulses. Like the similar radar technology, which uses radio waves instead of light, the range to an object is determined by measuring the time delay between transmission of a pulse and detection of the reflected signal. LiDAR has been used in particular as a transformative method of obtaining detailed topographic information. It is attractive because it can penetrate through many vegetation types, while obtaining information on vegetation height and density [*Robinson et al.*, 2008].

Precipitation data was obtained from the Ames Station (UTM (WGS84): 433776.8E, 4653416.7N) at the Iowa Environmental Mesonet (IEM) observing network (http://mesonet.agron.iastate.edu/). This station is about 10 km away from Brooks Field. Total precipitation amount for each growing season was summarized in Table 2.2.

2.2.2. Temporal characteristics of soil moisture

Basic characteristics of temporal and spatial variability of soil moisture were identified by using the standard deviation and coefficient of variation. The coefficient of variation was calculated by standard deviation of soil moisture, Θ, being scaled by the daily field mean soil moisture ($CV = SD(\theta)/\overline{\theta}$).

Nonparametric Spearman's test was applied here to identify the rank stable characteristics between two dates. The Spearman rank correlation coefficient was calculated by:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(R_{ij} - R_{ij'})^2}{n(n^2 - 1)} \quad (1)$$

Where, R_{ij} is the rank of the variable (soil moisture) θ_{ij} observed at location i on date j and $R_{ij'}$ is the rank of the variable $\theta_{ij'}$ at location i on date j', and n is the number of observations. A value $r_s = 1$ will correspond to identity of rank for any site, or perfect rank stability between dates j and j'. The closer r_s is to 1, the more stable the process will be.

2.2.3. Spatial characteristics of soil moisture

Geostatistical analysis was used to illustrate the commonality and the difference of the spatial patterns of soil moisture from day to day, and season to season.

In order to characterize the spatial variability across the field, the semivariogram of the spatially distributed data is required. The semivariogram measures the average dissimilarity between data separated by a certain distance. The estimated semivariogram for a given dataset is described by the equation:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[z(u_\alpha) - z(u_\alpha + h) \right]^2 \quad (2)$$

where $\gamma(h)$ is the empirical semivariogram, $N(h)$ is the number of pairs of data locations which are separated by the distance h, and $z(u_\alpha)$ is the data value at location u_α.

Semivariograms for each data set were computed, and subsequently a model was fit to each date's results according to the least square technique. Three types of permissible models, spherical, cubic, and wave models were used to describe the semivariograms. The spherical model with range **a** is described by the equation:

$$g(h) = \begin{cases} 1.5\frac{h}{a} - 0.5(\frac{h}{a})^3, if : h \le a \\ 0, otherwise \end{cases} \quad (3)$$

The cubic model with range **a** is described by the equation:

$$g(h) = \begin{cases} 7(\frac{h}{a})^2 - 8.75(\frac{h}{a})^3 + 3.5(\frac{h}{a})^5 - 0.75(\frac{h}{a})^7, if : h \le a \\ 0, otherwise \end{cases} \quad (4)$$

And, the wave model with range **a** is described by the equation:

$$g(h) = \begin{cases} \frac{a}{h} * \sin(\frac{h}{a}), if : h \leq a \\ 0, otherwise \end{cases} \quad (5)$$

The spherical model is the most widely used semivariogram model and is characterized by a linear behavior at the origin. The cubic model displays a parabolic behavior at the origin. The best fit semivariagram model would be finally selected for each sampling day.

Optimal cell size for precision management is calculated based upon the variogram parameters. *Han et al.* [1994] computed the maximum cell size equivalent to the mean correlation distance, which is expressed as

$$MCD = \int_0^{h_{max}} \rho(h) dh \quad (6)$$

Where h_{max} is the maximum distance between sampled locations. The normalized complement function $\rho(h)$ is related to the semivariogram $\gamma(h)$ through the equation

$$\rho(h) = \frac{(C_0 + C) - \gamma(h)}{C_0 + C} \quad (7)$$

Where, $(C_0 + C)$ is called sill which is the level at which the semivariogram flattens out, and C_0 is called nugget which represents small scale variability and measurement errors. The range is the distance beyond which the correlation between points is minimal. For situations in which h_{max} exceeds the semivariogram range a, the preceding equations (6) and (7) reduce to the following equation for a spherical semivariogram model,

$$MCD = \frac{3Ca}{8(C_0 + C)} \quad (8)$$

2.3. Temporal characteristics

2.3.1. Descriptive temporal statistics

The temporal average and standard deviation at individual sampled location describe the overall wetness conditions throughout the investigation period and the response of surface soil moisture to precipitation and evapotranspiration.

Figure 2.2 shows the temporal average versus standard deviation for regular-grid and two-transect locations for each growing season and compiled 5 or 4 seasons. In each season, there were different characteristics of soil moisture temporal variability. The standard deviation of soil moisture in time was generally stable without changing for the each separated investigation period except 2006. In 2006 growing season, the temporal variability increased with temporal mean soil moisture conditions. In 2004 and 2008, all the locations had less temporal variability than the rest of the seasons, with all locations except one in 2004 having standard deviation greater than 0.04 cm^3/cm^3. In 2006 growing season, there was least amount of precipitation within the investigation period, and there were the greatest temporal variability of soil moisture at the sampled locations. The temporal mean soil moisture at Transect 1 locations were always drier than that at Transect 2 locations. It is not possible to conclude that there were different characteristics of temporal variability across two scales in Brooks field.

Figure 2.3 shows the temporal average versus coefficient of variation for regular-grid and two-transect locations for each growing season and compiled 5 or 4 seasons. The coefficient of variation decreased with increasing temporal mean soil moisture conditions for 2004, 2005, 2007 and compiled seasons. In 2006 and 2008, there was no obvious relationship between coefficient of variation and temporal mean soil moisture.

2.3.2. Temporal correlation

The Spearman's rank correlation coefficients for regular-grid sampled locations were calculated. And the average of Spearman's coefficients and standard deviation versus each time lag were plotted in Figure 2.4. Generally, the spatial soil moisture was more rank stable throughout 2004 and 2008 growing seasons. It was not possible to characterize the temporal scales with the available soil moisture data in Brooks field.

2.4. Spatial characteristics

2.4.1. Descriptive spatial statistics

It was desirable to identify relationships between the standard deviation or the coefficient of variation of moisture content and its mean values within an area. because mean moisture content changes with time, such relationships can be used to characterize the spatial variability of soil moisture, to determine changes in the number of samples required to estimate the mean moisture content with a specified limit of error, to estimate changes in the limit of error associated with a prescribed number of samples, to estimate the variability of surface moisture content within an area of land surface given its remotely sensed mean, and to infer changes in the accuracy and precision of remote sensing.

Studies have been done to model standard deviation of soil moisture in space versus spatial mean soil moisture theoretically and empirically [*Western et al.*, 2002; *Famiglietti et* *al.*, 2008]. It has been observed in many study areas that the standard deviation increased with mean soil moisture under drier conditions and then decreased under wetter conditions. Figure 2.5 shows the standard deviation of soil moisture in space versus spatial mean soil moisture in Brooks. In 2005 and 2005, the relationships between standard deviation and mean soil moisture were similar to previous findings. In 2004, 2007, and 2008, there might not be sufficient data to discover such relationship due to the lack of data under either drier or wetter conditions.

Some researchers have discovered an exponential relationship between CV and spatial mean soil moisture across different study area at different scales [*Jacobs et al.*, 2004; *Choi et al.*, 2007; *Famiglietti et al.*, 2008]. Figure 2.6 shows CV in space versus spatial mean soil moisture at Brooks field. Although exponential model was not the best option to describe the relationship between CV and daily spatial mean soil moisture, CV decreased with increasing mean soil moisture for most of the investigation periods except 2008 at Brooks.

2.4.2. Spatial correlation

In order to assess the spatial correlation, semivariograms were calculated, and three permissible semivariogram models (spherical, wave, cubic) were compared using weighted sum squares (WSS) of errors between experimental semivariograms and the model semivariograms. Although the wave models yielded the smallest WSS for the most of the dates, the ranges given by the wave models, which were around 30 meters, tended to be too small for the real condition. The WSS of spherical and cubic models were close for every day's data, and the spherical models were selected for all the dates. Another justification for using the spherical model is that most other research seems to have used it, and thus if we use it, we enable comparison between our findings and others.

The shapes of the semivariogram on most of the sampling days were similar, and most of them reach a maximum around 150 m before dipping and fluctuating around a sill value, which was defined as "hole effect" by *Goovaerts* [2000]. A few examples were shown in Figure 2.7. The so-called "hole effect" typically reflects pseudo-periodic or cyclic phenomena [*Goovaert*， 2000；*Journel and Huijbregts,* 1978, p. 403]. In Brooks field, the hole effect relates to the existence of two lowest depressions in the field (Figure 2.1) which created two high valued areas of soil moisture.

Ranges from the fitted spherical models were plotted versus spatial mean soil moisture in Figure 2.8. There was no obvious relationship between range or correlation length and mean soil moisture conditions. For most of the sampling days, the ranges from the fitted models were stable except on a few days (either wetter or drier) when the ranges were much smaller. *Western et al.* [2004] had found out that the correlation length were related to the mean soil moisture in a few catchments. Their results did not apply to Brooks field.

The sills from the fitted spherical models were plotted versus mean soil moisture in Figure 2.9. The sills were more related to mean soil moisture than the ranges. In general, the sills increased and the decreased with mean soil moisture increasing.

2.4.3. Management zone size (Mean correlation Distance)

Mean correlation distances, derived from geostatistical analysis, were used as the upper limits of optimal cell sizes of management zones by *Han* [1994]. MCD is useful information for site-specific crop management. MCDs were plotted versus mean soil moisture in Figure 2.10. MCDs were pretty stable throughout each investigation periods except on a few days. In 2006, there were five days under drier conditions, when MCD were smaller. And in 2007, there was one day under wetter condition, when MCD were much smaller. Throughout 125 sampling days in 5 seasons, more than 80% of MCDs were within range from 40 to 60 meters. A maximum cell size of 50 meters might be appropriate scale for precision farming operations.

2.5. Relation of soil moisture to elevation

For the empirical modeling purpose, it is useful to relate the spatial variability of soil moisture to topographic variables, such as elevation, which is readily available and easily obtained information. Figure 2.11 shows the correlation coefficients between daily soil moisture and elevation for regular grids and two-transect locations. The elevation had different impact on daily soil moisture at different spatial scales. For regular grids, the correlation coefficients between daily soil moisture and elevation were always negative with different levels throughout all 5 growing seasons. The transects at two different landscape positions had different correlation relationships between soil moisture and elevation. Transect 1, located at concave landscape, were mostly positively related to elevation, and whereas, Transect 2, located at convex landscape, were mostly negatively related to elevation throughout 5 seasons.

2.6. Summary

Temporally, the standard deviation of soil moisture at each sampled location was small for all the investigation period except 2006. In 2006 the standard deviation increased with increasing mean soil moisture. There were no obvious different characteristics of temporal variability, represented by standard deviation, across two scales in Brooks field. The coefficient of variation at Transect 1 was always greater than Transect 2. For bigger spatial scale at the regular grids, the spatial soil moisture was more rank stable in 2004 and 2008 than the rest seasons.

Spatially, Transect 1, located at a concave landscape position, always had smallest spatial variability under various mean soil moisture conditions. The spatial variability, presented by standard deviation, had the highest values under the medium mean soil moisture conditions. The coefficient of variation decreased when increasing spatial mean soil moisture. The semivariagrams had similar shapes and parameters for most of the sampling days. Most of the ranges or correlation lengths were within the range of 120 m to 160 m, and only for a few days the ranges were much smaller either under the drier or wetter conditions. The sill had highest values at medium mean soil moisture condition. More than 80% of mean correlation distances, derived from fitted semivariogram models, were within 40 m to 60 m, which indicated the upper limits of the management zone sizes in Brooks field.

There were different correlation relationships between elevation and soil moisture at different scales. At the transect scale, these correlation relationship also changed through time.

2.7. References

Choi, M., J.M. Jacobs, and M.H. Cosh, 2007. Scaled spatial variability of soil moisture fields. *Geophys. Res. Lett.,* 34, L01401, doi:10.1029/2006GL028247.

Crave, A. and Gascuel-Odoux, C., 1997. The influence of topography on time and space distribution of soil surface water content. *Hydrol. Process.,* 11, pp. 203–210

Famiglietti, J. S., J. A. Devereaux, T. J. Jackson, S. T. Graham, C. A. Laymon, M. Rodell, T. Tsegaye, P. R. Houser, and P. J. van Oevelen, 1999. Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resour. Res.,* 35 (6), 1839–1851.

Famiglietti, J. S., D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson, 2008. Field observations of soil moisture variability across scales, *Water Resour. Res.*, 44, W01423, doi:10.1029/2006WR005804.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.*, 228, 113-129.

Grayson, R.B., and A.W. Western, 1998. Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *J. Hydrol.*, 207, 68-82.

Han, S., J. W. Hummel, C. E. Goering, and M. D. Cahn, 1994. Cell size selection for site-specific crop management. *Trans. ASABE,* 37(1), 19-26.

Hupet, F., and M. Vanclooster, 2002. Intraseasonal dynamics of soil moisture variability within a small agricultural maize cropped field. *J. Hydrol.*, 261:86–101.

Jacobs, J. M., B. P. Mohanty, E. Hsu, and D. Miller, 2004. SMEX02: Field scale variability, time stability and similarity of soil moisture. *Remote Sens. Environ.*, 92, 436-446.

Journel, A.G. and Huijbregts, C.J., 1978. . Mining Geostatistics Academic Press, New York.

Kaleita, A. L., J. L. Heitman, S. D. Logsdon. 2005. Field calibration of the Theta Probe for Des Moines Lobe soils. *Applied Engineering in Agriculture*, 21(5), 865-870.

Western, A.W., R.B. Grayson, and G. Blöschl, 2002. Scaling of soil moisture: a hydrologic perspective. *Ann. Rev. Earth Planet. Sci.*, 205: 20–37.

Western, A.W., S. Zhou, R.B. Grayson, T.A. McMohon, G. Blöschl, and D.J. Wilson, 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *J. Hydrol.,* 286, 113-134.

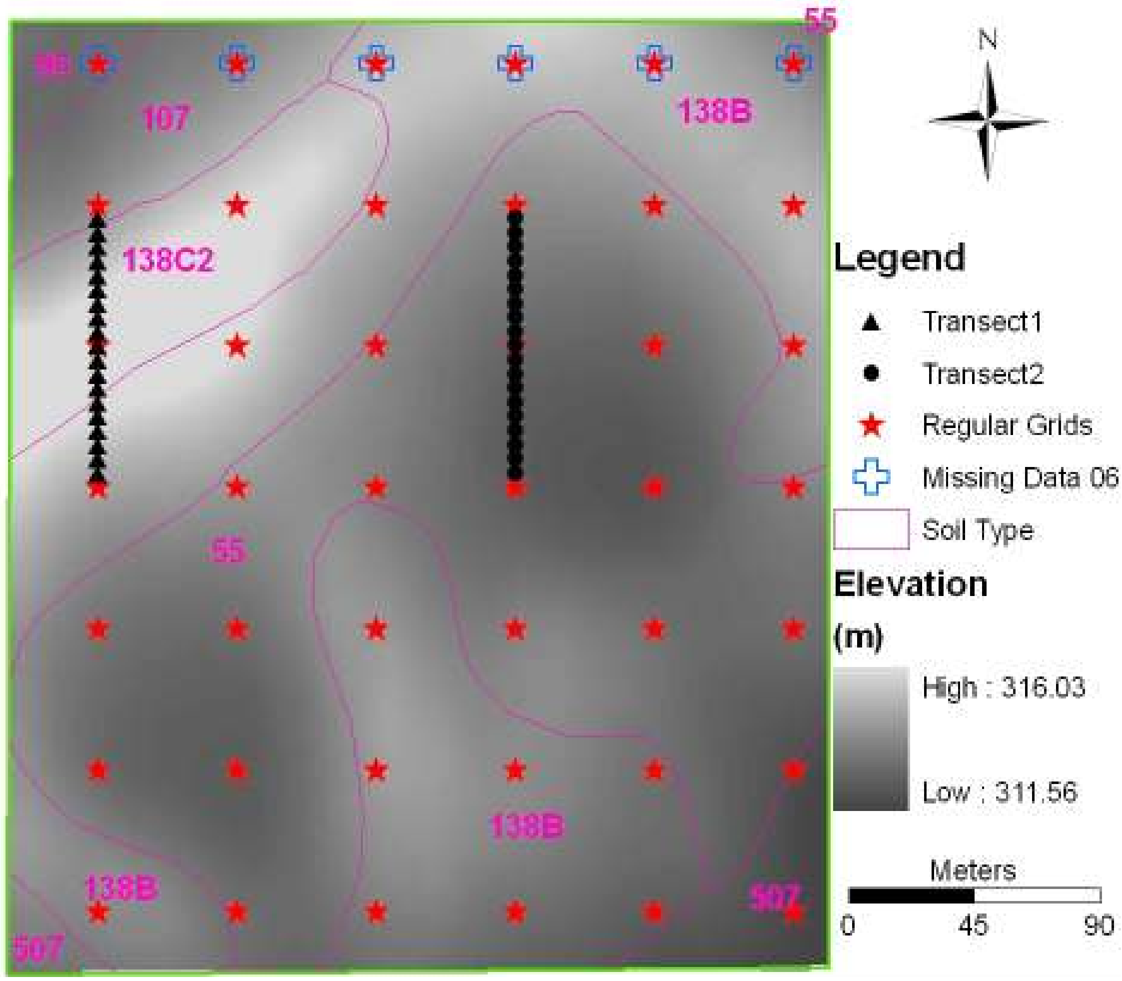

Figure 2.1 Soil Moisture Sampling Locations at Brooks Field.

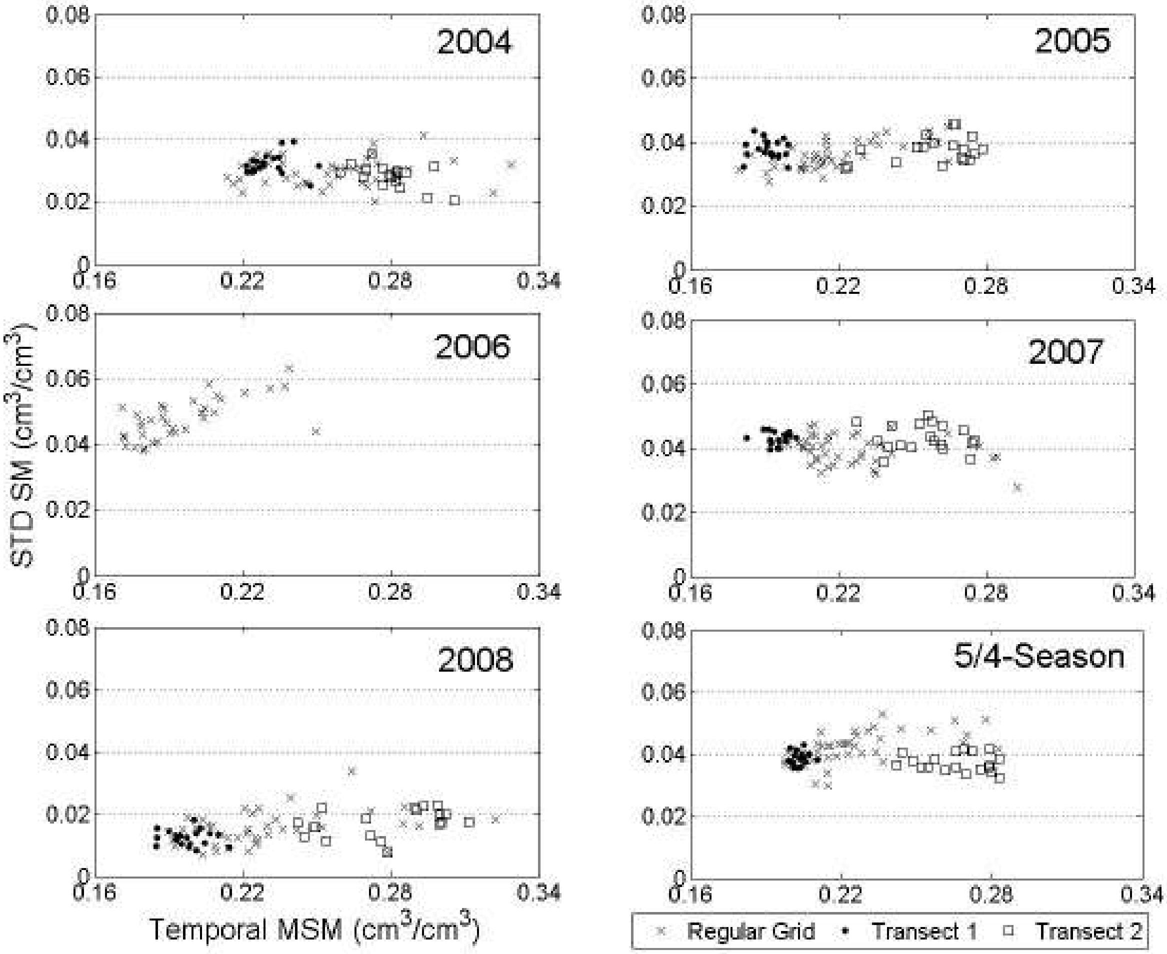

Figure 2.2 Standard deviation (STD) of soil moisture (SM) in time vs. temporal mean soil moisture (MSM) for each sampled location in every growing season and compiled 5 seasons (Note: 2006 had 6 missing sampled locations)

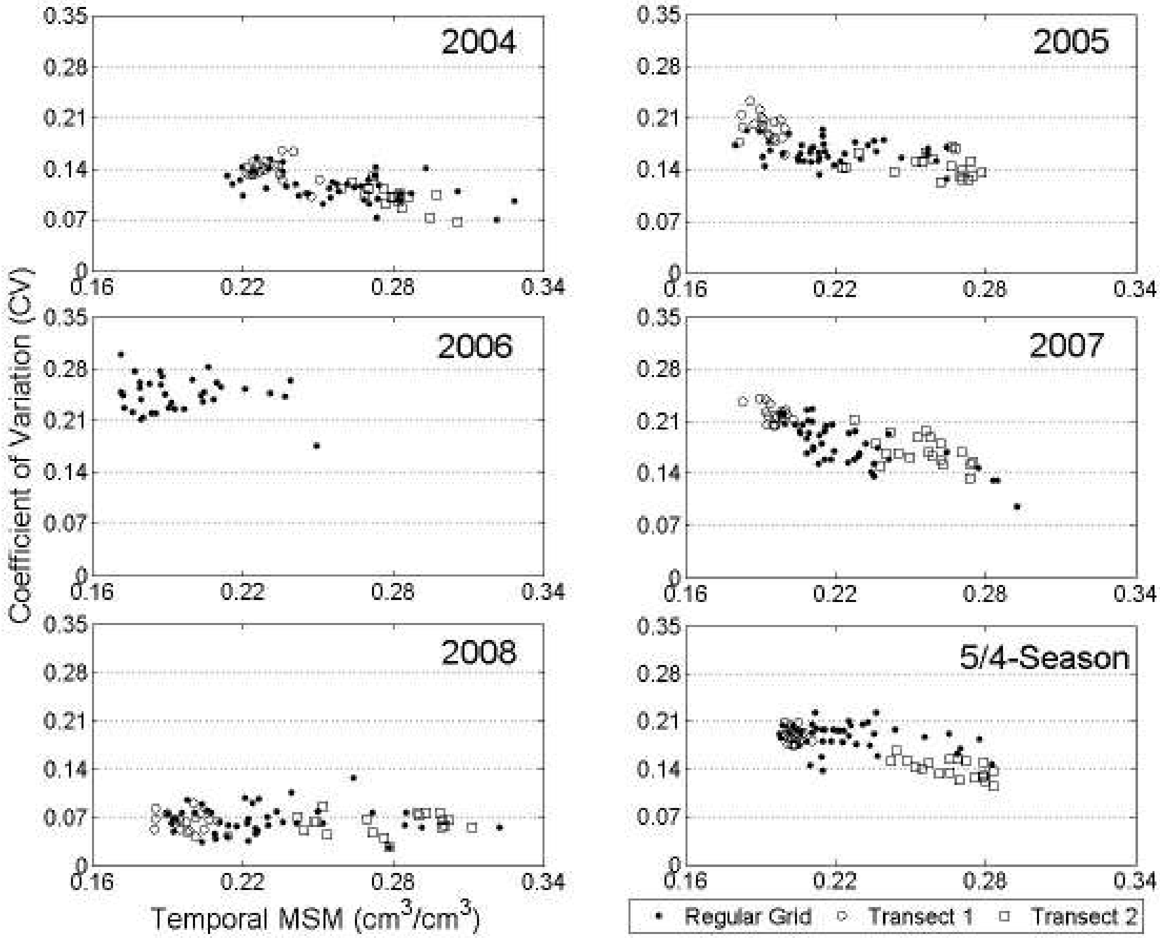

Figure 2.3 Coefficient of Variation (CV) of soil moisture (SM) in time vs. temporal mean soil moisture (MSM) for each sampled location in every growing season and compiled 5 seasons (Note: 2006 had 6 missing sampled locations)

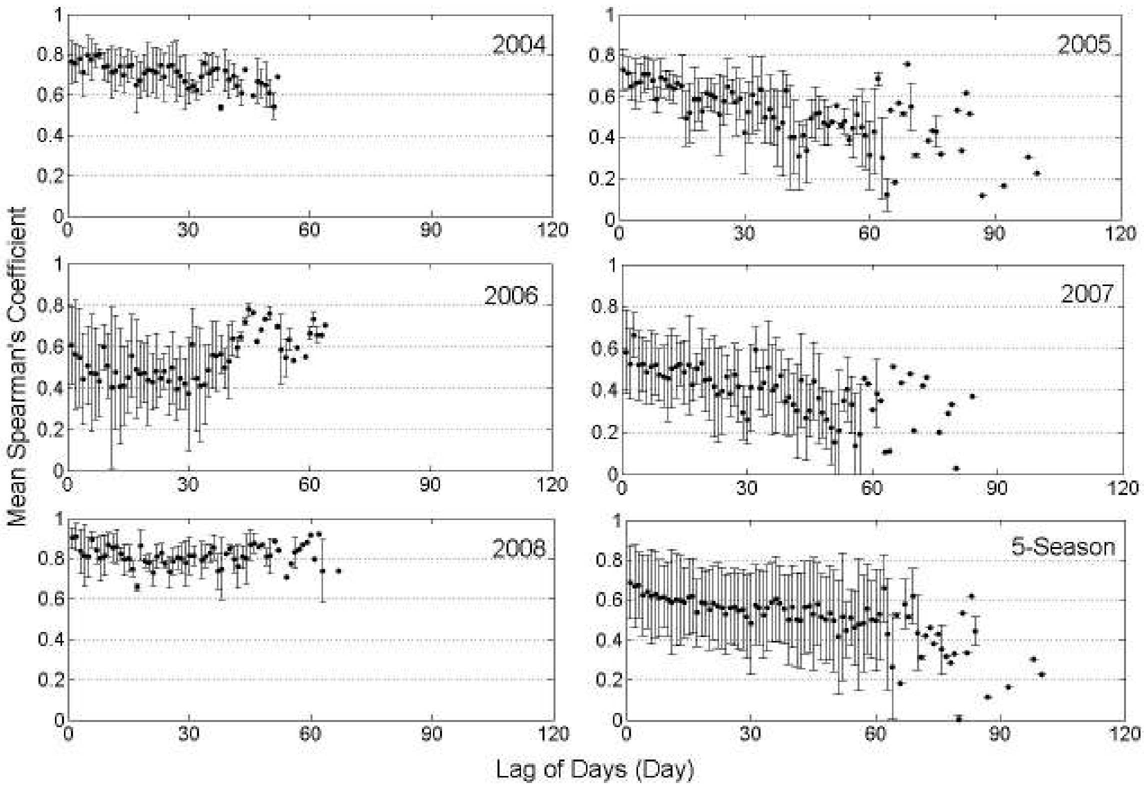

Figure 2.4 Spearman's Rank Correlation Coefficient vs. time lags in every growing season and compiled 5 seasons (Note: 2006 had 6 missing sampled locations)

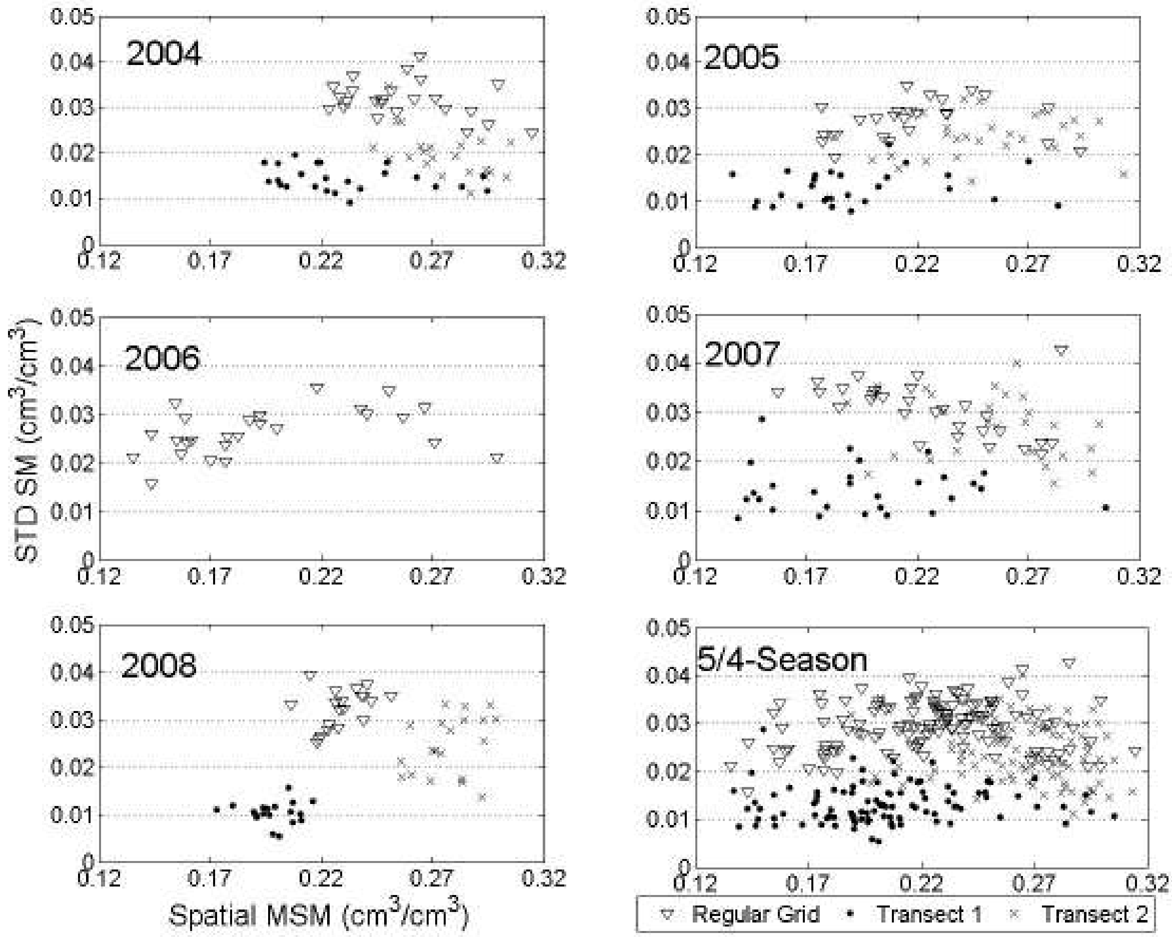

Figure 2.5 Standard deviation (STD) of soil moisture (SM) in space vs. spatial mean soil moisture (MSM) for each sampling day in every growing season and 5-season compiled.

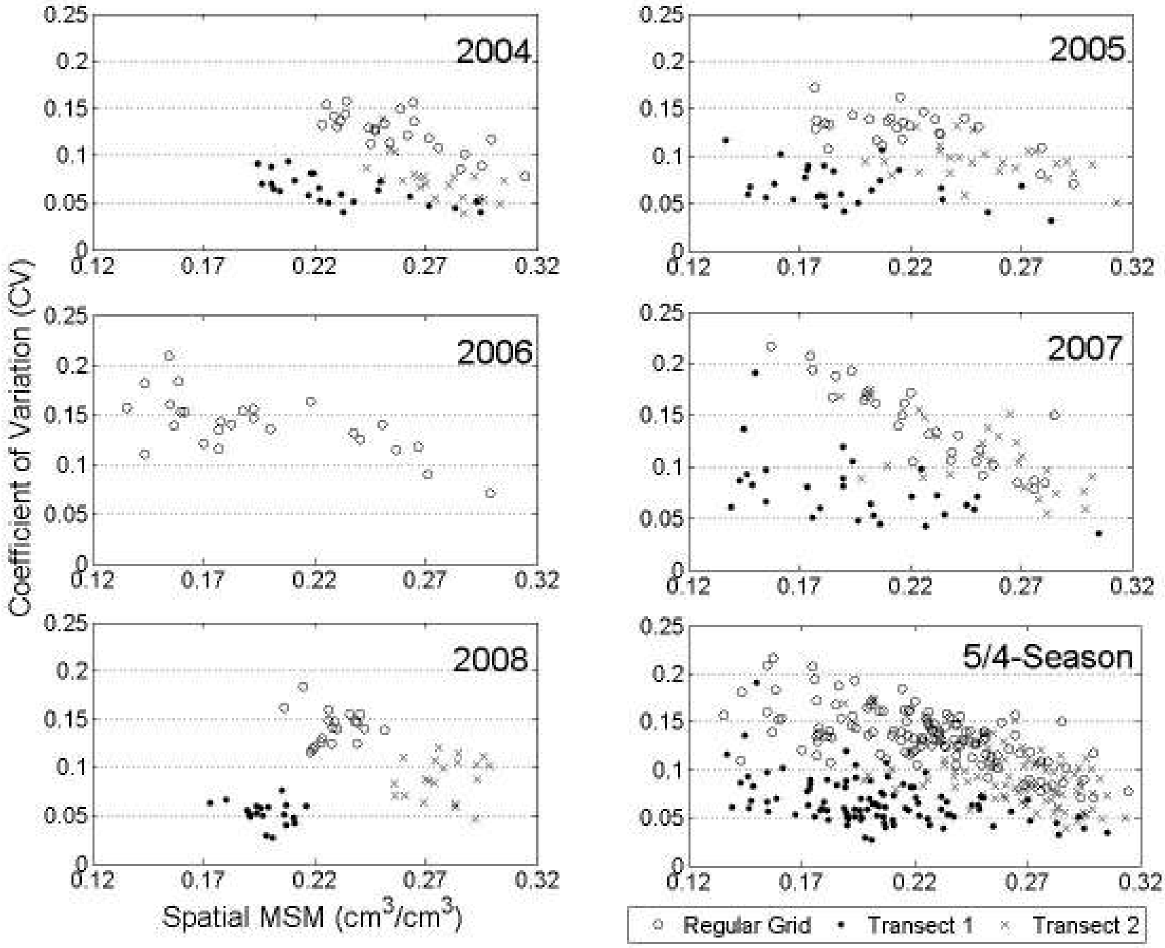

Figure 2.6 Coefficient of Variation (CV) of soil moisture (SM) in space vs. spatial mean soil moisture (MSM) for each sampling day in every growing season.

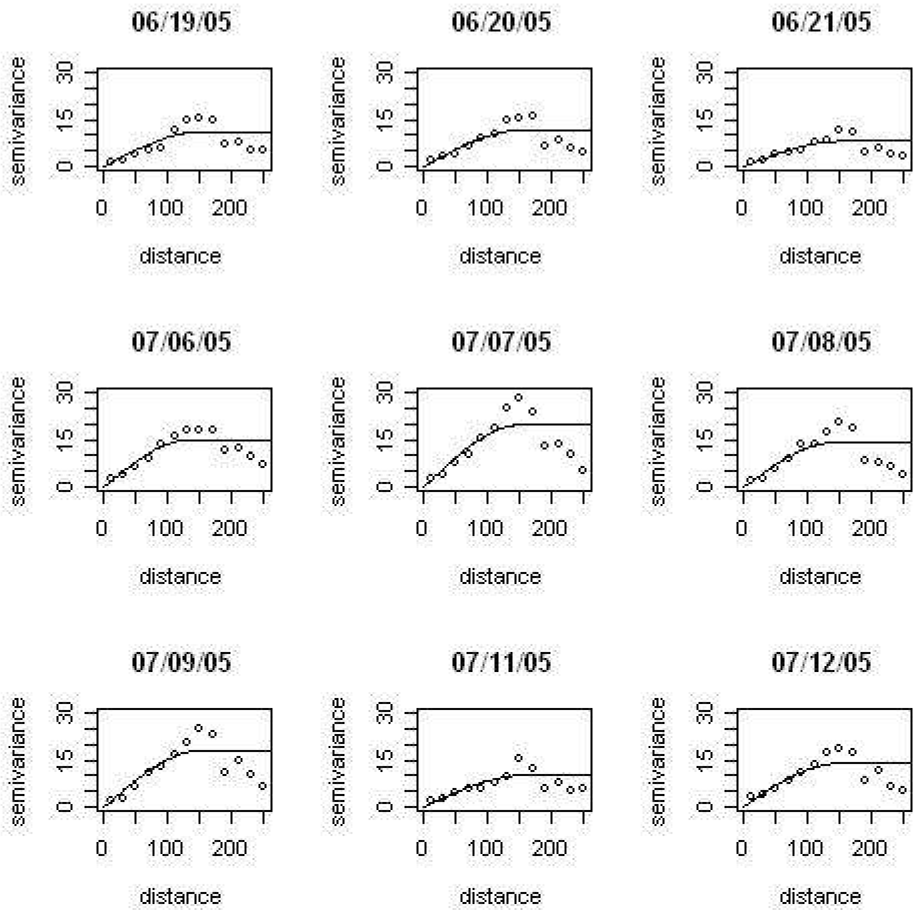

Figure 2.7 Computed semivariograms in Brooks Field

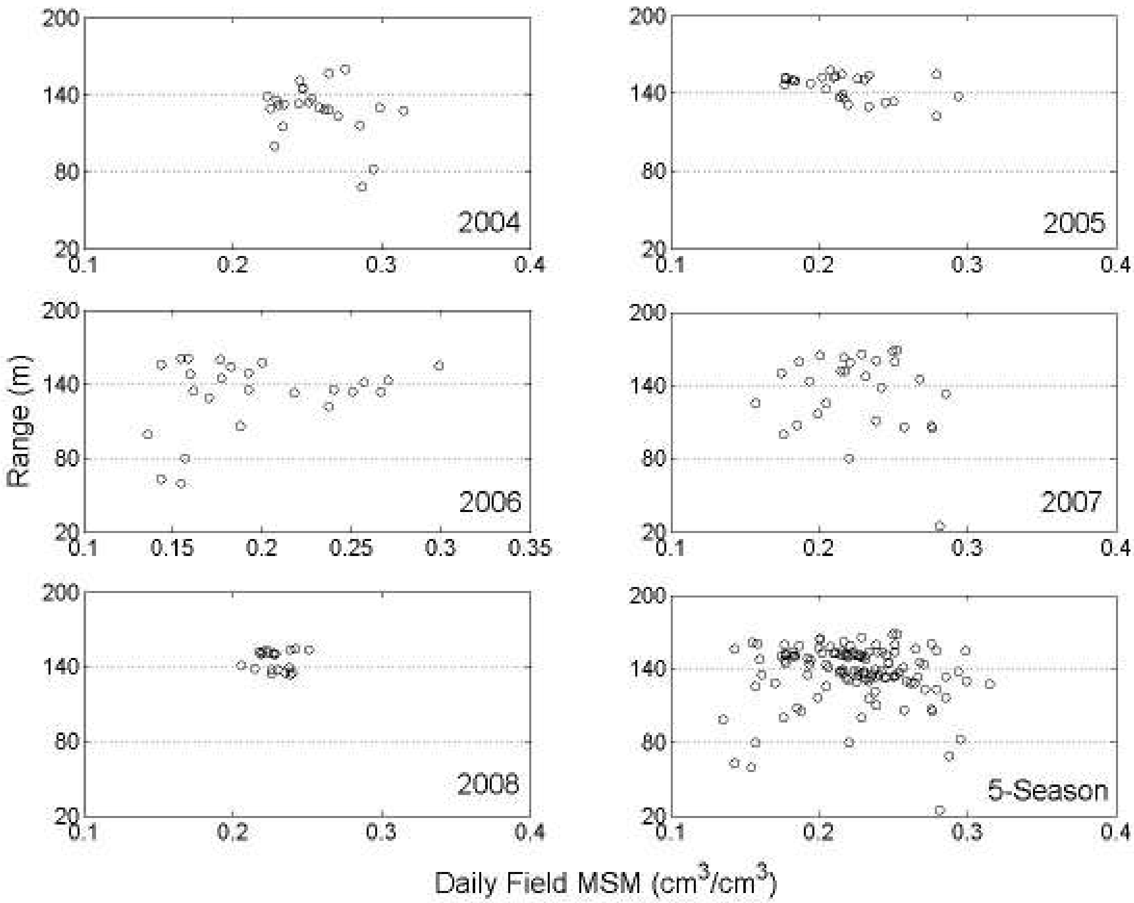

Figure 2.8 Ranges from spherical semivariograms vs. spatial mean soil moisture (MSM) for each sampling day in every growing season.

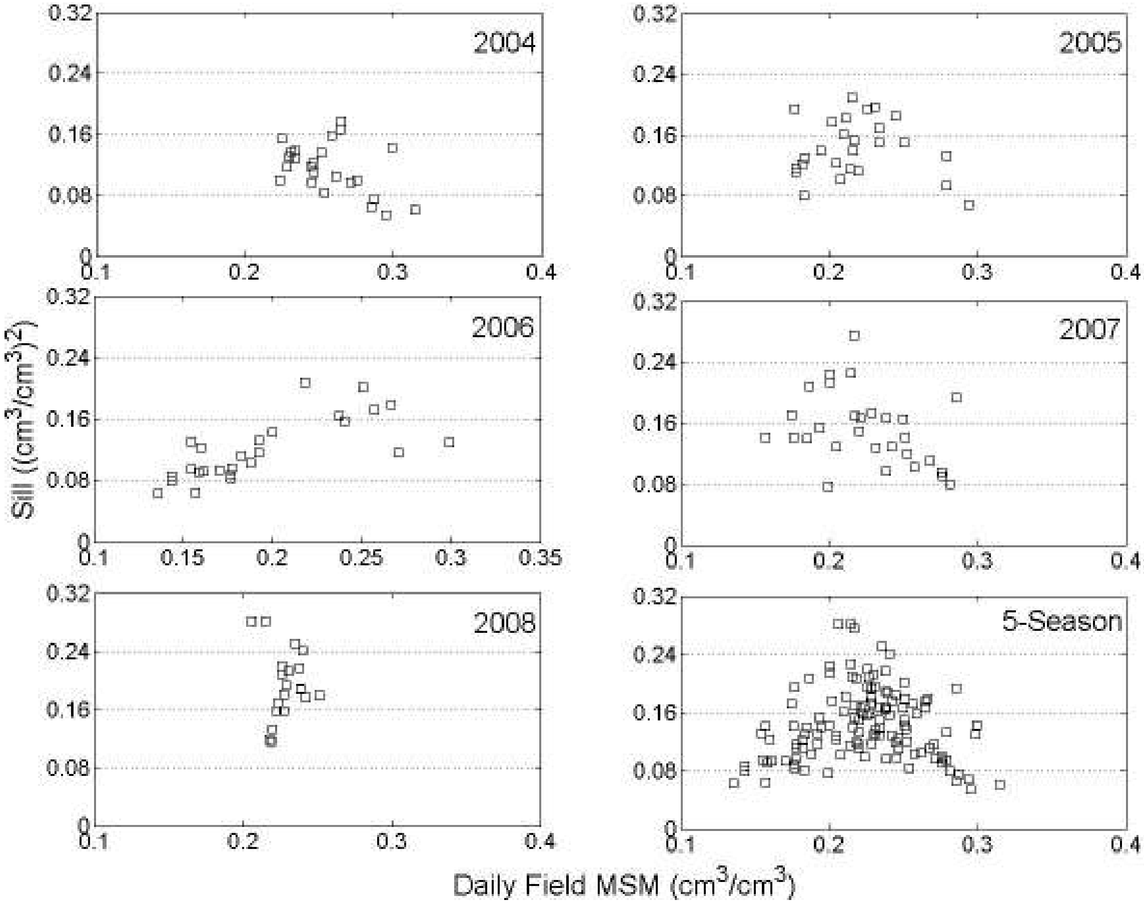

Figure 2.9 Sills from spherical semivariograms vs. spatial mean soil moisture (MSM) for each sampling day in every growing season.

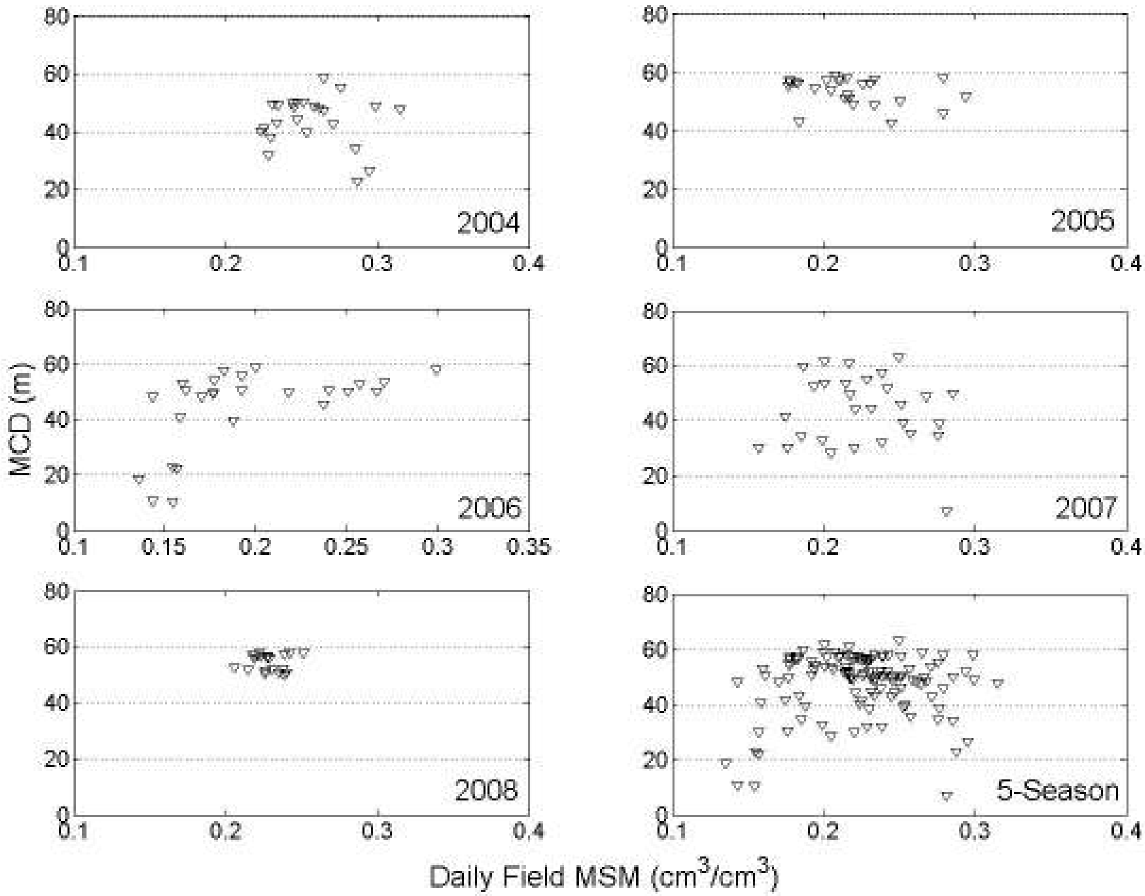

Figure 2.10 Mean Correlation Distances (MCD) from spherical semivariograms vs. spatial mean soil moisture (MSM) for each sampling day in every growing season.

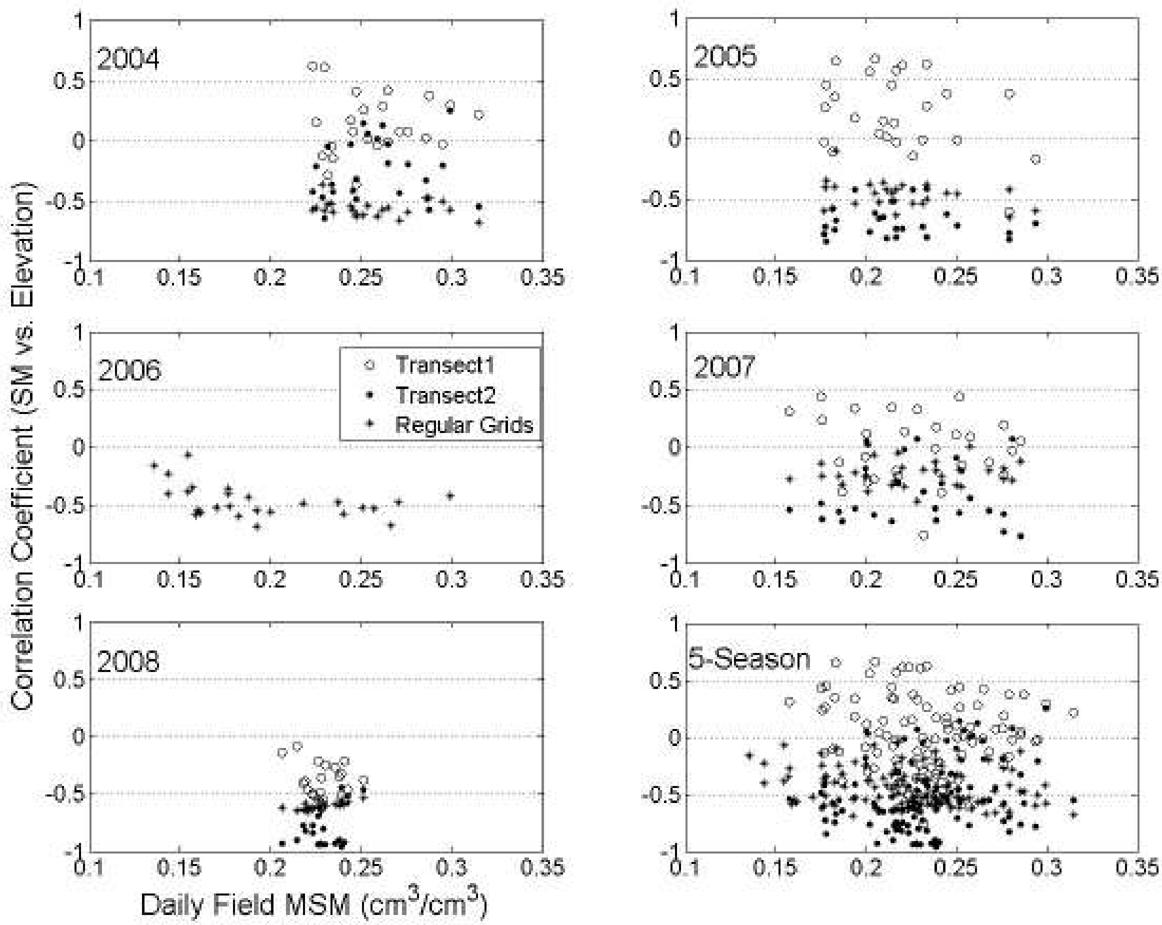

Figure 2.11 Correlation Coefficient between elevation and soil moisture for regular grids and transects.

Table 2.1 Description of soil types in Brooks

Soil Type	ID	Characteristics	Taxonomic Class	Typical Pedon
Nicollet loam	55	Very deep, somewhat poorly drained soils that formed in calcareous loamy glacial till on till plains and moraines	Fine-loamy, mixed, superactive, mesic Aquic Hapludolls	On a 1-3 percent plane slope in a cultivated field
Harps loam	95	Very deep, poorly drained soils formed in till or alluvium derived from till. Harps soils are on narrow rims or shorelines of depressions on till plains and moraines	Fine-loamy, mixed, superactive, mesic Typic Calciaquolls	A nearly level rim of a depression, in a cultivated field
Webster clay loam	107	very deep, poorly drained, moderately permeable soils formed in glacial till or local alluvium derived from till on uplands	Fine-loamy, mixed, superactive, mesic Typic Endoaquolls	On a concave slope of about 0 to 2 percent gradient in a cultivated field.
Clarion loam	138B	Very deep, moderately well drained soils on uplands. These soils formed in glacial till	Fine-loamy, mixed, superactive, mesic Typic Hapludolls	On a convex upland with a slope of 2-5 percent, in a cultivated field
	138C2			On a convex upland with a slope of 5-9 percent, in a cultivated field
Canisteo clay loam	507	Very deep, poorly and very poorly drained soils that formed in calcareous, loamy till or in a thin mantle of loamy or silty sediments and the underlying calcareous, loamy till. These soils are on rims of depressions, depressions and flats on moraines or till plains.	Fine-loamy, mixed, superactive, calcareous, mesic Typic Endoaquolls	Nearly level to slightly convex slope (0 to 2 percent), on a ground moraine, in a cultivated field.

Table 2.2 Summary of the total precipitation (Precip.) during soil moisture monitoring periods

Year	**Crop Type**	**Total Precip. (mm)**
2004	Soybean	252.2
2005	Corn	311.4
2006	Soybean	102.6
2007	Corn	239.3
2008	Soybean	278.6

CHAPTER 3. OPTIMAL SAMPLING LOCATIONS OF SOIL MOISTURE FOR VALIDATING REMOTE SENSING AT FIELD SCALE ACROSS MULTIPLE SEASONS

A paper submitted to Journal of Hydrology

Lingyuan Yang and Amy Kaleita

3.1. Abstract

Remote sensing has been used to monitor soil moisture at various scales. One of the challenges is to validate remotely sensed data with limited ground-based measurements. This study aimed to identify the locations on the ground to represent the mean soil moisture at the field scale across five different growing seasons from 2004 to 2008. The rank stability analysis had been used for that purpose. Optimal sampling locations (OSLs), giving accurate estimates of field mean soil moisture for each season, were selected using the rank stability analysis. The results indicated that there were OSLs in the field for each growing season and for the compiled five-season, and these locations were different from each sampling season to the next. The spatial patterns of soil moisture exhibited certain consistence across multiple seasons. The OSLs all tended to be located at those locations with higher elevation. These results suggest that it is not sufficient to use only one year or a few years' data to identify the soil moisture rank stable behavior using rank stability analysis, and that random sampling is possibly as good as targeted sampling in validating remotely sensed soil moisture data.

Key Words: Rank stability, Spatial variability, Temporal variability, Iowa

3.2. Introduction

The ability to monitor surface soil moisture over extended time periods and areas would provide valuable information for numerous applications in hydrology (*Jackson et al.,* 1996). Remote sensing has been used to capture the temporal and spatial characteristics of surface soil moisture at various scales [*Mohanty et al.*, 2001; *Bosch et al.*, 2006; *Cosh et al.*, 2006; *Choi et al.*, 2008]. However, there are limitations to the remote sensing approach. Among those limitations, remote sensors do not measure soil moisture directly, and the remote sensing signals have to be converted to soil water content values using retrieval algorithms. These algorithms need to be validated with *in situ* soil moisture measurements. At the same time, obtaining accurate and representative ground-based measurements can be time- and resource-consuming. Moreover, remote sensing instruments give spatially averaged measurements of soil moisture in a large area, known as footprint, while *in situ* measurements are point observations [*Robinson et al.*, 2008]. Thus, there is mismatch in scales between remote sensing footprints and ground samples [*Western and Blöschl*, 1999]. It is not practical to do exhaustive ground sampling. Guidance or prior knowledge is critical for choosing a limited number of optimal sites on the ground to use to accurately scale up the point-basis soil water content to the areal average soil moisture across a remote sensing footprint.

Vachaud et al. [1985] found that some locations consistently represented the mean and extreme values respectively of field water content, introducing the concept of temporal stability in soil moisture monitoring. Following the previous work, *Kachanoski and de Jong* [1988] defined temporal stability as the persistence of the spatial pattern of soil water storage in an area over time. *Chen* [2006] suggested that "rank stability" is a more precise term than "temporal stability," therefore, we will use the term rank stability in this paper.

If the rank stability exists in one field, reliable estimates of areal mean soil moisture in complex terrain could be obtained from a single or a limited number of sampling locations, which makes it possible to accurately scale up just a few point-scale soil moisture measurements to the areal mean. This could be used to overcome some of the challenges in validating remotely sensed soil moisture by reducing the number of *in situ* measurements needed to characterize a footprint's soil moisture content. Rank stable locations have been successfully used to estimate the areal mean soil moisture for validating corresponding remote sensing data [*Cosh et al.*, 2004; *Cosh et al.*, 2006; *Vivoni et al.*, 2008].

Most studies of soil moisture spatial or temporal patterns have been done on watershed or hillslope scales [*Famiglietti et al.,* 1998; *Grayson el al.,* 1998; *Cosh et al.,* 2004; *Jacobs et al.,* 2004; *Western et al.,* 2004; *Cosh et al.,* 2006; *Choi et al.,* 2008; *Penna et al.,* 2009]. There is a gap in our ability to routinely measure soil moisture at intermediate scales (subwatershed or catchment) [*Robison et al.*, 2008]. *Zielinski* [2002] has defined approximate areas of subwatershed (1-80 km^2) and catchment (0.1-1 km^2). The field scale, as we will use it in this study, is within these two categories. *Kaleita et al.* [2007] also suggested that additional studies on the field scales are still needed.

The development of global observing systems has become a common goal for hydrologists across the world. As part of this effort, soil moisture remote sensing missions have been launched or are in development, such as SMOS (Soil Moisture and Ocean Salinity) from the European Space Agency and SMAP (Soil Moisture Active and Passive) from the United States' NASA. These missions have called for a long-term *in situ* soil moisture network on the ground. Rank stable locations will provide an important basis for in situ soil moisture network deployment and operation with regards to remote sensing validation [*Cosh and Jackson, 2009*].

To make the rank stability analysis valuable for long-term soil moisture monitoring, more studies are needed to determine if the rank stable locations are consistent across the multiple years. Only a few studies, such as *Martinez-Fernandez and Ceballos* [2003] and *Schneider et al.* [2008], have analyzed rank stability over multiple seasons or years. These studies have suggested that the rank stable locations might lose their temporal rank persistence from year to year. *Schneider et al.* [2008] remain open to debate whether more accurate rank-stable locations can be obtained with a time series longer than two years. Therefore, further studies are needed to confirm the applicability of the rank stability concept for study periods longer than two years.

In order to understand more about the usage of rank stability concept and its potential value in the process of validating remotely sensed soil moisture, the objectives of this study were: 1) to determine if the rank stability of soil moisture exists at the field scale within each growing season and for five consecutive growing seasons together, in other words, if the optimal sampling locations (OSL), which give the most information with least amount of work, exist; and 2) to compare rank stable analysis to other sampling strategies. This study differs from the previous studies in that it focuses on the field scale and multiple growing seasons.

3.3. Materials and Methods

3.3.1. Study Area and Data

Distributed near-surface soil moisture data was collected across an agricultural field, Brooks Field, southwest of Ames, Iowa, during the growing seasons in 2004-2008. Brooks is a 10-ha corn-soybean rotation field with moderate topographic variation.

There were a total of 42 sampling locations in the entire field for 2004, 2005, 2007, and 2008 (Figure 3.1). These sampling locations were on a regular 50 meter grid. Six of those locations labeled differently in Figure 3.1, were not sampled in 2006. The sampling time window during the day was consistent within each season and slightly different across the seasons, but in all cases was limited to a maximum of two hours in order to minimize soil moisture differences due to drying.

At each sampling location, three readings for 0-6 cm depth were taken using a Theta probe moisture meter (Delta-T Devices, Cambridge UK, marketed in the United States by Dynamax, Inc., Houston, Tex.). The Theta probe output voltage values were converted to percentage volumetric soil moisture using the calibrated relationship for Des Moines Lobe soils provided by *Kaleita et al.* [2005]; the field component of that study was also conducted at the Brooks field. The average value of the three soil moisture contents at each location was used as the representative soil moisture at that location on that date for the analysis in this study.

In each season data collection was begun only after planting (usually May); data collection in 2008 was delayed because heavy spring rains delayed planting. In every growing season, there were missing data points on some sampling days due to the presence of standing water on the site; missing data was generally only in the two “potholes” in the field, where water tended to pond. For the days when data was missing for more than one sampling location, the sampling day was not included in further analysis for the sake of consistency. In this study, one day was excluded for this reason in 2004, four days in 2005, and four days in 2006. This may bias the results slightly, since generally the days excluded from analysis were very wet. When these data are excluded, there remain a total of twenty four days' data (from mid-May to early July) in 2004, twenty six days' data (from early-May to Mid-August) in 2005, twenty six days' data (from mid-May to mid-July) of data in 2006, twenty nine days' data (late May to Mid-August) in 2007, and twenty days' data in 2008 (late June to late August) available for analysis. On average, soil moisture data were collected twice a week in 2005, 2007 and 2008, and three times a week in 2004 and 2006. Data collection overall captured days with wet and with dry soil moisture conditions, as well as the dry out process.

According to NRCS web soil survey (http://websoilsurvey.nrcs.usda.gov/app/), Brooks has five soil types (Figure 3.1): Nicollet loam, Harps loam, Webster clay loam, Clarion loam, and Canisteo clay loam. The description of characteristics of each soil type is included in Table 3.1.

Elevation data was obtained by a LiDAR (**Li**ght **D**etection **a**nd **R**anging) survey at a 2-m horizontal resolution at Brooks Field (Figure 3.1). LiDAR is an optical remote sensing technology that measures properties of scattered light to find range and/or other information of a distant target. The prevalent method to determine distance to an object or surface is to use laser pulses. Like the similar radar technology, which uses radio waves instead of light, the range to an object is determined by measuring the time delay between transmission of a pulse and detection of the reflected signal. LiDAR has been used in particular as transformative method of obtaining detailed topographic information. It is attractive because it can see through many vegetation types, while obtaining information on vegetation height and density (Robinson et al., 2008).

Precipitation data was obtained from the Ames Station (UTM (WGS84): 433776.8E, 4653416.7N) at the Iowa Environmental Mesonet (IEM) observing network (http://mesonet.agron.iastate.edu/). This station is about 10 km away from Brooks Field.

3.3.2. Methodology

Rank stability analysis, provided by *Vachaud et al.* [1985], was used to find the optimal sampling locations (OSL) at Brooks field in the monitoring period. The mean relative difference $\bar{\delta}_i$ and its standard deviation $\sigma(\bar{\delta}_i)$ were calculated for each season individually, and for all seasons together. These parameters are defined by equation 1 through 3:

$$\delta_{ij} = \frac{\theta_{ij} - \bar{\theta}_j}{\bar{\theta}_j} * 100 \qquad (1)$$

$$\bar{\delta}_i = \frac{1}{m} \sum_{j=1}^{m} \delta_{ij} \qquad (2)$$

$$\sigma(\bar{\delta}_i) = \left[\sum_{j=1}^{m} \frac{\delta_{ij} - \bar{\delta}_i}{m-1} \right]^{1/2} \qquad (3)$$

where θ_{ij} is the moisture at the ith location on the jth sampling occasion, $\bar{\theta}_j$ is the field average of all θ_{ij} on the jth sampling occasion, and *m* is the total number of sampling occasions. Locations with mean relative difference $\bar{\delta}_i$ close to zero indicate sites having soil moisture close to field mean. Locations with negative mean relative difference underestimated the field average soil moisture, meaning those locations are relatively drier.

Locations with positive ones overestimated the field average, meaning those locations are relatively wetter. Locations with small $\sigma(\bar{\delta}_i)$ are considered to be temporally rank stable, because they have consistent behavior over time with respect to the field average soil moisture content. If such locations exist, using just several of them would provide a good alternative for sampling many points with acceptable accuracy to track field mean moisture conditions. .

The field mean soil moisture can be estimated using the following equation:

$$\bar{\theta}_{EST} = \frac{\theta_{OSL}}{1 + \bar{\delta}_{OSL}/100} \tag{4}$$

Where θ_{OSL} is the measured volumetric soil moisture content from an OSL on a given day, $\bar{\delta}_{OSL}$ is the mean relative difference from this OSL (determined from equation 2), and $\bar{\theta}_{EST}$ is the estimated field mean soil moisture from this OSL on the given day.

3.4. Results and Discussion

Time series of soil moisture (Figure 3.2) shows the temporal development of daily field mean volumetric soil moisture and the standard deviation as error bar, as well as the precipitation. The amount and distribution of precipitation, and the range of mean soil water content and the variability changed from season to season. In 2006, the investigation period had the least amount rain (102.6 mm), and it had the most amount rain in 2005 (311.4 mm) (Table 3.2). The field mean soil moisture had relatively greater variability in 2006 than the rest seasons, and 2008 had the least variability.

The temporal development of mean soil water content was highly influenced by the precipitation. The daily field average soil moisture responded to the precipitation as expected. Average soil moisture increased after a rain event, and dried out thereafter until the next event. We were able to capture the soil moisture changing processes using our sampling strategies.

If a region exhibits rank stable characteristics, selection of a few stable sampling points offers an efficient alternative to many points in order to estimate the areal mean soil moisture. Figure 3.3 shows the mean relative difference $\bar{\delta}_i$ values of all locations ranked in ascending order, with $\sigma(\bar{\delta}_i)$ as error bars, for each sampling season from 2004 to 2008 and the five compiled seasons. Generally, in Brooks field the wetter locations ($\bar{\delta}_i$>0) in each investigation time had greater variability than the drier locations ($\bar{\delta}_i$<0). *Jacobs et al.* [2004] observed a similar trend in SMEX02 (Soil Moisture EXperiment 2002) study area which was in the same watershed as Brooks Field. There have also been similar findings in other study areas [*Martínez-Fernández and Ceballos*, 2003]. However, it has also been found that there was no clear relationship between $\bar{\delta}_i$ and $\sigma(\bar{\delta}_i)$ in some other areas [*Schneider et al.*, 2008]

In this study, the four locations with the smallest standard deviation were selected as the OSLs for every sampling season and the compiled seasons; four locations were selected because this represents approximately 10% of the full data set, and also because it four samples strikes a balance between the time and resources invested and the value of information obtained. The OSLs are circled with location ID labeled in Figure 3.3.

The specific OSLs in each sampling season were different (Figure 3.3). Only a few points appeared as OSLs in more than one season, but no OSL appeared more than twice in all the years. The results are in agreement with the findings in the study of *Martinez-Fernandez and Ceballos* [2003], who found that $\sigma(\bar{\delta}_i)$ of rank stable locations changed from year to year. *Schneider et al.* [2008] discovered the similar results. In this study, the shift of these rank stable locations did not have any direct relationship with either corn-soybean rotation, precipitation amount, or the mean soil moisture conditions. These results suggest that it may not be sufficient to identify the OSLs from only one year's data or even a few years' data, like this case, for the purpose of using OSLs to estimate the field mean average soil moisture conditions across seasons.

Despite the year-to-year differences in exact locations, the OSLs exhibited some common attributes. 69% of OSLs were located on higher elevations (above the median elevation) (Figure 3.4), while 86% of the least rank stable locations, which had the largest standard deviation of mean relative difference, were located on elevations below the median (Figure 3.5). Though we did not have sampling locations in the middle of the closed depressions or potholes in the field, results of the rank stability analysis suggests that those locations are likely to be the least rank stable locations in the field. It might be because that these locations tend to have poorly drained soils, and they tend to be perpetually wet. When the field is wet in general, the soil moisture conditions at these locations are close to the field mean, while when the field is drying, these locations are often still wet, which makes their moisture conditions far from the mean conditions. Therefore, they have higher temporal variability relative to the field mean, making them the least rank stable locations.

This may provide useful instruction for future soil moisture sampling campaigns in this and similar areas. If representation of field average conditions with a minimum number of samples is an objective; campaigns in geomorphologically similar study regions should target locations with higher elevation, and avoid sampling in and around closed depressions.

Although the optimal sampling locations were not the same from season to season, the overall spatial patterns of soil moisture were somewhat consistent across multiple seasons. The mean relative difference of each sampling site for every season and the compiled seasons were highly correlated to each other, with all the correlation coefficients above 0.62 and more than half above 0.80 (Table 3.3).

In order to assess the impact of OSL selection on mean soil moisture estimation, the four OSLs (point 10, 20, 22, 39) in 2004 were used to estimate the field mean soil moisture content for 2004 and the other seasons according to equation 4. The results (Table 3.4) indicated that the estimated 2004 field average soil moisture from each OSL matched well to the measured data set, and more than 83.3% of them were within the +/-0.02 cm^3/cm^3 soil moisture error line and all of them fell into +/- 0.04 cm^3/cm^3 soil moisture error line. This is to be expected, as the 2004 data was used to select OSLs. The 2004-based estimates of field mean soil moisture for the subsequent years, however, were less robust (Table 3.4). *Schneider et al.* [2008] found in their study that the selected rank stable locations predict soil moisture within 0.05 cm^3/cm^3 error at each sampling date with just a few exceptions. By looking at the estimates of mean soil moisture for each season from the individual 2004-based OSL (Figure 3.6), it was found that the soil moisture were either overestimated or underestimated except 2004 (Table 3.5). The root mean square errors (RMSEs) of the estimated volumetric soil moisture from 2004-based OSLs ranged from 0.009 to 0.013 cm^3/cm^3 volumetric soil moisture for 2004, 0.017 to 0.041 cm^3/cm^3 for 2005, 0.017 to 0.025 cm^3/cm^3 for 2006, 0.017 to 0.054 cm^3/cm^3 for 2007, 0.007 to 0.019 cm^3/cm^3 for 2008, and

0.017 to 0.029 cm^3/cm^3 for 5-season compiled. The average estimated soil moisture from 4 OSLs gave a better prediction of field mean soil moisture with less RMSE than the individual OSL (Table 3.5) for every year.

For comparison, 5 sets of four random points were selected for each investigation period to estimate the field mean soil water contents. Every time, four random numbers were generated using a random number generator, and the locations with point ID matching these numbers were selected. In addition to those 5 sets of random points, four locations (point 15, 17, 20 and 38) were also selected using the general guidance resulting from the preceding rank stability analysis, that is, to focus on the areas with higher elevation (Figure 3.4). These guided locations were with the same as some of the rank stable locations from different growing seasons.

In any case, the RMSE was almost always smaller than 0.04cm^3/cm^3 (all but one) and frequently smaller than 0.02cm^3/cm^3 error if using the single rank stable location to predict field mean soil moisture (Table 3.5). If the average of 4 rank stable points were used, the error was almost always smaller than 0.02cm^3/cm^3 (Table 3.5). At the same time, RMSEs of the estimated mean soil water contents were always smaller than 0.01cm^3/cm^3 by averaging four random locations, and almost always smaller than 0.01cm^3/cm^3 (except one) by averaging four guided locations (Table 3.5). The average of both four guided and 5 sets of four randomly selected locations gave less error than that of the four 2004-based rank stable locations in predicting mean soil moisture (Table 3.5). Figure 3.7 shows the comparison of the errors of mean soil moisture prediction using three sampling strategies (four 2004-based rank stable locations, 5 sets of four random locations and four guided locations) for 125 days over the five seasons. Overall, the random sampling strategy was as good as guided sampling, and both were better than the rank stability method in Brooks Field. The 5 sets of four random points performed as well as four guided sampling locations in predicting mean soil moisture except the random strategy had more outliers. The four 2004-based rank stable locations had more overestimation than underestimation of mean soil moisture and had the most outliers of three methods, and the median of the error was about 3.4% higher than 0%. These results imply that the average effects of several rank stable locations might have less error than only one rank stable location in estimating field mean soil moisture if short-term analysis identifies rank stable locations. Furthermore, even if rank stable location does not exist in the study region, the average effects of several locations with random selection or with guidance can give acceptable estimates of field mean soil moisture.

For the application of validating remotely sensed soil moisture data by generating an average from a single or a limited number of locations determined to be stable on the basis of one year's analysis, this study indicates that from time to time the rank stable locations may not perform as well in estimating field mean soil moisture as they did within the period used to determine their location. In order to accurately account for error in the remotely sensed estimates, it is important to be aware of this potential source of validation error.

3.5. Conclusions

The application of the rank stability method [*Vachaud et al.*, 1985] to a Midwest agricultural field in central Iowa, for the 2004 to 2008 growing seasons indicated that rank stability existed in each season, but the rank stable locations or optimal sampling locations (OSLs) were not consistent from season to season. In spite of the inconsistency of the OSLs across multiple seasons, the spatial patterns were highly correlated (correlation coefficients were above 0.62) from season to season. Furthermore, most of the stable locations were located on higher elevations. The results suggested that it is not sufficient to use only one year or a few years' data to capture rank stable soil moisture behavior with rank stable locations, for the short-term remote sensing validation or any other similar application. However, random sampling is possibly as good as targeted sampling for validating remotely sensed soil moisture using ground-based measurements. Further studies are needed for the application and limitation of rank stability analysis.

3.6. Acknowledgements

This work was supported by NASA grant NNG06GC63G from the Land Surface Hydrology program, and by the Iowa State University Environmental Science Fellowship and the Iowa Space Grant Consortium Program for Space-borne and Earthbound System Sustainability.

3.7. References

Bosch, D.D., V. Lakshmi, T.J. Jackson, M. Choi, and J.M. Jacobs, 2006. Large scale measurements of soil moisture for validation of remotely sensed data: Georgia soil moisture experiment of 2003. *J. Hydrol.*, 323, 120-137.

Chen, Y., 2006. Letter to the Editor on "Rank Stability or Temporal Stability". *Soil Sci. Soc. Am. J.*, 70 (1), 306.

Choi, M., J.M. Jacobs, and M.H. Cosh, 2007. Scaled spatial variability of soil moisture fields. *Geophys. Res. Lett.,* 34, L01401, doi:10.1029/2006GL028247.

Choi, M., J.M. Jacobs, and D.D. Bosch, 2008. Remote sensing observatory validation of surface soil moisture using Advanced Microwave Scanning Radiometer E, Common Land Model, and ground based data: Case study in SMEX03 Little River Region, Georgia, U.S.. *Water Resour. Res.*, 44, W08421.

Cosh, M.H., T.J.Jackson, R. Bindlish, and J.H. Prueger, 2004. Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sens. Environ.*, 92, 427-435.

Cosh, M.H., T.J. Jackson, P. Starks, and G. Heathman, 2006. Temproal stability of surface soil moisture in the Little Washita River watershed and its appplicaions in satellite soil moisture product validation. *J. Hydrol.*, 323, 168-177.

Cosh, M.H. and T.J. Jackson, 2009. In situ validation of soil water content from satellite remote sensing. *ASA-CSSA-SSSA Annual Meeting*, Nov. 1-5.

Famiglietti, J.S., J.W. Rudnicki, and M. Rodell, 1998. Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas. *J. Hydrol.*, 210, 259-281.

Grayson, R.B., and A.W. Western, 1998. Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *J. Hydrol.*, 207, 68-82.

Jacobs, J. M., B. P. Mohanty, E. Hsu, and D. Miller, 2004. SMEX02: Field scale variability, time stability and similarity of soil moisture. *Remote Sens. Environ.*, 92, 436-446.

Kaleita, A. L., J. L. Heitman, S. D. Logsdon. 2005. Field calibration of the Theta Probe for Des Moines Lobe soils. *Applied Engineering in Agriculture*, 21(5), 865-870.

Kaleita, A. L., M. C. Hirschi, and L. F. Tian, 2007. Field-Scale Surface Soil Moisture Patterns and Their Relationship to Topographic Indices. *Transactions of the ASABE*, 50(2), 557-564.

Kachanoski, R.G., and E. de Jong. 1988. Scale dependence and the temporal persistence of spatial patterns of soil water storage. *Water Resour. Res.*, 24, 85–91.

Martínez-Fernández, J., and A. Ceballos. 2003. Temporal stability of soil moisture in a large-field experiment in Spain. *Soil Sci. Soc. Am. J.*, 67, 1647–1656.

Mohanty, B.P., and T.H. Skaggs, 2001. Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Adv Water Resour.*, 24, 1051-1067.

Penna, D., M. Borga, D. Norbiato, and G. D. Fontana, 2009. Hillslope scale soil moisture variability in a steep alpine terrain. *J. Hydrol.*, 364, 311-327.

Robinson, D. A., C. S. Campbell, J. W. Hopmans, B. K. Hornbuckle, S. B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth, 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7 (1), 358-389.

Schneider, K., J.A. Huisman, L. Breuer, Y. Zhao, and H.-G. Frede, 2008. Temporal stability of soil moisture in various semi-arid steppe ecosystems and its application in remote sensing. *J. Hydrol.*, 359 (1-2), 16-29.

Vachaud, G., A. Passerat de Silans, P. Balabanis, and M. Vauclin, 1985. Temporal stability of spatially measured soil water probability density function. *Soil Sci. Soc. Am. J.*, 49, 822-828.

Western, A.W., and G. Blöschl, 1999. On the spatial scaling of soil moisture. *J. Hydrol.,* 217, 203-224.

Western, A.W., S. Zhou, R.B. Grayson, T.A. McMohon, G. Blöschl, and D.J. Wilson, 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *J. Hydrol.,* 286, 113-134.

Vivoni, E.R., M. Gebremichael, C.J. Watts, R. Bindlish, and T.J. Jackson, 2008. Comparison of ground-based and remotely-sensed surface soil moisture estimates over complex terrain during SMEX04. *Remote Sens. Environ.*, 112, 314-325.

Zielinski, J., 2002. Watershed Vulnerability Analysis. http://www.cwp.org/Resource_Library/Center_Docs/USRM/Vulnerability_Analysis.pdf. *Center for Watershed Protection, Ellicot City, MD 21043.*

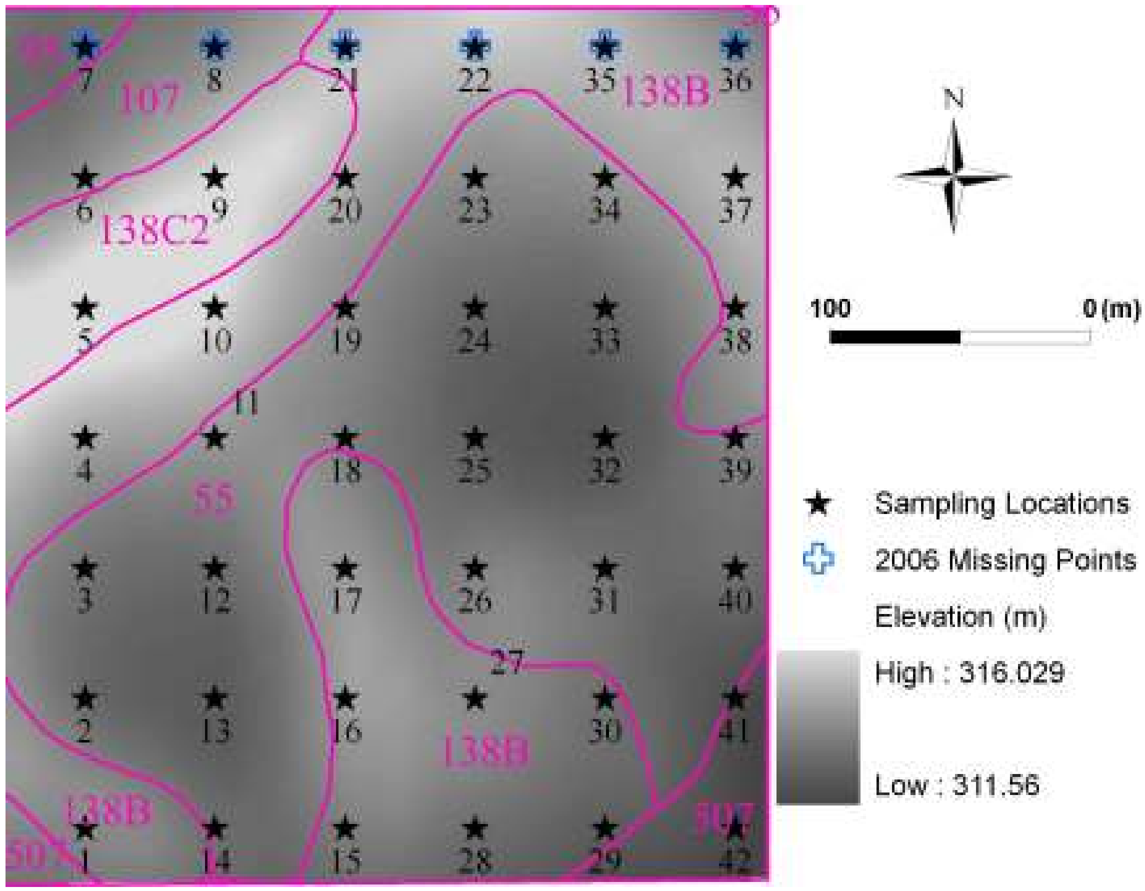

Figure 3.1 Soil Moisture Sampling Locations at Brooks Field.

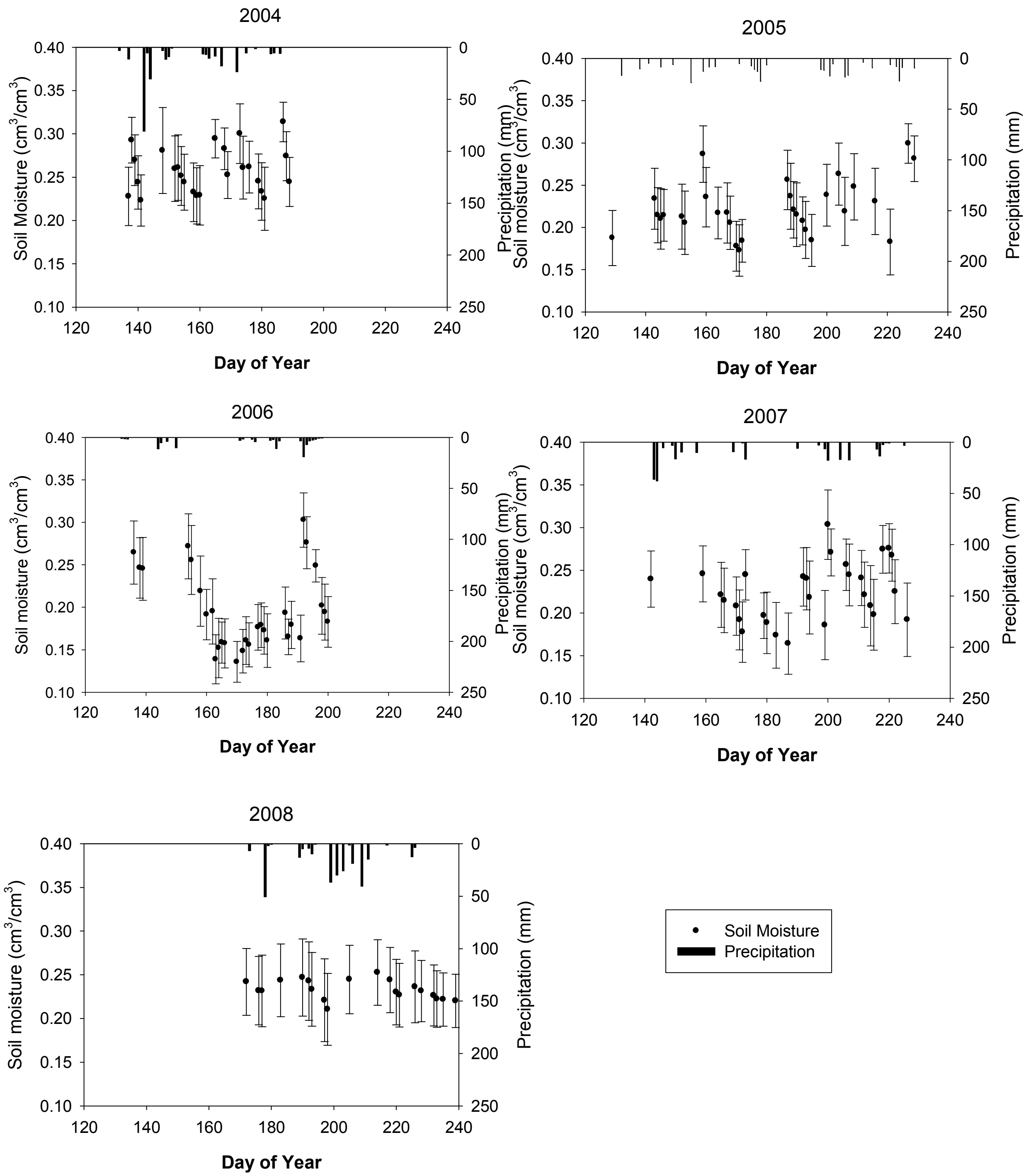

Figure 3.2 Mean and standard deviation of volumetric soil moisture content with precipitation by sampling seasons for Brooks field

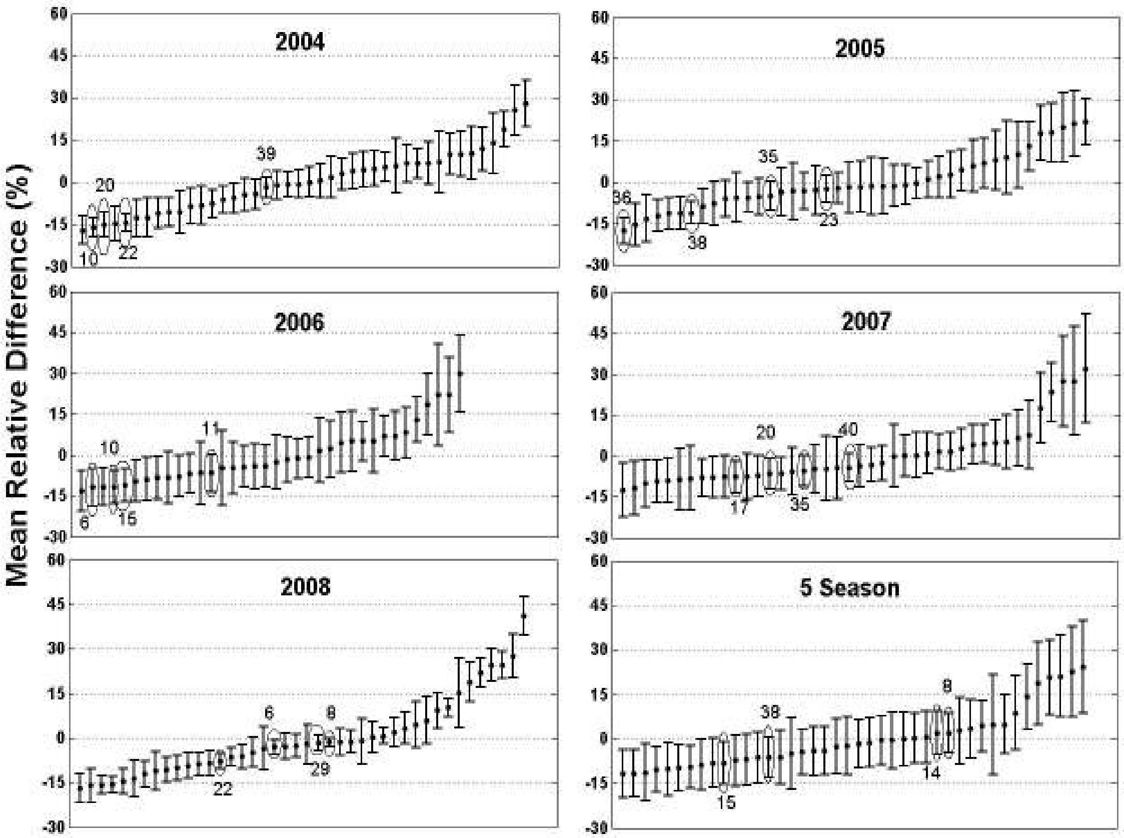

Figure 3.3 The mean relative difference in soil moisture at various locations within Brooks Field for 2004 to 2008, and five seasons combined. The black dots indicate the mean relative difference of soil moisture, and the error bars are the standard deviation of mean relative difference. The circled locations with point ID labeled represent the rank stable locations (or OSLs) in each data set.

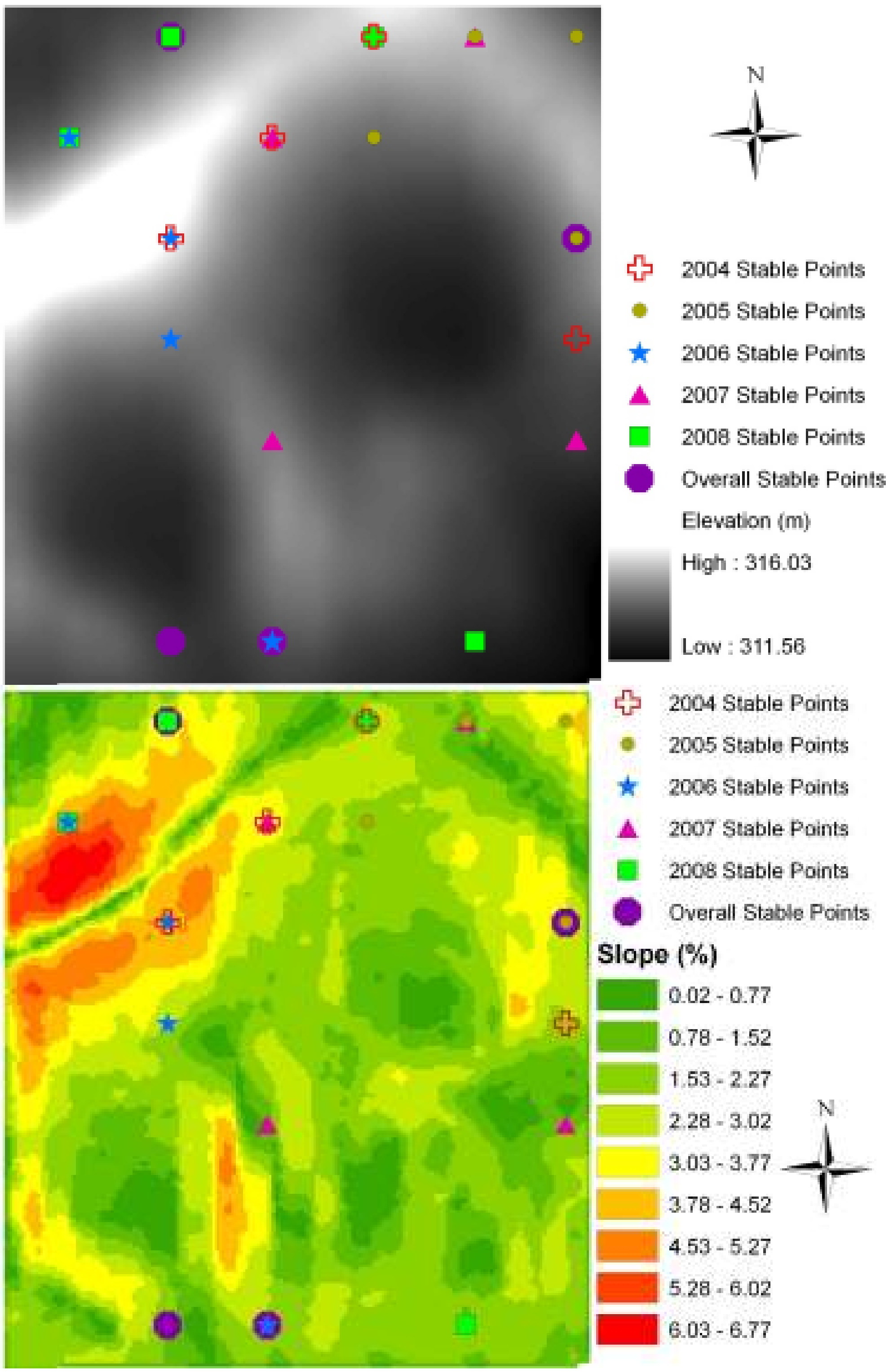

Figure 3.4 The stable locations with elevation and slope for each sampling season and compiled seasons

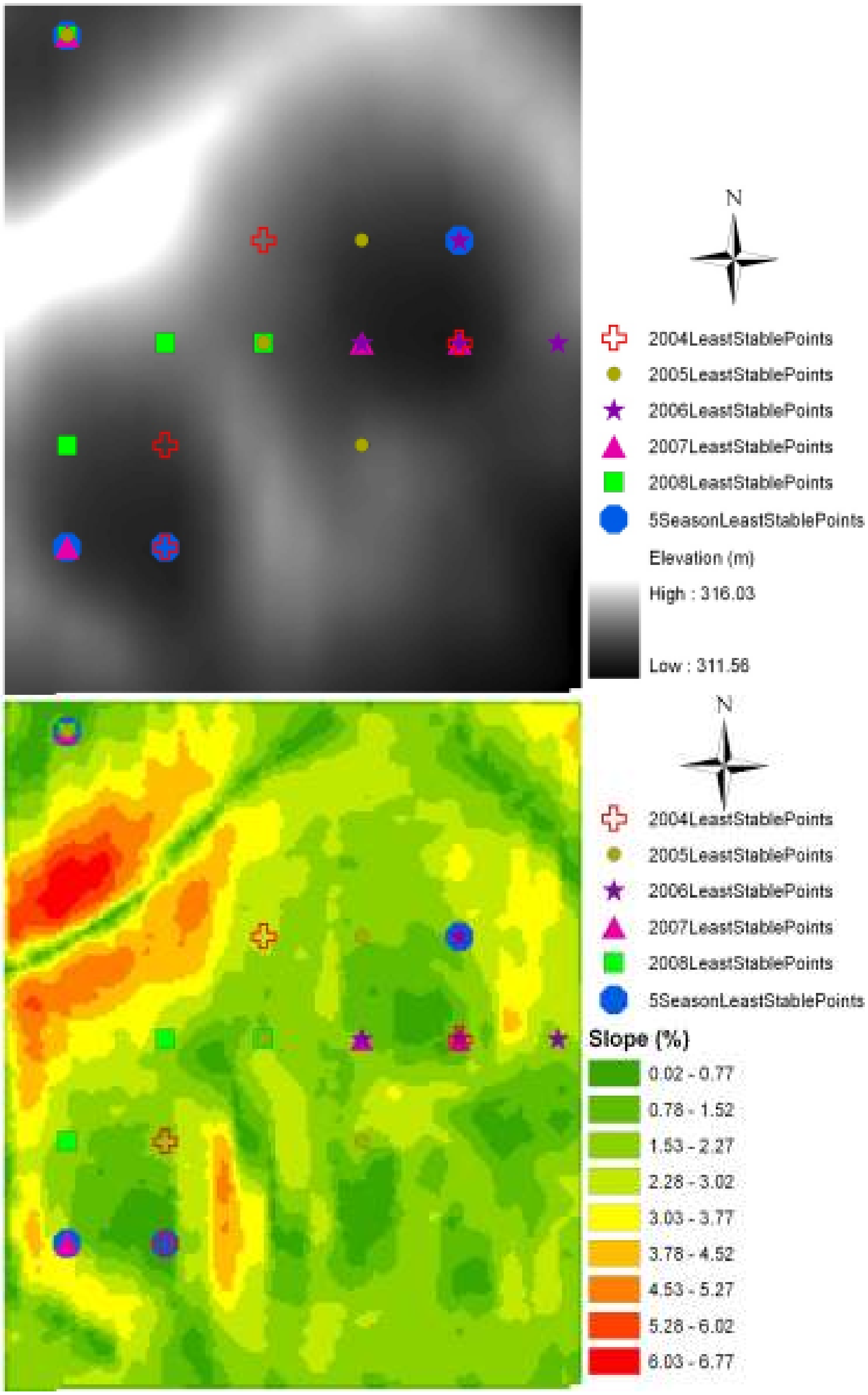

Figure 3.5 The least stable locations with elevation and slope for each sampling season and compiled seasons

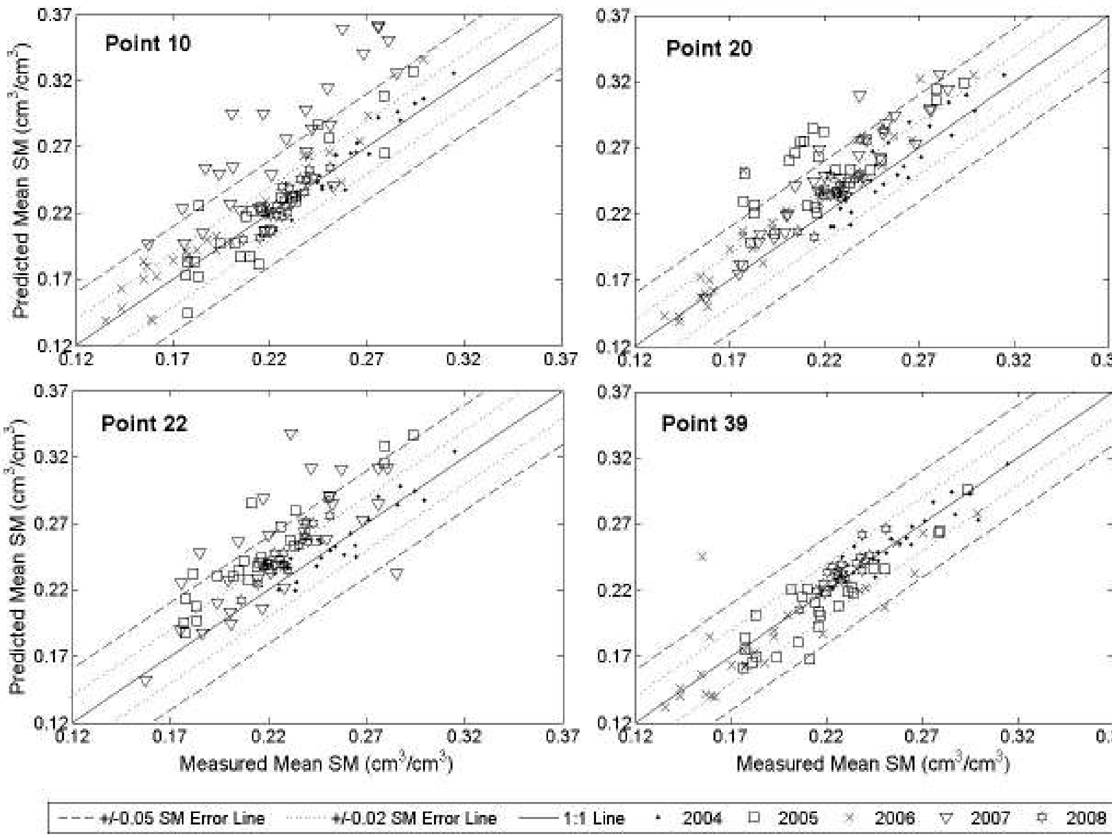

Figure 3.6 The estimated field mean soil moisture for each sampling season from 2004 to 2008 and using the OSLs determined with 2004 data.

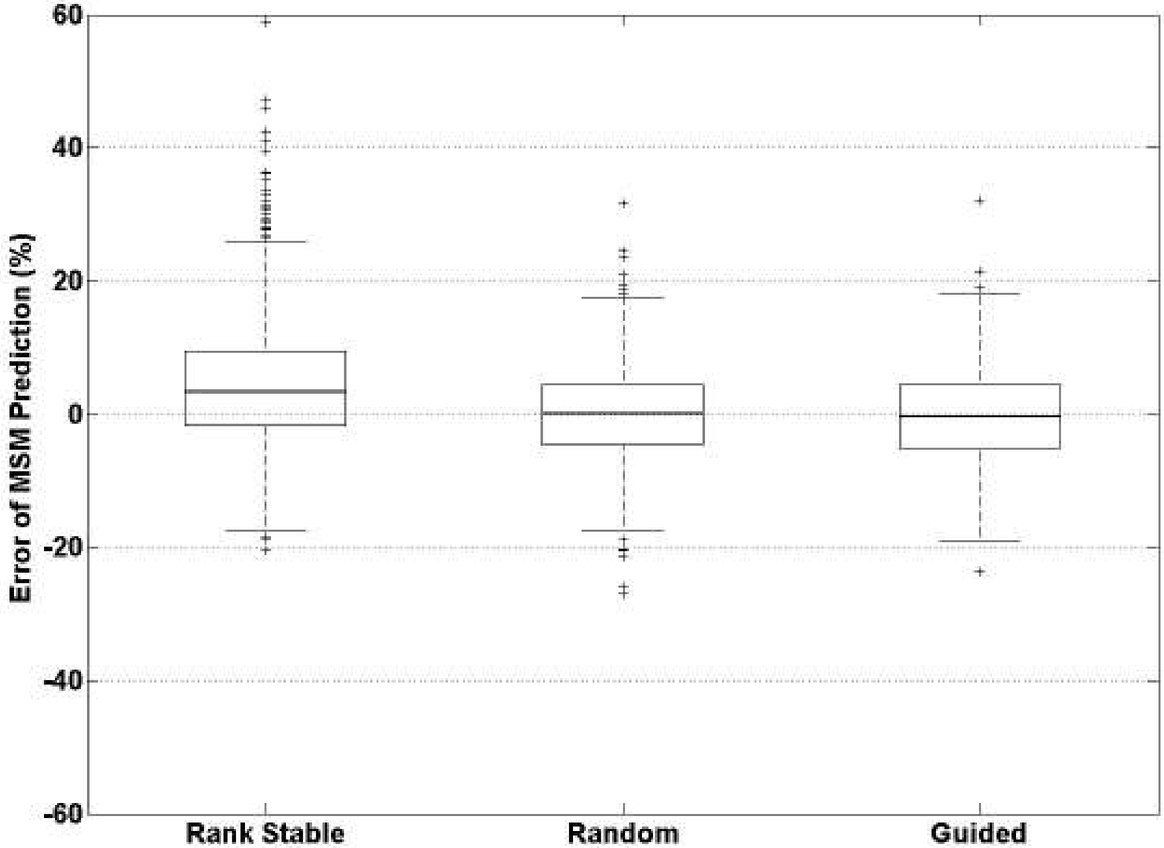

Figure 3.7 Comparison of the errors of field mean soil moisture (MSM) prediction for all 5 seasons by using three different sampling strategies: 2004-based rank stable locations, Random sampling locations, and Guided sampling locations).

Table 3.1 Description of soil types in Brooks

Soil Type	ID	Characteristics	Taxonomic Class	Typical Pedon
Nicollet loam	55	Very deep, somewhat poorly drained soils that formed in calcareous loamy glacial till on till plains and moraines	Fine-loamy, mixed, superactive, mesic Aquic Hapludolls	On a 1-3 percent plane slope in a cultivated field
Harps loam	95	Very deep, poorly drained soils formed in till or alluvium derived from till. Harps soils are on narrow rims or shorelines of depressions on till plains and moraines	Fine-loamy, mixed, superactive, mesic Typic Calciaquolls	A nearly level rim of a depression, in a cultivated field
Webster clay loam	107	very deep, poorly drained, moderately permeable soils formed in glacial till or local alluvium derived from till on uplands	Fine-loamy, mixed, superactive, mesic Typic Endoaquolls	On a concave slope of about 0 to 2 percent gradient in a cultivated field.
Clarion loam	138B	Very deep, moderately well drained soils on uplands. These soils formed in glacial till	Fine-loamy, mixed, superactive, mesic Typic Hapludolls	On a convex upland with a slope of 2-5 percent, in a cultivated field
	138C2			On a convex upland with a slope of 5-9 percent, in a cultivated field
Canisteo clay loam	507	Very deep, poorly and very poorly drained soils that formed in calcareous, loamy till or in a thin mantle of loamy or silty sediments and the underlying calcareous, loamy till. These soils are on rims of depressions, depressions and flats on moraines or till plains.	Fine-loamy, mixed, superactive, calcareous, mesic Typic Endoaquolls	Nearly level to slightly convex slope (0 to 2 percent), on a ground moraine, in a cultivated field.

Table 3.2 Summary of the total precipitation (Precip.) during soil moisture monitoring periods

Year	Crop Type	Total Precip. (mm)
2004	Soybean	252.2
2005	Corn	311.4
2006*	Soybean	102.6
2007	Corn	239.3
2008	Soybean	278.6

Table 3.3 Correlation of the mean relative difference of soil moisture from each year

	2004	**2005**	**2006***	**2007**	**2008**	**5 Seasons**
2004	1.00					
2005	0.76	1.00				
2006*	0.82	0.83	1.00			
2007	0.62	0.65	0.65	1.00		
2008	0.78	0.77	0.80	0.64	1.00	
5 Seasons	0.89	0.90	0.93	0.83	0.89	1.00

*: When calculating the correlation coefficients between 2006 and the other years, only the data points available in 2006 were used.

Table 3.4 Estimates of field mean soil moisture within +/-0.02 and +/-0.04cm^3/cm^3 tolerance using the 2004-based OSLs.

	% Estimates of mean soil moisture within +/- 0.02cm^3/cm^3 tolerance				**% Estimates of mean soil moisture within +/- 0.04cm^3/cm^3 tolerance**			
Year	**P10**	**P20**	**P22**	**P39**	**P10**	**P20**	**P22**	**P39**
2004	95.8	83.3	100.0	95.8	100.0	100.0	100.0	100.0
2005	69.2	38.5	23.1	84.6	92.3	61.5	76.9	96.2
2006	73.1	61.5	------*	69.2	100.0	92.3	------	92.3
2007	24.1	48.3	44.8	72.4	44.8	89.7	69.0	100.0
2008	100.0	80.0	55.0	95.0	100.0	100.0	100.0	100.0
5-Season	69.6	60.8	54.5	82.4	85.6	88.0	84.8	97.6

*: P22 is one of the missing data points in 2006.

Table 3.5 The root mean square error of predicted field mean soil moisture using the rank stable locations

Year	Root Mean Square Error (RMSE) (Volumetric soil moisture (cm^3/cm^3))						
	P10	**P20**	**P22**	**P39**	**Average of 2004-based OSLs**	**Average of 4 random locations**	**Average of 4 guided locations**
2004	0.010	0.013	0.009	0.009	0.007	0.007	0.009
2005	0.020	0.041	0.034	0.017	0.016	0.007	0.011
2006	0.017	0.024	------*	0.025	0.013	0.007	0.008
2007	0.054	0.028	0.040	0.017	0.023	0.008	0.009
2008	0.007	0.018	0.019	0.009	0.012	0.005	0.007
5-Season	0.029	0.027	0.029	0.017	0.016	0.010	0.009

*: P22 is one of the missing data points in 2006.

CHAPTER 4. GENETIC ALGORITHM FOR PARAMETER SELECTION TO PREDICT SOIL MOISTURE PATTERN USING TOPOGRAPHIC INDICES

A paper to be submitted to Transactions of ASABE

Lingyuan Yang and Amy Kaleita

4.1. Abstract

Despite the well-known fact that soil moisture highly varies in both space and time, the recurring spatial soil moisture pattern in time exists, which provide unique opportunity to understand complex soil moisture variability. In this study, near-surface soil moisture data were collected across two working fields, Brooks and Iowa Validation Site (IVS), southwest of Ames, Iowa, in multiple growing seasons in order to identify the recurring patterns. Multiple linear regression was used to predict recurring soil moisture patterns using topographic indices at optimal resolutions. A genetic algorithm was developed to select the input independent variables over a range of resolutions for multiple linear regression models. Using this approach, not only were the primary influential topographic indices to soil moisture patterns uncovered, but the most appropriate resolutions for each influential index was identified. The recurring patterns in two fields, Brooks and IVS, were well predicted by the combination of static topographic indices at optimal resolutions.

Keywords: Evolutionary computation, scaling, soil water content

4.2. Introduction

The surface soil moisture is an important media for the water and energy exchanges between atmosphere and the land surface. It controls the processes of exchanging the partitioning solar energy into latent sensible heat, and the processes of turning the precipitation into surface runoff and infiltration.

Soil moisture highly varies in both space and time, due to the complexity of hydrologic/natural systems and the factors (such as atmosphere forcing, topography, vegetation, soil heterogeneity, etc.) governing the movement of water into and through the soil. Even though high spatio-temporal variability of soil moisture is a well known fact, results from numerous studies have implied that some spatial patterns of soil moisture repeat to appear in time, which we define as underlying/recurring spatial patterns of soil moisture. It has been found in different study regions, some locations are always drier than the regional average soil moisture condition, some are always wetter, and some are always close to the regional average [*Vachaud et al.,* 1985, *Cosh et al.*, 2004; *Cosh et al.*, 2006; *Vivoni et al.*, 2008]. These recurring soil moisture patterns provide a unique opportunity to understand the complex soil moisture variability much easier by decomposing the soil moisture variability into deterministic/recurring and random components.

Numerous studies have been done on the spatio-temporal variability/patterns of soil moisture [*Charpentier and Groffman,* 1992; *Western et al.,* 1999; *Gómez-Plaza et al.,* 2001; *Wilson et al.,* 2004; *De Lannoy et al.,* 2006; *Brocca et al.,* 2007; *Choi and Jacobs,* 2007; *Das and Mohanty*, 2008; *Penna et al.,* 2009]. Many studies have related the soil moisture patterns to various factors/variables for different study regions in different seasons. [*Western et al.,* 1999; *Gómez-Plaza et al.,* 2001*; De Lannoy et al.,* 2006]. At the intermediate scales (subwatershed or catchment) and hillslope scales, it has been found that soil moisture varies spatially due to the characteristics of topographic indices [*Hawley et al.,* 1983; *Charpentier and Groffman,* 1992; *Moore et al.,* 1993; *Western et al.,* 1998; *Western et al.,* 1999; *Western* *and Blöschl,* 1999; *Mohanty et al.,* 2000a,b], soil properties [*Famiglietti et al.,* 1998], vegetation [*Hupet and Vanclooster*, 2002] or the combination of the above factors [*Mohanty and Skaggs,* 2001; *Jacobs et al.,* 2004]. *Mohanty and Skaggs* [2001], *Famiglietti et al.* [1998], and *Jacobs et al.* [2004] have provided in details concerning every influential factor and how each one impacts soil moisture variability. In summary, the soil characteristics affect the fluid transmission and retention properties and thus soil moisture content, through the variations in texture, organic matter content, soil structure, Aledo, and the presence of macro-pores. The vegetation characteristics including canopy, roots influence the runoff, interception, evapotranspiration processes and soil hydraulic conductivities, in turn soil moisture variability. The topographic indices (elevation / relative elevation, slope, aspect (or slope orientation), curvature, and upslope contributing area, etc) all affect soil moisture dynamics by different mechanisms. The slope affects the infiltration, drainage and runoff processes; the aspect (or slope orientation) influences the solar radiation, thus evapotranspiration, and then soil moisture; the curvature is a measurement of convexity or concavity of the landscape; and the upslope contributing are control the potential volume of the subsurface moisture flowing past a particular point on the landscape. Most of the studies mainly focused on the qualitative explanation of the relationship between soil moisture variability and the influential factors.

Prediction or estimation of soil moisture patterns is useful for reliable inputs for hydrologic models, design of in situ soil moisture field campaign or networks, validating remotely sensed soil moisture, and evaluating land surface models. There are studies which investigated the soil moisture variability as a quantitative function of those influential factors affecting the soil moisture, including empirical, physical and statistical models. These models require detailed information concerning characteristics of soil, vegetation, rainfall, etc. For example, *Isham et al.* [2005], *Rodriguez-Iturbe et al.* [2006] and *Manfreda and Rodriguez-Iturbe* [2006] have used analytically derived covariance function to characterize spatial pattern of soil moisture, which accounted in great details for soil characteristics, vegetation patterns, and rainfall dynamic in a region but neglected the topography. However, it is still a key challenge to find better ways to characterize the soil and vegetation [*Wilson et al.,* 2004]. The topography is the only readily available information in great details [*Wilson et al,* 2004], relatively inexpensive to obtain and only needs to obtain once. The technology (for example, ARCGIS) and digital elevation models [*Grayson et al.,* 1997] became also available and mature to deal with all the topographic data in great details. It is critical to look for a way to use the readily available information and technology to predict the recurring static soil moisture pattern in prior.

A fewer studies have used topographic indices to predict soil moisture patterns. *Western et al.* [1999] used multiple linear regression to model soil moisture pattern with terrain indices (wetness index, tangent curvature and potential radiation index) at Tarrawarra catchment in Australia, which captured 61% soil moisture spatial variation under wet conditions and up to 22% under dry conditions.

The resolution of topographic factors to be used in predicting recurring soil moisture patterns has never been considered. It is essential to know to which level the available detailed topographic information to be included, also important to provide guidance for the input sub-grid size for land surface modeling. In this paper, not only the influential factors are considered, but their products are included, as well as the optimal combination of resolutions for these factors. It is difficult to identify the best factors at the best resolutions for the functions using tradition multiple linear regression because the amount of possible combination between the factors and their resolutions is huge. Therefore, an effective optimization method is needed.

Genetic Algorithms (GAs) are adaptive heuristic search algorithms premised on the evolutionary ideas of natural selection and genetics. The basic concept of GAs is designed to simulate the Darwinian concept of "survival of the fittest." GAs are computationally simple yet powerful to provide robust search for difficult combinational search problems in complex spaces, without being stuck in local extremes [*Goldberg,* 1989a; *Tang, et al.,* 1999; *Steward, et al.,* 2005]. Therefore, GAs are powerful alternative tools to traditional optimization methods [*Goldberg,* 1989a].

In order to isolate and quantify the deterministic component of soil moisture variability which can be explained by topographic factors, the objective of this study is to develop and use a genetic algorithm to identify the influential topographic factors (slope, aspect, elevation and curvature) and their interactions at various resolutions. In doing so we hope to develop a tool, to identify not only the most important topographic factors for recurring soil moisture pattern development, but also the operative resolutions of each of those factors.

The rest of the paper is arranged as follows. First, in the section of materials and methods, we briefly describe the study area, the long-term surface soil moisture data and topographical data from two study sites used in the analysis, the method used to identify the recurring spatial pattern of collected soil moisture data, and the structure of developed genetic algorithm. Then the results and discussion are presented, followed by the conclusion.

4.3. Materials and Methods

4.3.1. Study Area and Data

In order to identify the recurring soil moisture patterns at the field scale, long-term, distributed near-surface soil moisture data was collected across two agricultural fields, Brooks Field and Iowa Validation Site (IVS), southwest of Ames, Iowa. Brooks is located directly south of IVS. Soil moisture was taken in the 2004, 2005, 2006, 2007, and 2008 growing seasons for Brooks Field, and in 2007 and 2008 growing seasons for IVS. Brooks' area is about 0.10km^2 and IVS' is around 0.77km^2. Both of them are corn-soybean rotation fields with moderate topographic variation. The climate in the study area is humid continental type with severe winter and hot summer. There is no dry season. January mean temperature is -7.0^{0}C and July mean temperature is 23.5^{0}C. Total average annual precipitation is 833 mm.

In Brooks, there were a total of 42 sampling locations for 2004, 2005, 2007, and 2008 (Figure 4.1). These sampling locations were on a regular 50 meter grid. Six of those locations were not sampled in 2006, which were labeled differently in Figure 4.1. In IVS, there were 33 sampling locations with 100-m or 200-m interval (Figure 4.1). The sampling time window during the day was consistent within each season and slightly different across the seasons.

At each sampling location in both fields, three readings for 0-6 cm depth were taken using a Theta probe moisture meter (Delta-T Devices, Cambridge UK, marketed in the United States by Dynamax, Inc., Houston, Tex.). The Theta probe output voltage values were converted to percentage volumetric soil moisture using the calibrated relationship for Des Moines Lobe soils provided by *Kaleita et al.* [2005]; the field component of that study was also conducted at the Brooks field. The average value of the three soil moisture contents at each location was used as the representative soil moisture at that location for the analysis in this study.

In each season data collection was begun only after planting; data collection in 2008 was delayed because heavy spring rains delayed planting. In every growing season, there were missing data points on some sampling days due to the presence of standing water on the site; missing data was generally only in the two “potholes” in the field, where water tended to pond. For the days when data was missing for more than one sampling location, the sampling day was not included in further analysis for the sake of consistency. In this study, one day was excluded for this reason in 2004, four days in 2005, and four days in 2006. This may bias the results slightly, since generally the days excluded from analysis were very wet. When these data are excluded, there remain a total of 24 days’ data (from mid-May to early July) in 2004, 26 days’ data (from early-May to Mid-August) in 2005, 26 days’ data (from mid-May to mid-July) of data in 2006, 29 days’ data (late May to Mid-August) in 2007, and 20 days’ data in 2008 (late June to late August) available for analysis. On average, soil moisture data were collected twice a week in 2005, 2007 and 2008, and three times a week in 2004 and 2006. Data collection overall captured days with wet and with dry soil moisture conditions, as well as the dry out process.

According to NRCS web soil survey (http://websoilsurvey.nrcs.usda.gov/app/), Brooks has 5 soil types (Figure 4.1): Nicollet loam, Harps loam, Webster clay loam, Clarion loam, and Canisteo clay loam. IVS has 6 types of soil, which include the above 5 types and Okoboji silty clay loam. The description of characteristics of each soil type is included in Table 4.1.

Elevation data was obtained by LiDAR (**Li**ght **D**etection **a**nd **R**anging) survey at a 2-m horizontal resolution at Brooks Field (Figure 4.1). LiDAR is an optical remote sensing technology that measures properties of scattered light to find range and/or other information of a distant target. The prevalent method to determine distance to an object or surface is to use laser pulses. Similar to radar technology, which uses radio waves instead of light, the range to an object is determined by measuring the time delay between transmission of a pulse and detection of the reflected signal. LiDAR has been used in particular as transformative method of obtaining detailed topographic information. It is attractive because it can see through many vegetation types, while obtaining information on vegetation height and density (*Robinson et al.,* 2008).

These LiDAR elevation data were used to create maps of topographic indices (slope, curvature, and aspect). The slope, defined as the first-order derivative of the topography, describes the rate of elevation change. The curvature, defined as the second derivative of the topography, describes the acceleration or deceleration of water flow over that surface. Negative curvatures mean concave surfaces or potholes, and positive curvatures mean convex surfaces or hilltop [*Kravchenko and Bullock,* 2000]. Aspect is defined as the direction where the slope is facing with unit of degree. 0 degree aspect corresponds to North, and the values of aspect increase clockwise. 2-m DEM (Digital Elevation Model) was generated in ArcGIS, then aggregated to larger resolutions (4 m, 8 m …… 80 m). Aspect, curvature, and slope at the various resolutions were calculated from the individual DEMs.

4.3.2. Spatial pattern of soil moisture

Mean relative difference, provided by *Vachaud et al.* [1985], was used to represent the soil moisture patterns at Brooks and IVS in the monitoring periods. The mean relative difference $\bar{\delta}_i$ and were calculated for all seasons together in each field. These parameters are defined by equation 1 and 2:

$$\delta_{ij} = \frac{\theta_{ij} - \bar{\theta}_j}{\bar{\theta}_j} * 100 \qquad (1)$$

$$\bar{\delta}_i = \frac{1}{m}\sum_{j=1}^{m}\delta_{ij} \qquad (2)$$

where θ_{ij} is the moisture at the ith location on the jth sampling occasion, $\bar{\theta}_j$ is the field average of all θ_{ij} on the jth sampling occasion, and m is the total number of sampling occasions. Locations with mean relative difference $\bar{\delta}_i$ close to zero indicate sites having soil moisture close to field mean. Locations with negative mean relative difference underestimated the field average soil moisture, meaning those locations are relatively drier. Locations with positive ones overestimated the field average, meaning those locations are relatively wetter.

4.3.3. Genetic algorithm structure

In order to determine the role of topography in the development of the common spatial distribution of soil moisture in the study fields, a genetic algorithm was developed and employed. The ultimate goal of the GA was to develop a model for prediction of mean relative difference based on a combination of important topographic indices computed at the most appropriate resolutions, respectively.

The technical details of the GA are as follows (Figure 4.2):

• Chromosome: Chromosomes are the abstract representations of candidate solutions to an optimization problem evolving toward better solutions. In this study, solutions are the predictive models for mean relative difference based on topographic data. The chromosomes were composed of 18 genes that described variables which were selected by MLR (Multi-Linear Regression) models. There are two parts of genes in each chromosome. The first part, including 4 genes, was encoded in integers representing: (1) scale of aspect; (2) scale of curvature; (3) scale of elevation; (4) scale of slope, the second part, including 14 genes, was encoded in binary string representing: (5) – (18) whether the variables (aspect, curvature, elevation, and slope), their squares, and their interactions (or products) were selected in the model or not. "1" represented selected, and "0" meant not selected.

• Population size: One of the advantages of genetic algorithms over traditional optimization and search procedures is that GAs search from a population of solutions, not a single solution. The rule of thumb, suggested by *Goldberg* [1989b], a population size is approximately equal to the chromosome length. For our study with a chromosome length of 38 bits for Brooks IVS, a population size of 40 was used for both Brooks and IVS for the convenience of the following steps.

• Fitness: A fitness value is a particular type of objective function value that quantifies the optimality of a solution (that is, a chromosome) in a genetic algorithm so that that particular chromosome may be ranked against all the other chromosomes. The fitness for an individual was the 1/RMSECV (Root Mean Square Error of Cross Validation) associated with that individual calculated through a MLR cross validation procedure. Analysis was replicated 24 times with randomly generating 24 different starting populations. Cross validation was accomplished using leave-one-out procedure. Model performance was measured using RMSECV. The model with the minimum RMSECV (maximum fitness) in the population was the best one.

• Selection: During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (as measured by a fitness function) are typically more likely to be selected. In this study, the population of chromosomes was ranked by fitness value. Absolute fitness replacement was implemented by sorting the population by fitness and discarding the less fit half of the population. The upper half was then replicated to form a new lower half [*Steward et al.,* 2005], processed for the following crossover and mutation.

• Crossover and mutation: The next step is to generate a second generation population of solutions from those selected through genetic operators: crossover (also called recombination), and/or mutation. For each new solution to be produced, a pair of "parent" solutions is selected for breeding from the pool selected previously. By producing a "child" solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its "parents". New parents are selected for each child, and the process continues until a new population of solutions of appropriate size is generated. For the population selected from above step, the children were generated using double-point crossover, which makes sure both integer and binary parts have cross-over point. Each child was selected for mutation. Four numbers (0, 1, 2, 3) were randomly generated, and 0, 1, 2, 3 means no mutation, mutation of the first part of the child, mutation of the second part, mutation of both parts, respectively.

• Stopping Criteria: The generational process is repeated until a stopping condition has been reached. This genetic algorithm stopped when one of these two conditions was satisfied: first, the program terminated when more than half of the population contained duplicate individuals; second, the process stopped if the generation number was greater than 500.

4.4. Results and Discussion

4.4.1. Mean Relative Difference (MRD) represented the reoccurring soil moisture patterns

The overall soil moisture conditions were summarized in Table 4.2. Among 125 sampling days within 5 growing seasons in Brooks, the minimum and maximum daily field mean soil moisture was 0.14 and 0.31 cm^3/cm^3, respectively. And the minimum and maximum standard deviation of daily field mean soil moisture was 0.02 and 0.04 cm^3/cm^3, respectively. Among 34 sampling days within 2 growing seasons in IVS, the minimum and maximum daily field mean soil moisture was 0.16 and 0.32 cm^3/cm^3, respectively. And the minimum and maximum standard deviation of daily field mean soil moisture was 0.02 and 0.04 cm^3/cm^3, respectively. The field wetness conditions in both fields, indicated by field mean soil moisture, were changing throughout the investigation periods. For the purpose of empirical modeling for the deterministic component of soil moisture variability, the mean soil moisture is not appropriate, since it changes with other varying factors such as precipitation or evapotranspiration. To represent the long-term underlying, recurring soil moisture patterns, using MRD is better than the average value of soil moisture in many aspects. MRD is the relative wetness condition to the field mean soil moisture regardless of the actual mean soil moisture value. Using MRD, the mean soil moisture condition is not required to be considered in many different years with different precipitation amount. While, the mean soil moisture value of multiple years/seasons is biased towards either dry or wet conditions depending on which one is dominating.

The MRDs in both fields represented the common spatial patterns of soil moisture throughout the investigation periods, with some locations wetter than the field average, some locations drier than the field average, and some locations around the field average (Figure 4.3). According to results of the coefficient of determination (R^2), on average MRD could explain 48% variance of daily field soil moisture for Brooks and 36% for IVS throughout their own investigation period.

4.4.2. Genetic Algorithm (GA) selected topographic index and their resolutions for multiple linear regression (MLR) models

After a few replications of GA program, top 20 best models with best fitness values have been selected for Brooks field. GA not only selected the topographic indices included in the models but their resolutions. Among the 20 models, cos(Aspect) and curvature were the most popular ones, which have been included in the 20 models for 20 and 18 times, respectively (Table 4.3). They maintained their popularities when moving from top 20 models to top 10 and the best one. The other topographic indices, like slope, products of cos(Aspect) and curvature, cos(Aspect) and elevation, cos(Aspect) and slope, curvature and elevation, elevation and slope, have increasing popularities when moving from top 20 models to top 10 and the best one (Table 4.3), which means they are more important for better model performance.

Figure 4.4 (left side) shows the histograms of all the resolutions for each topographic index selected in the models for Brooks. The resolution of slope was always 38 m for all 20 models, and the resolutions for the rest indices had less consistency among the 20 models. But, when the 20 models were divided into different levels according to their performance, the resolutions for each level had high consistency. Among the top 5 models for Brooks (Table 4.4), the selected resolutions of aspect, elevation and slope were always the same, and the selected resolutions of curvature were 62 m and 64 m. Among another 5 models from model 8 to 12, the resolutions for each index were persistent except the resolutions of elevation were 58 m and 60 m. For model 15 to 17, the selected resolutions for each index were the same. Model 18 and 19 had consistency in the resolution selections.

Similarly, the topographic indices and their resolutions for IVS selected by GA were presented in Table 4.3 and Figure 4.4. Slope, $\cos^2$(Aspect), products of cos(Aspect) and slope, curvature and elevation, curvature and slope are the most popular ones, which had been included in the 20 models for 18, 20, 20, 19 and 19 times, respectively (Table 4.3). cos(Aspect), elevation, elevation^2, products of cos(Aspect) and elevation, elevation and slope had increasing popularities when narrowing down to smaller amount of models.

On the right side of Figure 4.4, histograms of all the resolutions for each topographic index selected in 20 models for IVS were presented. The resolutions of curvature and slope were more consistent in comparison to those of the other two indices. When the 20 models were divided into different levels according to their performance, the resolutions for each level had high consistency (Table 4.5). Among the top 10 models, the selected resolutions of aspect were 24m (twice), 60m (4 times) and 82m (4 times), and ones for curvature were 54m (9 times) and 98m (once), and ones for slope were 4m (8 times), 8m (once) and 22m (once).

The resolutions of elevation were the least consistent ones. For the bottom 10 models for IVS (Table 4.5), the resolutions of three indices did not change. The resolutions of elevation were scattered along the spectrum, since the elevation were not selected in the bottom 10 models except in the form of product of elevation and curvature.

The selected resolutions of each topographic index for Brooks and IVS were completely different, although two fields are next to each other. The possible reason for the difference of selected resolutions of two fields might be that the extent of IVS is about 6 times bigger than that of Brooks, and the spatial intervals of soil moisture sampling in IVS were as twice or four times big as the ones in Brooks.

4.4.3. Topographic indices predicted the reoccurring soil moisture patterns

The topographic indices selected by the best MLR model (model 1) for Brooks were cos(Aspect), curvature, slope, and products of cos(Aspect) and curvature, cos(Aspect) and elevation, cos(Aspect) and slope, curvature and elevation, and elevation and slope (Table 4.3). And the resolutions associated with aspect, curvature, elevation and slope were 36m, 62m, 72m, and 28m (Table 4.4), respectively. For IVS, the selected indices by the best model were cos(Aspect), elevation, $\cos^2$(Aspect), elevation^2, products of cos(Aspect) and elevation, cos(Aspect) and slope, curvature and elevation, curvature and slope, and elevation and slope. And the resolutions associated with aspect, curvature, elevation and slope were 60, 54, 68 and 4 meters.

Predicted MRDs by the best model were shown in Figure 4.5, with top left for Brooks and top right for IVS. The coefficients of determination (R-Square) from multiple regression analysis between MRD and selected topographic indices were 0.88 for Brooks and 0.80 for

IVS, respectively. The scatter plots between MRD and the selected influential indices were presented in Figure 4.6 (Brooks) and Figure 4.7 (IVS). There were no obvious trends for most of them. For Brooks, the plot between MRD and curvature, the plot between MRD and product of curvature and elevation might show trends of negative relationships (Figure 4.6). For IVS, the plot between MRD and the product of elevation and slope might show trend of negative relationship (Figure 4.7). Although there was no direct relationship between the recurring soil moisture patterns and most of the topographic indices or there products, the combination of all the selected variables could predict well the recurring soil moisture patterns in both fields, which means each topographic index might not influence soil moisture patterns individually but interactively.

The selected topographic indices and their resolutions by the best model for Brooks have been tested on IVS. The coefficient of determination was 0.60 and RMSECV was 6.94 (Figure 4.5, bottom left). The best model for Brooks did not really work the best for IVS. However, by the combination the topographic indices selected by the best model for Brooks and the resolutions selected by the best model for IVS, the coefficient of determination increased to 0.74 and RMSECV decreased to 5.29 (Figure 4.5, bottom right), which got closer to the best model for IVS. The results proved that there were optimal resolutions for each topographic index to predict the recurring soil moisture patterns. Therefore, the impacts of topographic indices on the recurring patterns of soil moisture were similar in Brooks and IVS, since there are in the same topographic region. But the resolutions of influential indices are different for the two fields, which may imply that the resolutions associated with topographic indices are site-specific and could be controlled by some other factors.

As a framework for discussion, the scale triple to characterize the scale of measurements by Bloshchl and Sivapalan [1995] and Western et al. [2002] is considered. According to their opinions, there are three scale aspects: support (the area over which a measurement averages the heterogeneities), spacing (the distance between two measurements), and extent (the overall coverage of study area). The increasing support will cause the loss of small-scale features, the larger spacing will cause the loss of smaller-scale variability, and the smaller extent will cause the loss of large-scale features [Western et al., 2002]. In this study, IVS has larger spacing and extent than Brooks, therefore, the results from IVS may include more larger-scale features but less detailed variability for smaller-scale features. And the aggregated topographic indices (aspect, curvature, elevation and slope) at different resolutions provided different levels of details concerning information of each variable. The different selection of influential topographic indices at different resolutions for Brooks and IVS in this study may reveal that with different spacing and extent, not only the relationships of influential topographic variables on soil moisture patterns have changed, but the resolutions of the influential variables. This illustrated that the changing relationships between soil moisture patterns and topographic indices being observed across different scales. Similar results have been found by Robinson et al. [2008] that there were changing coefficients of regression equations between soil moisture and sand content not only with different spacing but also whether aggregated or point measurements were considered.

4.5. Conclusions

There were deterministic recurring soil moisture patterns in the long-term observation period in the study area, captured by mean relative difference of soil moisture which represented the relative wetness condition of the field. The recurring patterns in two fields, Brooks and IVS, were well predicted by the combination of static topographic indices at optimal resolutions. The developed genetic algorithm was able to not only select the influential topographic indices for predicting recurring soil moisture patterns, but identify their optimal resolutions.

4.6. Acknowledgements

This work was supported by NASA grant NNG06GC63G from the Land Surface Hydrology program, and by the Iowa State University Environmental Science Fellowship and the Iowa Space Grant Consortium Program for Space-borne and Earthbound System Sustainability.

4.7. References

Beven, K.J. and M. J. Kirkby, 1979. A physically based, variable contributing area model of basin hydrology. *Hydrol Sci. Bull,* 24 (1): 43-69.

Brocca, L., R. Morbidelli, F. Melone, T. Moramarco, 2007. Soil moisture spatial variability in experimental areas of central Italy. *J. Hydrol.*, 333, 356–373.

Charpentier, M. A., and P. M.Groffman, (1992). Soil moisture variability within remote sensing pixels. *J. Geophys. Res.,* 97 (D17), 18987–18995.

Choi, M., J. M. Jacobs, and M. H. Cosh, 2007. Scaled spatial variability of soil moisture fields. *Geophys. Res. Lett., 34,* L01401, doi:10.1029/2006GL028247.

Cosh, M. H., T. J. Jackson, R. Bindlish, J. H. Prueger, 2004. Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sens. Environ.*, 92: 427-435.

Das, N. N., and B. P. Mohanty, 2008. Temporal dynamics of PSR-based soil moisture across spatial scales in an agricultural landscape during SMEX02: A wavelet approach, *Remote Sens. Environ.*, 112(2), 522–534,doi:10.1016/j.rse.2007.05.011.

David J. W., A. W. Western, R. B. Grayson, A. A. Berg, M. S. Lear, M. Rodell, J. S. Famiglietti, R. A. Woods, T. A. McMahon, 2003. Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. *J. Hydrol.*, 276, 254–274.

De Lannoy, G. J. M., N. E. C.Verhoest, P. R. Houser, T. J. Gish, and M. Meirvenne, 2006. Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE3), *J. Hydrol.*, 331, 719–730.

Famiglietti, J. S., J. W. Rudnicki, and M. Rodell, 1998. Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas. *J. Hydrol.*, 210, 259–281.

Famiglietti, J. S., J. A. Devereaux, T. J. Jackson, S. T. Graham, C. A. Laymon, M. Rodell, T. Tsegaye, P. R. Houser, and P. J. van Oevelen, 1999. Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resour. Res.,* 35 (6), 1839–1851.

Farnham, I., K. Johannesson, A. Singh, V. Hodge, and K. Stetzenbach (2003), Factor analytical approaches for evaluating groundwater trace element chemistry data, *Anal. Chim. Acta,* 490, 123–138.

Goldberg, D. E. 1989a. Genetic Algorithms in Search, Optimization and Machine Learning. Reading, Mass.: Addison-Wesley Publishing Co., Inc.

Goldberg, D. E., 1989b, Sizing populations for serial and parallel genetic algorithms. In Proc. 3rd Int. Conference on Genetic Algorithms, 70-79, San Mateo, Calif: Morgan Kaufman.

Gómez-Plaza, A., M. Martınez-Mena, J. Albaladejo, and V. M. Castillo, 2001. Factors regulating spatial distribution of soil water content in small semiarid catchments. *J. Hydrol.*, 253, 211–226.

Grayson, R. B., A. W. Western, F. H. S. Chiew, and G. Blöschl, 1997. Preferred States in Spatial Soil Moisture Patterns: Local and Nonlocal Controls, *Water Resour. Res.*, 33(12), 2897–2908.

Hawley, M. E., T. J. Jackson, and R. H. McCuen, 1983. Surface soil moisture variation on small agricultural watersheds. *J. Hydrol.*, 62: 179–200.

Isham, V., D. R. Cox, I. Rodríguez-Iturbe, A. Porporato, and S. Manfreda, 2005. Mathematical characterization of the space-time variability of soil moisture, *P. R. Soc. A*, 461(2064), 4035 – 4055, doi:10.1098/rspa.2005.1568.

Jacobs, J. M., B. P. Mohanty, E. Hsu, and D. Miller, 2004. SMEX02: Field scale variability, time stability and similarity of soil moisture. *Remote Sens. Environ.*, 92, 436-446.

Johnson, R. and D. Wichern, 2002. Applied Multivariate Statistical Analysis, 5th ed., Prentice-Hall, Upper Saddle River, N. J.

Kaleita, A. L., J. L. Heitman, S. D. Logsdon. 2005. Field calibration of the Theta Probe for Des Moines Lobe soils. *Applied Engineering in Agriculture*, 21(5), 865-870.

Kaleita, A. L., B. L. Steward, R. P. Ewing, D. A. Ashlock, M. E. Westgate, and J. L. Hatfield, 2006. Novel analysis of hyperspectral reflectance data for detecting onset of pollen shed in maize*, Trans. ASABE,* 49 (6), 1947-1954.

Kravchenko A. N. and D. G. Bullock, 2000. Correlation of corn and soybean grain yield with topography and soil properties. *Agron. J.,* 92: 75-83.

Manfreda, S., and I. Rodríguez-Iturbe, 2006. On the spatial and temporal sampling of soil moisture fields, *Water Resour. Res.*, 42, W05409, doi:10.1029/2005WR004548.

Mohanty, B. P., J. S. Famiglietti, and T. H. Skaggs, 2000 (a). Evolution of soil moisture spatial structure in a mixed vegetation pixel during the Southern Great Plains 1997 (SGP97) hydrology experiment. *Water Resour. Res.,* 36(12), 3675–3686.

Mohanty, B. P., T. H. Skaggs, and J. S. Famiglietti, 2000 (b). Analysis and mapping of field-scale soil moisture variability using high-resolution ground based data during the Southern Great Plains 1997 (SGP97) hydrology experiment. *Water Resour. Res.,* 36, 1023–1032.

Mohanty, B. P., and T. H. Skaggs, 2001. Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing foot- prints with varying soil, slope, and vegetation. *Adv Water Resour.*, 24, 1051–1067.

Moore, I. D., G. J. Burch, and D. H. Mackenzie, 1988. Topographic effects on the distribution of surface soil water and the location of ephemeral gullies. *Trans. ASAE,* 31(4): 1908-1107.

Moore, I. D., T. W. Morton, and J. E. Williams, 1993. Modelling environmental heterogeneity in forested landscapes, *J. Hydrol.*, 150, 717-747.

Penna, D., M. Borga, D. Norbiato, G. D. Fontana, 2009. Hillslope scale soil moisture variability in a steep alpine terrain. *J. Hydrol.*, 364, 311–327.

Robinson, D. A., C. S. Campbell, J. W. Hopmans, B. K. Hornbuckle, S. B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth, 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7 (1), 358-389.

Rodríguez-Iturbe, I., V. Isham, D. R. Cox, S. Manfreda, and A. Porporato, 2006. An analytical framework for the space-time variability of soil moisture fields with heterogeneous vegetation, *Water Resour. Res.*, 42, W06D05, doi:10.1029/2005WR004497.

Starks, P. J., G. C. Heathman, T. J. Jackson, M. H. Cosh, 2006. Temporal stability of soil moisture profile. *J. Hydrol.*, 324, 400-411.

Steward, B.L., A.L. Kaleita, R.P. Ewing, and D.A. Ashlock, 2005. Genetic algorithms for hyperspectral range operator selection, *ASABE paper No. 053063.* St. Joseph, Mich.: ASABE.

Syed, T. H., V. Lakshmi, E. Paleologos, D. Lohmann, K. Mitchell, and J. S. Famiglietti (2004), Analysis of process controls in land surface hydro- logical cycle over the continental United States, *J. Geophys. Res.,* 109, D22105, doi:10.1029/2004JD004640.

Tang, L., L. F. Tian, and B. L. Steward, 2000, Color image segmentation with genetic algorithm for in-field weed sensing*, Trans. ASAE,* 43, 1019-1027.

Vachaud, G., A. Passerat de Silans, P. Balabanis, and M. Vauclin, 1985. Temporal stability of spatially measured soil water probability density function. *Soil Sci. Soc. Am. J.*, 49, 822-828.

Western, A. W., G. Blöschl, and R. B.Grayson, 1998. Geostatistical characterization of soil moisture patterns in the Tarrawarra catchment. *J. Hydrol.*, 205, 20-37.

Western, A. W., and G. Blöschl, 1999. On the spatial scaling of soil moisture. *J. Hydrol.*, 217, 203–224.

Western, A. W., R. B. Grayson, G. Blöschl, G. R. Willgoose, and T. A. McMahon, 1999. Observed spatial organization of soil moisture and its relation to terrain indices, *Water Resour. Res.,* 35(3), 797–810.

Western, A.W., R.B. Grayson, and G. Blöschl, 2002. Scaling of soil moisture: a hydrologic perspective. *Ann. Rev. Earth Planet. Sci.*, 205: 20–37.

Western, A. W., S. Zhou, R. B. Grayson, T. A. McMohon, G. Blöschl, D. J. Wilson, 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *J. Hydrol.*, 286: 113-134.

Wilson, D. J., A. W. Western, and R. B. Grayson, 2004. Identifying and quantifying sources of variability in temporal and spatial soil moisture observations, *Water Resour. Res.*, 40, W02507, doi:10.1029/2003WR002306.

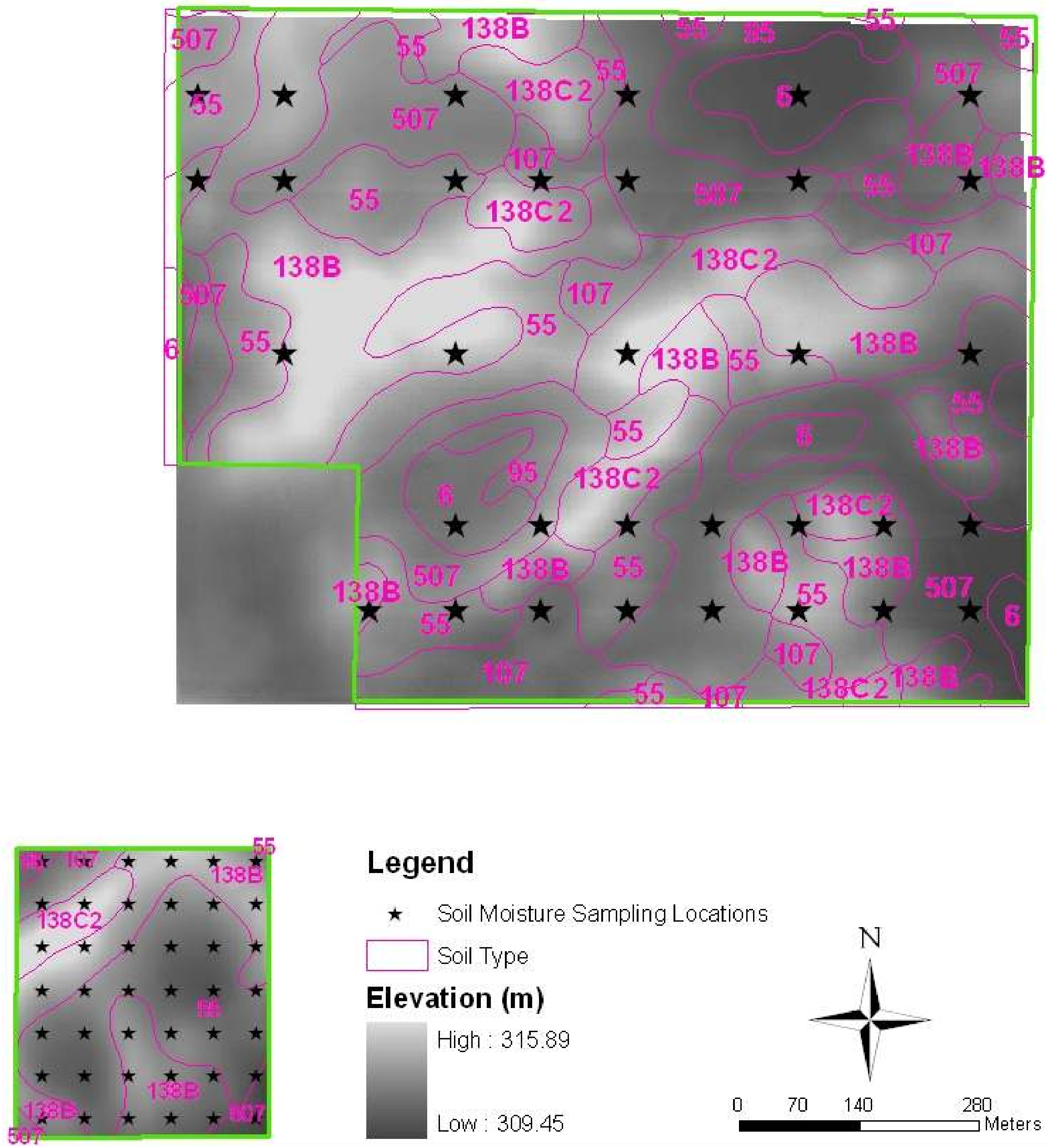

Figure 4.1 Soil Moisture Sampling Locations at Brooks Field (Left bottom) and IVS (Right top).

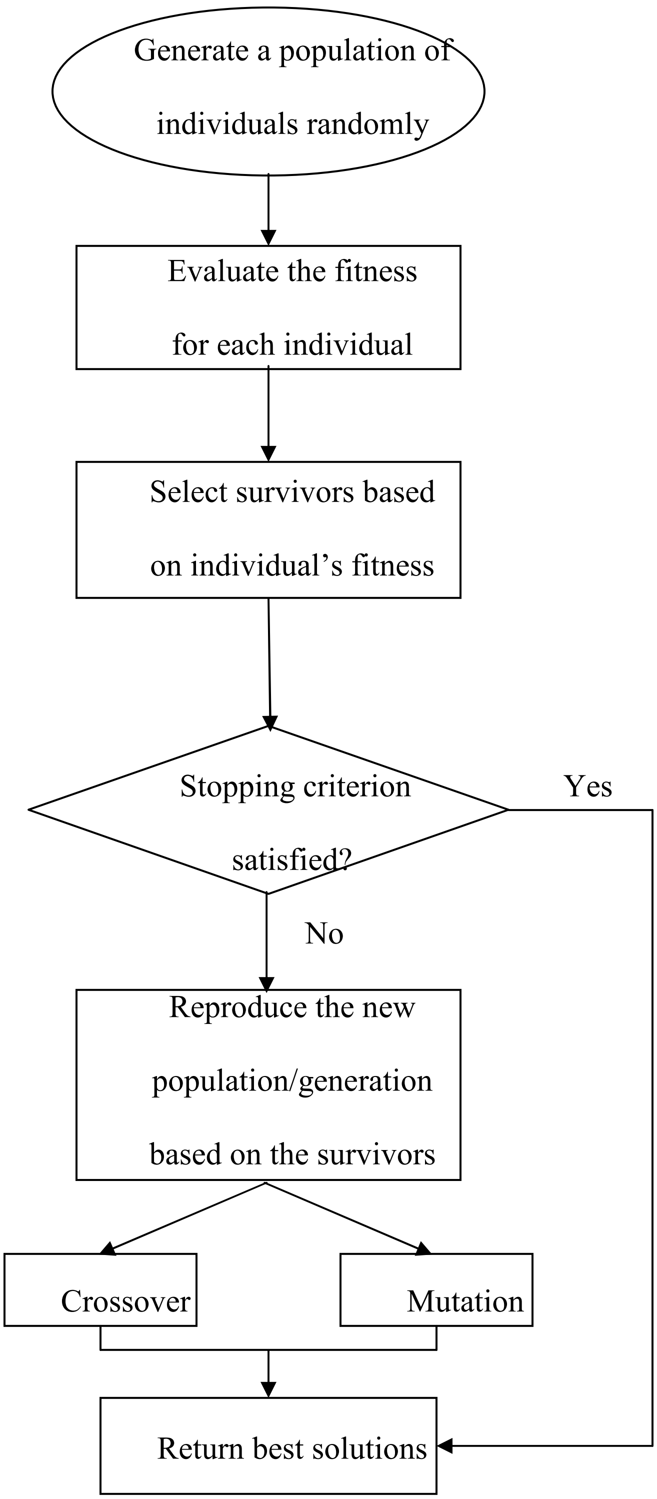

Figure 4.2 Flowchart of Genetic Algorithm (GA)

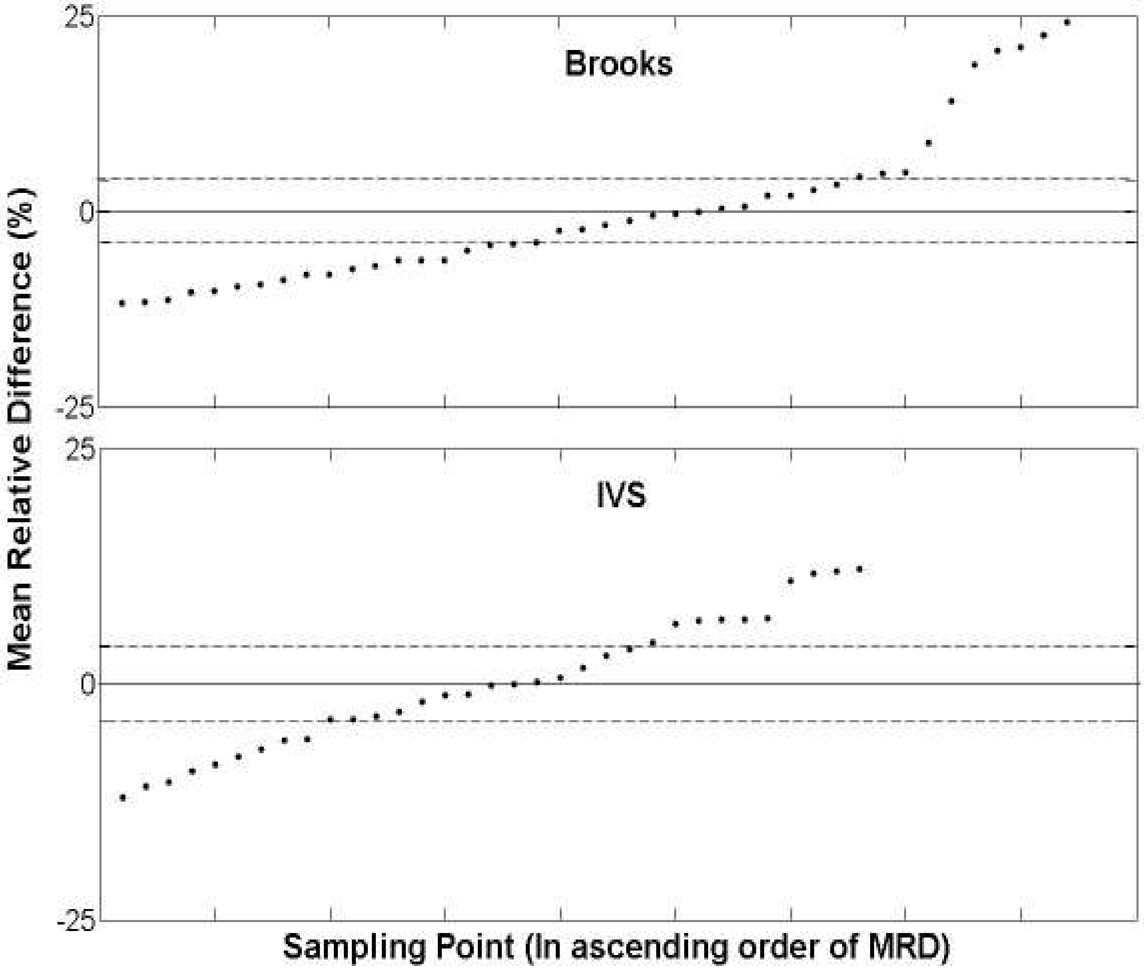

Figure 4.3 Mean Relative Difference (MRD) for Brooks (42 sampling locations) and IVS (33 sampling locations) in ascending order, and dashed line represent +/-4% MRD.

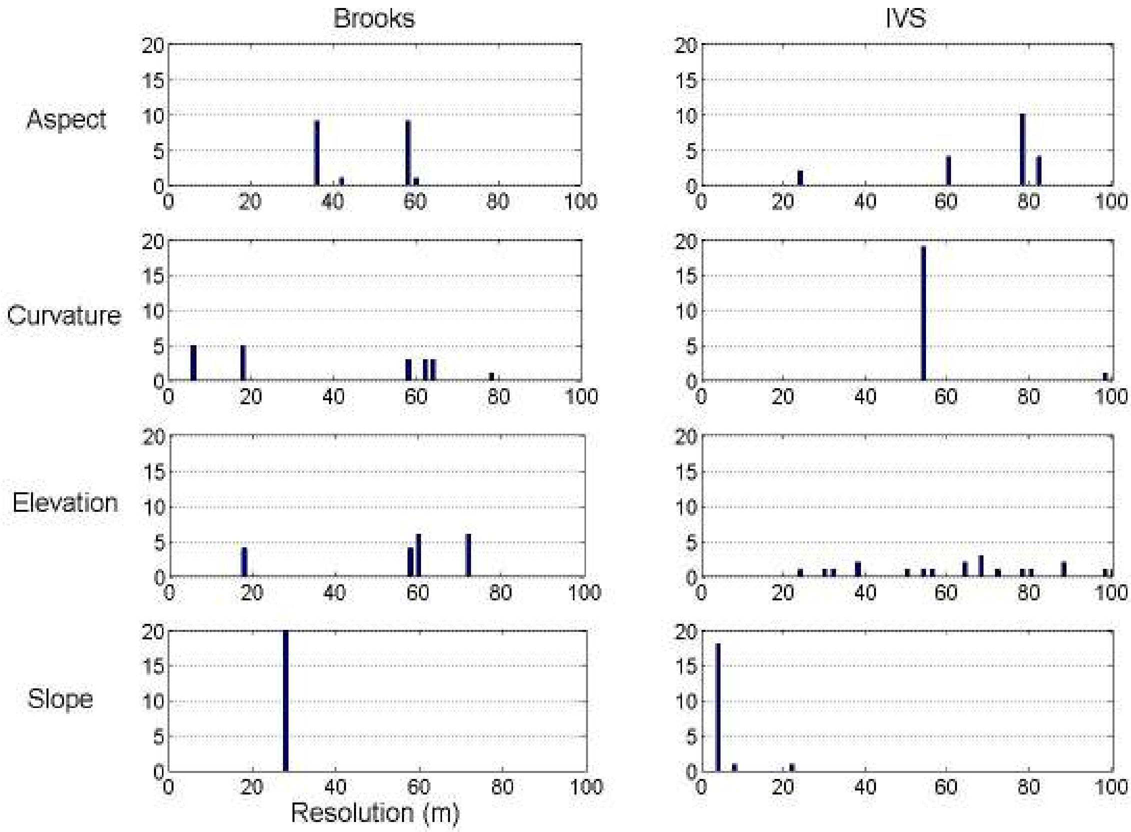

Figure 4.4 Histograms of selected resolutions for each topographical factor (each row) out of the best 20 models identified by GA for Brooks Field (Left) and IVS (Right)

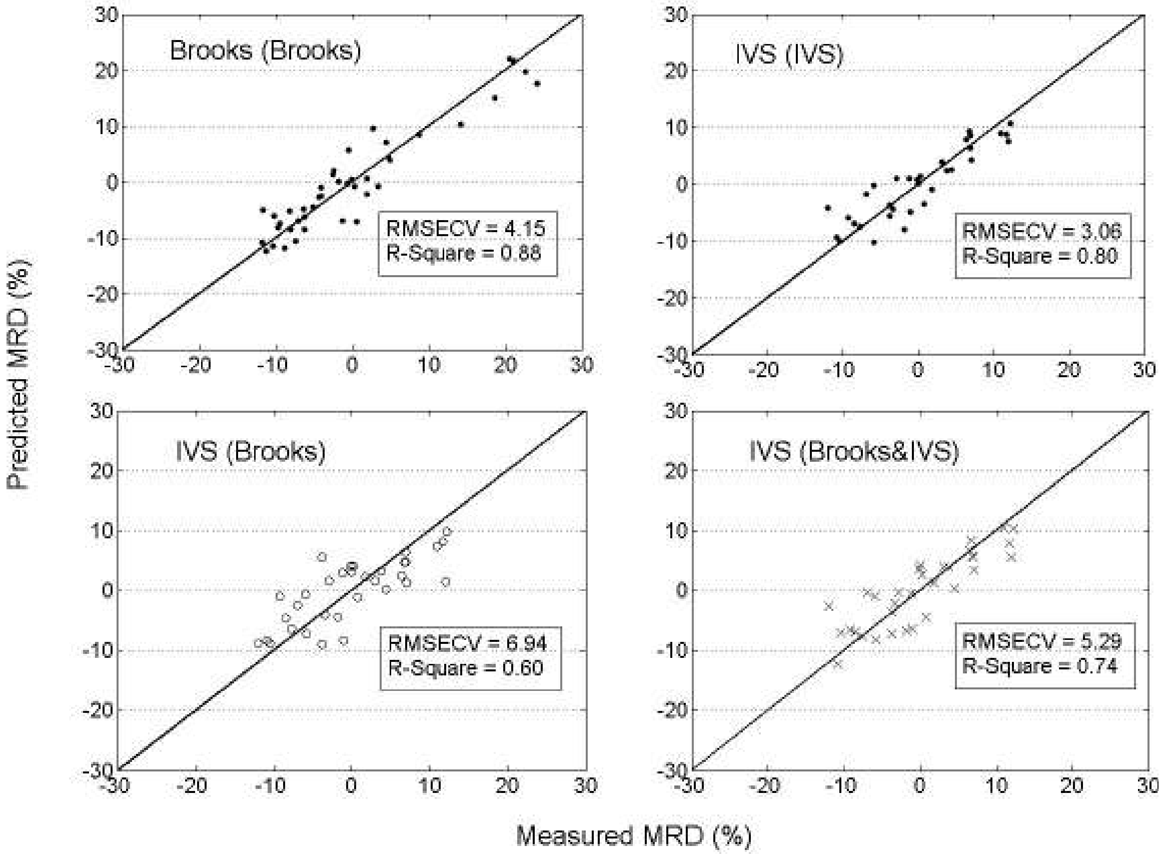

Figure 4.5 Model predicted MRD vs. Measured MRD (Top left: predicted MRD of Brooks based upon the model (both topographic indices and their resolutions) selected for Brooks; top right: predicted MRD of IVS based upon the model (both topographic indices and their resolutions) selected for IVS; bottom left: predicted MRD of IVS based upon the model selected for Brooks; bottom right: predicted MRD of IVS based upon the topographic indices selected for Brooks and the resolutions selected for IVS).

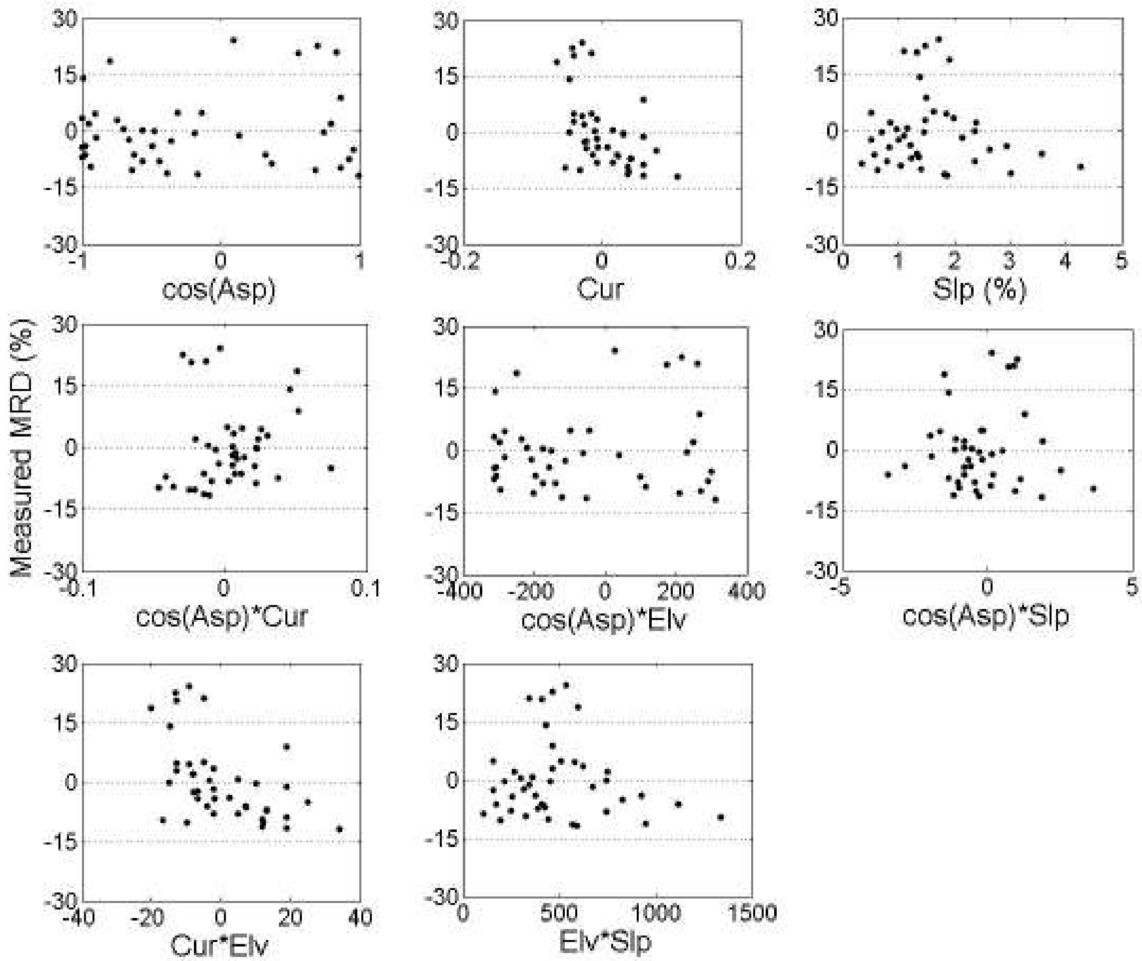

Figure 4.6 The scatter plot (MRD vs. The best model selected topographic indices) for Brooks (The abbreviations in the figure are, Asp: Aspect; Cur: Curvature; Elv: Elevation; Slp: Slope)

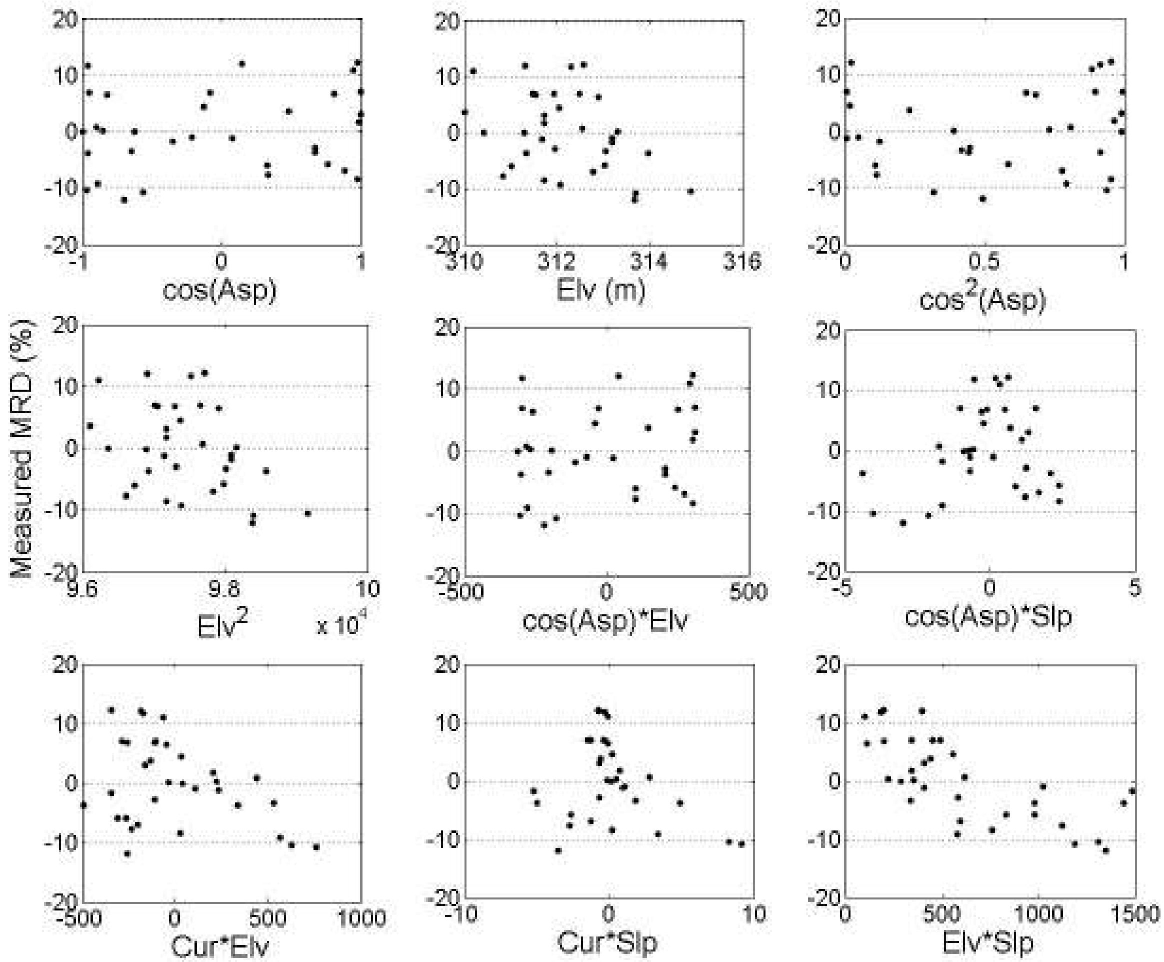

Figure 4.7 The scatter plot (MRD vs. The best model selected topographic indices) for IVS (The abbreviations in the figure are, Asp: Aspect; Cur: Curvature; Elv: Elevation; Slp: Slope)

Table 4.1 Description of soil types in Brooks and IVS

Soil Type	ID	Characteristics	Taxonomic Class	Typical Pedon
Okoboji silty clay loam	6	Very deep, very poorly drained soils formed in alluvium or lacustrine sediments. These soils are in closed depressions on till plains and moraines.	Fine, smectitic, mesic Cumulic Vertic Endoaquolls	In a depression, in a cultivated field, on 0-1 percent slope.
Nicollet loam	55	Very deep, somewhat poorly drained soils that formed in calcareous loamy glacial till on till plains and moraines	Fine-loamy, mixed, superactive, mesic Aquic Hapludolls	On a 1-3 percent plane slope in a cultivated field
Harps loam	95	Very deep, poorly drained soils formed in till or alluvium derived from till. Harps soils are on narrow rims or shorelines of depressions on till plains and moraines	Fine-loamy, mixed, superactive, mesic Typic Calciaquolls	A nearly level rim of a depression, in a cultivated field
Webster clay loam	107	very deep, poorly drained, moderately permeable soils formed in glacial till or local alluvium derived from till on uplands	Fine-loamy, mixed, superactive, mesic Typic Endoaquolls	On a concave slope of about 0 to 2 percent gradient in a cultivated field.
Clarion loam	138B	Very deep, moderately well drained soils on uplands. These soils formed in glacial till	Fine-loamy, mixed, superactive, mesic Typic Hapludolls	On a convex upland with a slope of 2-5 percent, in a cultivated field
	138C2			On a convex upland with a slope of 5-9 percent, in a cultivated field
Canisteo clay loam	507	Very deep, poorly and very poorly drained soils that formed in calcareous, loamy till or in a thin mantle of loamy or silty sediments and the underlying calcareous, loamy till. These soils are on rims of depressions, depressions and flats on moraines or till plains.	Fine-loamy, mixed, superactive, calcareous, mesic Typic Endoaquolls	Nearly level to slightly convex slope (0 to 2 percent), on a ground moraine, in a cultivated field.

Table 4.2 Summary of field soil moisture content in the sampling periods at Brooks and IVS

	# of days	Minimum Daily Field Mean VSM (cm^3/cm^3)	Maximum Daily Field Mean VSM (cm^3/cm^3)
Brooks	125	0.14	0.31
IVS	34	0.16	0.32
		Minimum STD of Daily Field VSM (cm^3/cm^3)	Maximum STD of Daily Field VSM (cm^3/cm^3)
Brooks	125	0.02	0.04
IVS	34	0.02	0.04

Table 4.3 Influential parameters in the top 20, 10 and 1 best models for Brooks Field and IVS

Parameter	**Popularity for Brooks (%)**			**Popularity for IVS (%)**		
	Out of 20	Out of 10	Out of 1	Out of 20	Out of 10	Out of 1
cos(Aspect)	100	100	100	25	50	100
Curvature	90	90	100	5	10	0
Elevation	75	70	0	35	70	100
Slope	65	90	100	90	80	0
$\cos^2$ (Aspect)	50	40	0	100	100	100
Curvature^2	55	40	0	15	30	0
Elevation^2	45	70	0	30	60	100
Slope^2	80	60	0	5	10	0
cos(Aspect)*Curvature	45	60	100	60	20	0
cos(Aspect)*Elevation	55	70	100	25	50	100
cos(Aspect)*Slope	35	60	100	100	100	100
Curvature*Elevation	55	80	100	95	90	100
Curvature*Slope	5	0	0	95	90	100
Elevation*Slope	55	50	100	30	60	100

Table 4.4 Fitness values of the models and resolutions for each topographic index of the best 20 models identified by Genetic Algorithm (GA) for Brooks Field

Model ID	**RMSECV (%)**	**Resolution of aspect (m)**	**Resolution of curvature (m)**	**Resolution of elevation (m)**	**Resolution of Slope (m)**
1	4.15	36	62	72	28
2	4.26	36	62	72	28
3	4.35	36	64	72	28
4	4.36	36	64	72	28
5	4.36	36	64	72	28
6	4.39	60	18	60	28
7	4.42	36	62	72	28
8	4.44	58	6	58	28
9	4.44	58	6	58	28
10	4.44	58	6	60	28
11	4.45	58	6	58	28
12	4.45	58	6	60	28
13	4.46	58	18	58	28
14	4.48	36	58	18	28
15	4.49	58	18	60	28
16	4.49	58	18	60	28
17	4.49	58	18	60	28
18	4.49	36	58	18	28
19	4.49	36	58	18	28
20	4.51	42	78	18	28

Table 4.5 Fitness values of the models and resolutions for each topographic index of the best 20 models identified by Genetic Algorithm (GA) for IVS

Model ID	RMSECV (%)	Resolution of aspect (m)	Resolution of curvature (m)	Resolution of elevation (m)	Resolution of Slope (m)
1	3.06	60	54	68	4
2	3.06	60	54	68	4
3	3.22	24	54	88	22
4	3.22	82	54	38	4
5	3.22	60	54	64	4
6	3.22	82	54	38	4
7	3.22	82	54	32	4
8	3.22	60	54	64	4
9	3.23	24	98	72	8
10	3.25	82	54	30	4
11	3.26	78	54	68	4
12	3.26	78	54	100	4
13	3.26	78	54	98	4
14	3.26	78	54	88	4
15	3.26	78	54	80	4
16	3.26	78	54	78	4
17	3.26	78	54	56	4
18	3.26	78	54	24	4
19	3.26	78	54	54	4
20	3.26	78	54	50	4

CHAPTER 5. GENERAL CONCLUSIONS

5.1. Conclusions

This dissertation described efforts to answer a few challenging questions in designing *in situ* soil moisture observation network. In summary,

1) Rank stability analysis could assist to make decisions on where the *in situ* soil moisture stations should be, but it is not sufficient to use only one year or a few years' data to capture rank stable soil moisture behavior with rank stable locations, for the short-term remote sensing validation or any other similar application. Random sampling is possibly as good as targeted sampling for validating remotely sensed soil moisture using ground-based measurements. Further studies are needed for the application and limitation of rank stability analysis. Four locations at Brooks were able to estimate areal mean soil moisture condition at satisfactory level, but decision on the number of sampling locations might be site-specific.

2) It provides a unique opportunity to make understanding complex soil moisture variability easier by decomposing the variability into deterministic and random components, as well as to relate the deterministic component to static topographic variables. There were deterministic recurring soil moisture patterns in the long-term observation period in the study area, captured by mean relative difference of soil moisture which represented the relative wetness condition of the field. The recurring patterns in two fields, Brooks and IVS, were well predicted by the combination of static topographic indices at optimal resolutions. The developed genetic algorithm was able to not only select the influential topographic indices for predicting recurring soil moisture patterns, but identify their optimal resolutions.

5.2. Recommendations for *in situ* soil moisture network design

1) Three measurements of soil moisture were taken at each sampled location at both Brooks field and Iowa Validation Site (IVS), which helped to lower the risk of losing data. Replicated measurements of soil moisture at each sampling location are preferable for data quality control.

2) In the projects included in this dissertation, the knowledge of soil moisture patterns from Brooks field was used to plan the soil moisture sampling strategies in IVS. Therefore, pre-experiment in the study area is necessary for implementing permanent long-term *in situ* soil moisture network.

5.3. Prospects for future research

1) Four locations at Brooks were able to well estimate areal mean soil moisture condition, but more justification or methodologies are needed to in order to answer the question that how many samples are sufficient.

2) Not only topographic data is critical for predicting soil moisture patterns, but soil characteristics are also important. A greater emphasis is required on quantitative over qualitative soil data. Electromagnetic Induction (EMI) instruments can often provide useful quantitative soil information. One fundamental challenge is how to incorporate EMI data into prediction of soil moisture patterns. Future study on overcoming this challenge in modeling soil moisture patterns is suggested.

3) More efforts are suggested to move towards transfer the knowledge concerning spatio-temporal soil moisture patterns from gaged study area to ungaged watersheds and fields.

APPENDIX: GENETIC ALGORITHM CODE (IN MATLAB)

```
% The main loop, calling different functions in.
clear all; close all;

%Load independent variable data
aspect = load('C:\YANGS\Research Writings\2ndPaper\GAReRun\Brooks\
cos(Aspect)@Resolution.txt');
curvature = load('C:\YANGS\Research Writings\2ndPaper\GAReRun\Brooks\
Curvature@Resolution.txt');
elevation = load('C:\YANGS\Research Writings\2ndPaper\GAReRun\Brooks\
Elevation@Resolution.txt');
slope = load('C:\YANGS\Research Writings\2ndPaper\GAReRun\Brooks\
Slope@Resolution.txt');

%Load corresponding (dependent) variable data
MRD = load ('C:\YANGS\Research Writings\2ndPaper\GAReRun\Brooks\
5SeasonMRD.txt');

[npoint,nday] = size(MRD(:,:)); %size of the corresponding (dependent) variable data
if nday ~= 1
   sprintf('there is more than one dependent variable');
end

nvar = 4;   % number of independent variables
ncom = 14;  % number of production of independent variables
nvart = nvar+ncom; % the length of one individual
nscale = 40; % number of the scales;

%size of the independent variables
[npointA nscaleA] = size (aspect);
[npointC nscaleC] = size (curvature);
[npointE nscaleE] = size (elevation);
[npointS nscaleS] = size (slope);

 %check if the sizes of all the variables, including dependent and
%independent match
if npointA ~= npoint
   error ('The size of aspect does not match the dependent variable');
   else if npointC ~= npoint
         error ('The size of curvature does not match the dependent variable');
         else if npointE ~= npoint
               error ('The size of elevation does not match the dependent variable');
               else if npointS ~= npoint
```

```
                error ('The size of slope does not match the dependent variable');
            end
        end
    end
end

if nscaleA ~= nscale
  error ('The number of scaled aspect does not match the number of scales');
  else if nscaleC ~= nscale
      error ('The number of scaled curvature does not match the number of scales');
      else if nscaleE ~= nscale
          error ('The number of scaled elevation does not match the number of scales');
          else if nscaleS ~= nscale
              error ('The number of scaled slope does not match the number of scales');
            end
        end
    end
end

%organize the independent variables into one file
invar = zeros(npoint, nscale, nvar);
invar(:,:,1) = aspect;
invar(:,:,2) = curvature;
invar(:,:,3) = elevation;
invar(:,:,4) = slope;

%set up the parameters
npop = 40;  % number of individuals in the population
maxrep = 30;    %number of replication
maxgen = 500;   %maximum generation
convergence = 50; %convergence rate (%)
maxdups = ceil(npop*convergence/100); %allowed maximum duplication of individuals in
the population
crossover = 1; %crossover rate

%define the output files
bestfit = zeros(maxgen,maxrep);
meanfit = zeros(maxgen,maxrep);
bestind = zeros(maxgen,nvart,maxrep);
bestfit2 = zeros(maxgen, maxrep);
bestind2 = zeros(maxgen, nvart, maxrep);
predict = zeros(npoint, maxgen, maxrep);

%the loop starts
for rep = 1:maxrep %for each replication
```

```
    intChrom = crtrip(npop,nvar, [1 (nscale+1)], ncom);
                                  %Initialize the population (integer)
    gen = 0;
    dups = 0;
    while dups < maxdups

        for i = 1:npop
%           Check to see that model is not a repeat
            dups = 0;
            if i > 1
                for ii = 1:i-1
                    dif = sum(abs(intChrom(i,:)-intChrom(ii,:)));
                    if dif == 0
                        dups = dups +1;
                    end
                end
            end
        end

        var = scalesel(intChrom, invar);
                %Select the scales for each independent variables
        [press, cumpress,fit, rmsec,cvpred] = fitness(var,MRD(:,:));
                %Evaluate the fitness of each individual in the population
        [sfit, sortChrom, spredict] = sortfit(intChrom, fit, cvpred);
                %Sort the Chromosomes according to ascending fitness

        %Check to see if maxgen has been met
        gen = gen +1;
        if gen >= maxgen
            dups = maxdups;
        end

        %Insert the record of the fitness value.
        bfit = sfit(1);
        mfit = mean(sfit(:));
        bestfit(gen,rep) = bfit(:);
        meanfit(gen,rep) = mfit;
        bestind(gen,:,rep) = sortChrom(1,:);
        predict(:,gen, rep) = spredict(:,1);
        %display the fitness values
        disp(sprintf('The best fitness in generation %g for replication %g is %g', gen, rep, bfit));
        disp(sprintf('The average fitness in generation %g for replication %g of is %g', gen, rep,
mfit));

 % Figure of max/min/average fitness value for each generation
```

```
        figure(rep)
        subplot(2,1,1)
        plot(1:npop, fit,'og');
        xlabel('Number of Individual');
        ylabel('Fitness');
        title(sprintf('%g Replication', rep));

        subplot(2,1,2)
        plot(1:gen, bestfit(1:gen,rep), 1:gen, meanfit(1:gen,rep));
        xlabel('Generation');
        ylabel('Average and Best Fitness');
        title(sprintf('Evolution of Average and Best Fitness'));

        %Select the chromosomes
        parChrom = selectchrom(sortChrom);

        %perform crossover
        ChxChrom = xover2point(parChrom, crossover);

        %perform mutation
          MutChrom = mut2(ChxChrom, nscale);

        %selected scales
        intChrom = shuffle(MutChrom);

    end

%Write the best fitness and best individuals into Excel files.
xlswrite('BestfitMRD5SeasonMar11nigh', bestfit(:,rep), rep);
xlswrite ('BestindMRD5SeasonMar11nigh',bestind(:,:,rep),rep);
xlswrite ('PredictMRD5SeasonMar11nigh',predict(:,:,rep)',rep);
end


%------------------------------------------------------------------
% CRTRIP.M        (CReaTe an initial (Real-value) Integer Population)
%
% This function creates a population of given size of random real-values
% (integer).
%
% Syntax:       intChrom = crtrip(Nind, LindS, RangeV, LindC );
%
% Input parameters:
%
%    Nind      - A scalar containing the number of individuals in the new population.
```

```
%
%    LindS     - A scalar containing the number of variables
%
%    RangeV    - A vector of size 2 describing the boundaries of each variable.
%              Range = [LowerBoundary UpperBoundary]
%    LindC     - A scalar containing the number of production of variables
%
% Output parameter:
%    intChrom     - A matrix containing the random valued individuals of the new population
of size Nind by Lind.
%
% Author:     Lingyuan Yang
% Date:       Dec. 03, 2007

function intChrom = crtrip(Nind, LindS, RangeV, LindC)

% Check parameter consistency
  if nargin >= 1, [mN, nN] = size(Nind) ; end
  if nargin >= 2, [mL, nL] = size(LindS) ; end
  if nargin == 3, [mB, nB] = size(RangeV) ; end

  if (mN ~= 1 & nN ~= 1), error('Nind has to be a scalar'); end
  if (mL ~= 1 & nL ~= 1), error('LindS has to be a scalar'); end

% Compute Range and Lower value of variables
  Range = (RangeV(2)-RangeV(1));
  Lower = RangeV(1);

% Create initial population
% Each row contains one individual, the values of each variable uniformly
% distributed between lower and upper bound (given by Range)
  intChrom1 = floor(Range * rand(Nind,LindS)) + Lower;
  intChrom2 = floor(rand(Nind,LindC)*2);
  intChrom = [intChrom1 intChrom2];

% End of function


%-----------------------------------------------------------------------
%SCALESEL.M       (SCALE SELection)
%
% This function selects the variable with scales in the initial population

%Author:  Lingyuan Yang
%Date:    01.25.2008
```

```
function [var] = scalesel(inChrom,invar)
%Select the variable with scales in the initialized population
[npoint, novar, nscale] = size(invar);
[nopop,nvar] = size(inChrom);

indapt = inChrom(:,1);
indcur = inChrom(:,2);
indelv = inChrom(:,3);
indslp = inChrom(:,4);
apt = invar(:, indapt,1);
cur= invar(:, indcur,2);
elv = invar(:, indelv,3);
slp = invar(:, indslp,4);

%
apt2 = apt.^2;
cur2 = cur.^2;
elv2 = elv.^2;
slp2 = slp.^2;
aptcur = apt.*cur;
aptelv = apt.*elv;
aptslp = apt.*slp;
curelv = cur.*elv;
curslp = cur.*slp;
elvslp = elv.*slp;

%Organize the variables into a data set
var = zeros(npoint, nvar-4, nopop);
for i = 1:nopop
   var(:,1,i) = apt(:,i);
   var(:,2,i) = cur(:,i);
   var(:,3,i) = elv(:,i);
   var(:,4,i) = slp(:,i);
   var(:,5,i) = apt2(:,i);
   var(:,6,i) = cur2(:,i);
   var(:,7,i) = elv2(:,i);
   var(:,8,i) = slp2(:,i);
   var(:,9,i) = aptcur(:,i);
   var(:,10,i) = aptelv(:,i);
   var(:,11,i) = aptslp(:,i);
   var(:,12,i) = curelv(:,i);
   var(:,13,i) = curslp(:,i);
   var(:,14,i) = elvslp(:,i);
end
```

```
%
for i = 1:nopop;
   [mind, nind] = find(inChrom(i,:) ==0);
   var(:,nind-4,i) = 0;
end

%End of function


%-----------------------------------------------------------------------
% FITNESS.M       (Evaluate the FITNESS of each individual in the population)
%
% Syntax:      [fit, press, cumpress, rmsecv, rmsec, cvpredict]= fitness(popx, y)
%
% Input parameters:
%
%    popx      - A matrix containing the independent variables for the regression model
%
%    y        - A vector containing the dependent variables for the regression model
%
% Output parameter:
%    fit       - A vector containing the fitness value for each individual in the population
%
%    press     - predictive residual error sum of squares PRESS for each subset
%
%    cumpress  - cumulative PRESS
%
%    rmsecv    - root mean square error of cross-validation
%
%    rmsec     - root mean square error of calibration
%
%    cvpred    - cross-validation y-predictions
%
% Author:    Lingyuan Yang
% Date:      Dec. 30, 2007

function [pressO, cumpressO, fit, rmsecO, cvpredO]= fitness(popx,y)

clear m; clear n;

[npoint, nvar, nopop]=size(popx); %size of input independent variables

[m n] = size (y); %size of input dependent variables
if n ~=1
```

```
    error ('There are more than one dependent variables')
end

%create output data files
pressO = zeros(m, nopop);
cumpressO = zeros(nopop, 1);
fit = zeros(nopop, 1);
rmsecO = zeros(nopop, 1);
cvpredO = zeros(m, nopop);

%Select individuals from initialized population based upon fitness
clear i;
for i = 1:nopop
    [iind jind] = find(y(:,1)~=0);
    y = y(iind, 1);
    popx0 = popx(iind, :, i);
    [mind nind] = find(popx0(1,:) ~= 0);
    popx1 = popx0(:, nind);
    if ~isempty(popx1)
        pre = {preprocess('default'),preprocess('default')};
        opts.preprocessing = pre;
        [press, cumpress, rmsecv, rmsec, cvpred, reg] = crossval(popx1, y, 'mlr', {'loo'}, opts);
        fit(i, 1) = rmsecv;
        pressO(iind, i) = press;
        cumpressO(i, 1) = cumpress;
        rmsecO(i, 1) = rmsec;
        cvpredO(iind, i) = cvpred;
    else
        fit(i,1) = inf;
    end
end

%End of function


%------------------------------------------------------------------
%SORTFIT.M (Sort the fitness and initialized chromosome in ascending order)

% Author:     Lingyuan Yang
% Date:       Jan. 02, 2008

function [sfit, sortChrom, spredict] = sortfit(inChrom, fit, cvpred)

%sort intChrom in ascending order based upon fitness
[sfit,ind] = sort(fit(:));
```

```
sortChrom(:,:) = inChrom(ind,:);
spredict(:,:) = cvpred(:, ind);

%End of function


%-------------------------------------------------------------------
%SELECTCHROM.M (Discard the second half of the chromosome in the population, and
duplicate the first half)
% Author:    Lingyuan Yang
% Date:      Jan. 02, 2008

function [outChrom] = selectchrom(inChrom)

[nopop, nvar] = size(inChrom);
outChrom = zeros(nopop, nvar);
outChrom(1:nopop/2, :) = shuffle(inChrom(1:nopop/2, :));
outChrom(nopop/2+1:nopop, :) = outChrom(1:nopop/2, :);

%End of function


%-------------------------------------------------------------------
%XOVER2POINT.M   (Double-point crossover)
% Author:    Lingyuan Yang
% Date:      Jan. 02, 2008

function [outChrom] = xover2point(inChrom, cross)

[nopop, nvar] = size(inChrom);
outChrom = inChrom;
for i = 1:nopop/4
   otp = 0;
   tp = otp;
   for j = 1:cross
      %select twist point at random, while guaranteeing that offspring
      % are distinct from parents
      while (tp == otp)
         tp = ceil (rand*nvar);
      end
      otp = tp;
      %Twist pairs and replace
      chrom1 = (nopop/2)+(i*2)-1;
      chrom2 = (nopop/2)+(i*2);
      chrom1rep = [inChrom(chrom1, 1:tp) inChrom(chrom2,tp+1:nvar)];
      chrom2rep = [inChrom(chrom2, 1:tp) inChrom(chrom1,tp+1:nvar)];
```

```
        outChrom(chrom1,:) = chrom1rep;
        outChrom(chrom2,:) = chrom2rep;
    end
end

%End of function

%-----------------------------------------------------------------------
%MUT2 (Mutation)
% Author:    Lingyuan Yang
% Date:      Oct. 20, 2009

function [outChrom] = mut2(inChrom, varrange)

[nopop, nvar] = size(inChrom);
pop = inChrom;

for i = (nopop/2)+1:nopop
    mutrate = ceil(rand*3);
        clear num1; clear num2; clear num3;

        if mutrate == 1
            num = ceil(rand*4);
            pop(i,num) = ceil(rand(1,1)*varrange);
            else if mutrate == 2
                    num = 4+ceil(rand*14);
                    oldval = pop(i,num);
                    pop(i,num) = ~oldval;
                else if mutrate == 3
                        num1 = ceil(rand*4);
                        num2 = 4+ceil(rand*14);
                        oldval = pop(i,num2);
                        pop(i, num1) = ceil(rand(1,1)*varrange);
                        pop(i, num2) = ~oldval;
                    end
                end
        end
end
outChrom = pop;

%End of function
```