THESIS

SPATIAL ANALYSIS OF SOIL MOISTURE AT THE CATCHMENT SCALE WITH APPLICATIONS FOR ESTIMATION AND INTERPOLATION

Submitted by
Mark A. Perry
Department of Civil Engineering

In partial fulfillment of the requirements
For the Degree of Master of Science
Colorado State University
Fort Collins, Colorado
Summer 2006

COLORADO STATE UNIVERSITY

May 9, 2006

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY MARK A. PERRY ENTITLED SPATIAL ANALYSIS OF SOIL MOISTURE AT THE CATCHMENT SCALE WITH APPLICATIONS FOR ESTIMATION AND INTERPOLATION BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE.

Committee on Graduate Work

TIMOTHY R. GREEN

FREEMAN M SMITH

Advisor

Jeffrey D Niemann

Department Head

LUIS GARCIA

ABSTRACT OF THESIS

SPATIAL ANALYSIS OF SOIL MOISTURE AT THE CATCHMENT SCALE WITH APPLICATIONS FOR ESTIMATION AND INTERPOLATION

The spatial distribution of soil moisture is important to numerous applications in hydrology, agriculture, ecology and climatology. Soil moisture is a state variable for many physical processes in these fields such as infiltration and transpiration. Because these processes often have non-linear relationships with soil moisture, they depend on the spatial variation of soil moisture. The spatial variation of soil moisture can be complex because it can change through time. The goal of this thesis is to characterize the time varying properties of soil moisture patterns, from which to develop improved soil moisture estimation and interpolation methods. Here, soil moisture patterns are studied using Empirical Orthogonal Function (EOF) analysis. EOF analysis decomposes space-time variability into a series of time-invariant spatial patterns (EOFs) and spatially-invariant time series called expansion coefficients (ECs). This method is applied to soil moisture data from the 10.5 ha Tarrawarra catchment in Australia. High-resolution soil moisture patterns are available for 13 days, spanning 14 months. The analysis shows that three EOFs explain 70% of the dataset variation. Connections are drawn between these EOFs and hydrologic processes that affect soil moisture. In particular, the most important EOF (EOF1) is most highly correlated with the topographic wetness index, which is conceptually related to steady lateral flow. The second most important EOF (EOF2) is most highly correlated with the potential solar radiation index, which is related

to evapotranspiration. The third most important (EOF3) is most highly correlated to elevation and is related to the seasonal wetting-up and drying-down of the catchment.

The EOFs and ECs are used for the purposes of estimation and interpolation. In the estimation problem, an estimate of the soil moisture pattern is desired for a time when only the spatial average soil moisture is known. It is assumed that the site's EOFs can be derived from fine resolution soil moisture data collected in a previous, short field campaign. The EC values are estimated from empirical relationships with the average soil moisture. Here, only ECs 1 and 2 are considered to be predictable through time, but this may be due to the limited temporal size of the dataset. An estimated soil moisture pattern is constructed form the average soil moisture, the observed EOFs 1 and 2, and the estimated ECs 1 and 2. This EOF-based estimation method is shown to outperform other available methods. Likewise in the interpolation problem, soil moisture patterns are observed only at a coarse scale and a high resolution pattern is desired. In this case, the ECs from the coarse data are used directly, and the EOFs from the coarse data are interpolated to a higher resolution using either a distance-based method or multiple linear regression with topographic attributes. For spatial interpolation the number of useful EOFs is shown to vary with the coarse data spacing, but up to 4 EOFs are useful here. The EOF-based soil moisture interpolation provides better estimates of the fine-scale soil moisture patterns than direct soil moisture interpolation because the EOFs exhibit more consistent spatial behavior than measured soil moisture.

This study shows that EOFs exhibit stronger topographic dependence than soil moisture, because important variation at Tarrawarra is related to topography and is partitioned into low order EOFs. Less important sources of variation and random noise

are partitioned into high order EOFs. Low order EOFs are shown to exhibit distinct and higher linear correlations with common topographic attributes than soil moisture itself. Likewise in a geostatistical analysis, low order EOFs are shown to exhibit distinct and more consistent variogram functions than soil moisture. Previous studies have noted the difficulty of quantifying the time-varying relationship between dynamic soil moisture patterns and static topography. This study shows that time-invariant EOF patterns exhibit time-stable relationships to topography. The time-varying nature of the soil moisture-topography relationship can be quantified by the associated ECs.

Finally, this thesis presents opportunities for future research. The Tarrawarra catchment has a strong seasonal climate, as well as spatially uniform soils and vegetation. Future studies should apply similar EOF analysis to sites without seasonal variation, with non-uniform vegetation and with non-uniform soils. In addition, analysis of the temporal behavior of ECs was limited here due to the dataset's small temporal dimension. Unfortunately, there is a scarcity of soil moisture datasets with large space and time dimensions. One possible solution is computer simulation of soil moisture data. Simulation of large amounts of soil moisture data could allow better characterization of ECs. Based on results here, it is anticipated that ECs will exhibit more certain temporal behavior than soil moisture. This should allow better soil moisture forecasting when time-series modeling is done on ECs instead of on soil moisture itself.

Mark A. Perry
Civil Engineering Department
Colorado State University
Fort Collins, CO 80523
Summer 2006

ACKNOWLEDGMENTS

The financial support of the U.S. Army Research Office Terrestrial Sciences Program is gratefully acknowledged. Jeff Niemann provided invaluable guidance, ideas and enthusiasm for this research. Tim Green offered insights and many helpful suggestions based on his expertise and his previous work at the Tarrawarra catchment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

The general topic of this thesis is the spatial analysis of soil moisture at the catchment-scale. Understanding soil moisture spatial variation is important for hydrology, agriculture, ecology, and climatology. In hydrology, it has been shown that patterns of soil moisture in a catchment can result in very different hydrologic responses to rainfall, even if other statistical properties like the mean, variance and spatial autocorrelation are the same. For example, in a four-season climate, average soil moisture may be similar during spring and autumn, but the spatial distribution of wet and dry areas may be quite different as the landscape dries down or wets up [*Western et al.*, 1999a]. At the scale of hillslopes and drainage gullies, it has been observed that if wet soil areas are connected to a catchment outlet along flow lines there is a greater potential for surface runoff than if the same wet areas are distributed randomly due to decreased opportunity for run-on infiltration [*Western et al.*, 1999b]. In agriculture, crop yield can exhibit a complex response to the variability of soil moisture. Crop yield depends nonlinearly on soil moisture. In particular, at very high or very low soil moisture values, crop yield effectively shuts down [*Jaynes et al.*, 2003]. At the scale of ridges, hillslopes and local depressions, Jaynes et al. [2003] showed that crop yield changes based on the distribution of soil moisture, not just its spatial average. In ecology, forest stand heterogeneity was related to soil moisture, topography and temperature differences within a 21.65 km^2 forest [*Moore et al.*, 1993]. In climatology, the spatial distribution of soil

moisture affects Earth's energy balance. The distribution of moisture relative to that of solar radiation determines the partitioning of sensible versus latent energy flux [*Bras*, 1999]. At regional scales this partitioning has implications for climate, especially where atmospheric moisture is derived locally from land surface-atmosphere interactions (e.g., evapotranspiration) [*Eltahir and Bras*, 1994].

In this study, a high-resolution spatiotemporal soil moisture dataset is analyzed statistically in order to better understand its organization. The data are ground-based, point soil moisture measurements collected on 13 sampling dates, spanning 14 months, at the 10.5 ha Tarrawarra catchment in southeast Australia [*Western and Grayson*, 1998]. The catchment is pasture land, has significant topographic relief and duplex soils, where a relatively high permeability A horizon is underlain by a low permeability B horizon. The Tarrawarra has a seasonal climate that results in distinct wet and dry periods. During wet periods, spatial distribution of soil water is driven by the topographic gradient and water is distributed laterally in the A horizon [*Grayson et al.*, 1997]. During dry periods, it has been observed that lateral redistribution decreases and the spatial distribution of soil moisture is controlled mainly by the spatial variation in evapotranspiration, which depends on soil properties, vegetation, and topographic slope and aspect [*Grayson et al.*, 1997; *Western et al.*, 1999a; *Wilson et al.*, 2005]. The interaction of these and other processes result in complex soil moisture patterns. These patterns have proved difficult to characterize and even more difficult to estimate and interpolate due to their shifting nature relative to static site properties.

Herein, empirical orthogonal function (EOF) analysis is applied to this problem. EOF analysis has been used extensively in meteorology to characterize complex space-

time datasets [*Buell*, 1978; *Biau et al.*, 1999; *Liu*, 2003]. More recently it has been applied to soil moisture, but mostly at large scales where more time-stable patterns of soil moisture might be expected [*Kim and Barros*, 2002b; *Yoo and Kim*, 2004]. In this study EOF analysis is applied at the catchment scale where soil moisture variation has been observed to be highly dynamic. In general EOF analysis uses statistical covariance to identify common underlying patterns in complex data. The underlying spatial patterns (EOFs) are time-invariant (i.e., they occur on all sampling dates), but vary in their importance according to time-varying weights (expansion coefficients or ECs).

The body of this thesis consists of two distinct but related papers entitled "Analysis and Estimation of Soil Moisture at the Catchment Scale Using EOFs" (Chapter 2) and "Improved Spatial Interpolation of Soil Moisture at the Catchment Scale Using EOFs" (Chapter 3). The union of these two papers results in some overlap; however, in general they are very complimentary and share the overarching goal of better understanding and prediction of soil moisture spatiotemporal variation. In Chapter 2, patterns of soil moisture spatial variation are analyzed first, and common underlying patterns are identified. It is shown that just two spatial patterns describe almost 65% of observed variation over the 13 sampling dates. The relative importance of these two patterns varies through time to produce a surprising complexity of soil moisture distributions on different sampling dates. The two primary patterns, EOF1 and EOF2, are analyzed in detail and are interpreted in terms of their possible physical origins. In Chapter 2, EOFs 1 and 2 are analyzed enough to identify them as the wet and dry season modes of variation (associated with lateral drainage and ET) previously observed at

Tarrawarra [*Grayson et al.*, 1997]. This issue is revisited in Chapter 3 where physical significance of the EOFs is analyzed more rigorously.

Chapter 2 also exchanges the space and time axes in the EOF analysis to identify space-invariant patterns of temporal variation. It is observed that the vast majority of temporal variation (94%) is associated with a spatially uniform seasonal response. More specifically, the entire catchment wets up and dries down together. Overall, temporal variation is much larger than spatial variation. However, given its single, dominant modality, temporal variation is fairly easy to characterize. This analysis shows that the more significant problem with soil moisture at the catchment scale is characterizing its spatial distribution. For example, given a change in the catchment average soil moisture, what does the spatial distribution of soil moisture look like?

After the temporal EOF analysis, Chapter 2 returns to the spatial EOF analysis in order to develop an improved method of soil moisture estimation. For this method, it is assumed that a high-resolution soil moisture dataset like Tarrawarra is available over a limited time period for the study area. From this dataset, EOFs (time-invariant spatial patterns) and ECs (time-varying weights) are identified. Further study of time-weights (ECs) shows them to vary seasonally, as the spatial average soil moisture in the catchment changes. Each EC can be estimated using a cosine function of the spatial average, where the functions for EC1 and EC2 (weights for EOF1 and EOF2, respectively) show distinct phase-shifts, reflecting the alternating seasonal role of each EOF. With such functions defined, the importance of each EOF pattern can be estimated at an unobserved time given only the spatial average soil moisture for that time. A high-resolution soil moisture spatial pattern is then estimated for that time as a linear

combination of the observed EOFs, weighted by the estimated ECs. The remaining question is how to estimate the spatial average soil moisture at the unobserved time. Various methods of estimating spatial average soil moisture are discussed, including remote sensing (in which case the estimation method here could be described as downscaling), a simple water balance model (where the estimation might be referred to as disaggregation), or strategic ground-based monitoring to track the catchment average through time. Ground-based tracking of the spatial average has been discussed in literature [*Grayson and Western*, 1998]; however, it is developed further in Chapter 2 herein. It is demonstrated that given a space-time soil moisture dataset, locations can be identified that have a high linear correlation to the spatial average through time. At each such location, a regression relationship can be established between point soil moisture and the spatial average. In the future, such a location can be monitored and the regression relationship used to estimate the catchment average soil moisture. Further, it is shown that a similar approach can be taken to track the EC weights directly. Finally in Chapter 2, the proposed soil moisture estimation method is compared to several available methods from the literature. The method proposed herein is shown to outperform the others. It is proposed that the EOF characterization of soil moisture not only produces better estimates but also produces those estimates based on an explicit consideration of the underlying patterns of spatial variation.

In Chapter 3, the focus is on the decomposition of soil moisture for the purpose of interpolating soil moisture patterns from widely-spaced observations. Results from Chapter 2 indicated that soil moisture can be decomposed roughly according to underlying hydrological processes. In Chapter 3, it is hypothesized that the decomposed

variables exhibit distinct spatial structures associated with these processes. For example, we expect the EOF associated with lateral drainage to have a different relationship to topography than the EOF associated with evapotranspiration. In fact, it is shown that several of the EOFs exhibit distinct spatial behavior – distinct from each other and from the original soil moisture data from which they are derived. This is shown first with a thorough analysis of the relationships of EOFs and soil moisture patterns to topographic attributes. The EOFs are shown to exhibit different and more consistent dependence on topography than the soil moisture patterns themselves. This result has far-reaching implications because a large amount of effort in hydrology has been dedicated to relating soil moisture variation to topography. Next, a simple geostatistical analysis is performed on the EOF and soil moisture patterns. As with their relationships to topography, the EOFs are shown to exhibit spatial structures that are distinct from and more consistent than those of soil moisture. A connection is made between the spatial structure of EOFs and that of their associated hydrological processes. Higher order EOFs are shown to have similar spatial structure to noise, e.g., measurement error. This too has far reaching implications. It suggests that standard geostatistical interpolation techniques may perform better if soil moisture data are first decomposed using EOF analysis.

Next an EOF-based method of interpolation is developed and compared to analogous methods applied directly to soil moisture. In the EOF-based methods, we start with sparse soil moisture measurements from which ECs and sparse EOFs are computed. The sparse EOFs are interpolated over the entire spatial domain using both a multiple linear regression against topographic attributes and a simple inverse distance weighted interpolation method (note that although the former is technically an estimation method,

both are loosely referred to as interpolation methods here, because they are used to fill in sparse spatial observations). Next, high-resolution soil moisture is estimated over the entire domain using the observed ECs and the interpolated EOFs. It is shown that this method of soil moisture interpolation outperforms direct interpolation of soil moisture, because EOFs exhibit more consistent spatial behavior and thus are more predictable than measured soil moisture. The results of Chapter 3 are important in that a large problem with soil moisture is the cost and difficulty of measuring it. While the estimation method proposed in Chapter 2 and other comparable methods in the literature perform well, they have the overarching limitation of being extremely data intensive. In most circumstances outside of research, high-resolution soil moisture data will rarely be available. In contrast, the interpolation method in Chapter 3 has light data requirements. No high-resolution soil moisture data are required.

Finally, Chapter 4 summarizes the implications of this study and identifies major directions for further research.

2 Analysis and Estimation of Soil Moisture at the Catchment Scale Using EOFs

Abstract

Soil moisture patterns and dynamics are important for numerous applications such as flood forecasting, climate modeling, and management of agricultural lands. Unfortunately, widespread observations of soil moisture are not currently available at the spatial scale of most of these applications. Given these data limitations and the complexity of soil moisture dynamics, there is a need to gain a better understanding of soil moisture patterns and to develop methods that can efficiently estimate these patterns from limited observations. In this paper, we use Empirical Orthogonal Function (EOF) analysis to study the Tarrawarra soil moisture dataset from Australia. EOF analysis partitions the observed variation into a series of time-invariant spatial patterns (EOFs) that can be multiplied by temporally varying (but spatially constant) coefficients and summed to reconstruct observed soil moisture patterns. Using this approach, we identify two important spatial patterns underlying soil moisture variability at Tarrawarra, which supports previous contentions that the spatial patterns are controlled by local soil properties in wet and dry conditions and topographic characteristics during intermediate conditions. We also use the EOF analysis to identify points in the catchment whose

temporal variability is most representative of that of each underlying spatial pattern and thus can be used to monitor the distinct modes of variability. Finally, we show that the EOF approach can be used to estimate soil moisture patterns for unobserved times if a field campaign has collected detailed soil moisture observations for a limited time period in the past.

2.1 Introduction

Soil moisture is a fundamental state variable in hydrology, meteorology, and agriculture, and plays an important role in numerous applications in these fields. It is well known, for example, that soil moisture affects runoff production. Western et al. [1999b] showed that the spatiotemporal distribution of soil moisture has a strong impact on the response of stream discharge to rainfall events, and they found that poor representations of soil moisture patterns in hydrologic models can produce significant errors in the estimated runoff. Kitanidis and Bras [1980] showed that wet soil moisture conditions play a significant role in producing floods. Soil moisture also affects plant growth and evapotranspiration, and for locations where precipitation is largely derived from local evapotranspiration, soil moisture conditions can promote the persistence of droughts and wet periods as well [*Eltahir and Bras*, 1994]. Delworth and Manabe [1988] emphasized the need for good quality spatiotemporal soil moisture data for modeling of the Earth's climate, and Liu [2003] showed that spatial patterns of soil moisture are linked to precipitation patterns in a regional climate model of southeast Asia. Soil moisture also affects erosion from overland flow and the formation of gullies [*Zaslavsky and Sinai*, 1981; *Moore et al.*, 1988]. In agriculture, Jaynes et al. [2003] showed that

spatiotemporal variation in soil moisture was reflected in crop yield. For example, low-lying areas of an Iowa field had higher than average corn yields in dry years due to their relatively abundant soil moisture, whereas these same areas produced lower than average yields in wet years because they were too wet. Green and Erskine [2004] observed similarities in the spatial autocorrelation of crop yield and early season soil moisture. They interpreted this as showing the importance of soil moisture for seed germination for dryland farming.

The spatiotemporal patterns of soil moisture are quite complex because they depend on multiple physical processes that can produce very distinct patterns. In some cases, soil moisture is locally controlled [*Vachaud et al.*, 1985; *Western et al.*, 1999a], meaning that the soil moisture at a location is primarily controlled by properties at the same location. Locally controlled soil moisture can be caused by strong spatial variations in soil properties or by the dominant role of vertical fluxes like evapotranspiration and infiltration in determining the soil moisture. Vachaud et al. [1985] identified temporally persistent soil moisture patterns at the field scale, and Kachanoski and De Jong [1988] argued that these soil moisture patterns were locally controlled. In other cases, soil moisture is horizontally redistributed, which introduces non-local dependencies. Computing spatial autocorrelation, Kachanoski and De Jong [1988] demonstrated that soil moisture at a point can interact with the soil moisture at neighboring locations, which produces more complex temporal behavior. Burt and Butcher [1985] observed lateral redistribution of soil water on a 1.4 ha hillslope via saturated flow above shallow impermeable bedrock. They documented the development of saturated areas at downhill, low slope, and convergent locations, which implies accumulation of flow in these areas.

Soil moisture can move horizontally even where soils are deep and unsaturated. Zaslavsky and Sinai [1981] argued that lateral movement of soil water is expected from anisotropic permeabilities caused by soil layering. Under certain conditions, they showed that the horizontal flux is proportional to the vertical flux, the ground slope, and the degree of anisotropy. This leads to accumulation of soil moisture in areas of concave-up curvature.

The relative roles of vertical and horizontal fluxes can vary in time at a given location as well. Burt and Butcher [1985] calculated the correlation between soil moisture and the wetness index (WI). WI is defined as the natural log of the ratio of the land area that drains through a unit length of contour to the local slope and is therefore associated with horizontal fluxes [*Beven and Kirkby*, 1979]. Burt and Butcher [1985] found that the correlation varied with time and was generally better during wet conditions. For the small Tarrawarra catchment in southeastern Australia, Western et al. [1999a] showed that correlations between soil moisture and topographic characteristics were lowest for dry conditions, rose for moderately wet conditions, and decreased again for very wet conditions. For moderately wet conditions they found that soil moisture correlates best with the natural log of drainage area. They suggested that the topographic influence on soil moisture during moderately wet conditions was mainly due to the lateral redistribution of water in the soil. For dry austral summer dates, they found that soil moisture usually correlated best with potential solar radiation index (PSRI) [*Moore et al.*, 1993], which describes topographic aspect in comparison with the position of the sun. Here, Western et al. [1999a] suggested that topographic influence on soil moisture was mainly due to evapotranspiration. Florinsky et al. [2002] also found seasonally-varying

correlations between topographic characteristics and soil moisture at a site in the Canadian prairie. However, for this low relief site, topographic characteristics had a weaker relation with the soil moisture patterns.

As with topographic dependence, the spatial structure of soil moisture can vary due to the time-varying importance of multiple hydrologic processes acting at different spatial scales. Seyfried and Wilcox [1995] discussed soil moisture variation from the spatial scale of shrub "islands" in a sparsely vegetated landscape to a larger scale related to topographic and geological variations. They illustrated how wet areas associated with vegetation at small scales are nested in and subsumed by larger scale dry areas associated with certain underlying geologic formations. Green and Erskine [2004] found that soil moisture at the field scale in a semi-arid climate lacked a stationary correlation length, which is consistent with scaling-invariant behavior. However, others have found that soil moisture had a clear variogram sill [*Western et al.*, 1998]. Western et al. [1998] computed soil moisture variograms for different sampling dates at Tarrawarra and showed generally longer correlations lengths on dry dates than on wet dates. They argued that the longer correlation lengths are related to the scale of evapotranspiration, the dominant process affecting soil moisture during the dry periods. In contrast, the shorter correlation lengths during the wet periods are related to the smaller spatial scale of lateral redistribution. Western et al. [1999a] computed variograms for the residual errors of soil moisture estimates based on the WI and the PSRI. They showed that PSRI explains variation occurring at larger scales (>80 m), while WI explains variation at smaller scales (>10 m, where 10m was equal to the sample spacing). Likewise, comparison of two studies in Oklahoma reveals different controls acting at different

scales. Yoo and Kim [2004] analyzed space-time patterns of soil moisture for two sites in the Little Washita experimental watershed. They found that a single spatial pattern can explain more than 70% of the observed soil moisture variation. This pattern is correlated with topographic attributes. However, Jawson and Niemann (preprint, 2005) analyzed soil moisture patterns for the same location at a much larger spatial scale using remotely-sensed data from the Southern Great Plains 1997 field campaign [*Jackson et al.*, 1999]. They found that a single pattern could explain over 60% of the spatial variation, but this pattern was most highly correlated with soil properties (percent sand and percent clay). Kachanoski and De Jong [1988] related the scale of soil moisture variation to the interaction of controlling hydrologic processes acting at different scales. They showed that the spectral densities of soil moisture on two dry dates were highly correlated across all observed scales. However, when a dry date was compared to a wet date, on which lateral redistribution was thought to control soil moisture, the spectral densities were not correlated below 40 m. Kachanoski and De Jong [1988] concluded that processes act at different scales, with some processes acting locally and others acting over large distances. They said that large scale processes can be modified by local processes, producing nested patterns of variation.

Ultimately, the complex nature of soil moisture patterns makes them difficult to estimate from sparse observations. For example, both Western et al. [1999a] and Florinsky et al. [2002] stated that the time varying correlations they observed between soil moisture values and topographic attributes limit the applicability of traditional regression-based approaches in estimating soil moisture. Western et al. [1999a] emphasized that both lateral redistribution and evapotranspiration are important for

determining soil moisture patterns and that there is a need to develop estimation methods that account for both processes and their time-varying relative importance. They proposed using a weighted combination of topographic attributes where the weights are based on the average soil moisture at a particular time. Wilson et al. [2005] developed and applied such a technique using a dynamic multiple linear regression. They proposed their method as an improvement to Topmodel [*Beven and Kirkby*, 1979], because Topmodel uses only WI as a predictor variable and assumes that it has a time invariant and perfect correlation with soil moisture.

Based on this brief literature review, there is clearly a need to gain a better understanding of soil moisture patterns and to develop methods that can estimate these patterns from limited observations. In particular, a method is needed to identify distinct underlying patterns of spatial variation in soil moisture and to estimate soil moisture in a way that accounts for these patterns. In this paper, we propose Empirical Orthogonal Function (EOF) analysis as an effective way to characterize and estimate soil moisture patterns in a catchment where the soil moisture exhibits dynamic statistical properties. EOF analysis is commonly used in the meteorological field to characterize spatiotemporal variables [*Preisendorfer*, 1988; *Liu*, 2003]. It has also been used by Yoo and Kim [2004] to characterize soil moisture for agricultural sites and by Jawson and Niemann (preprint, 2005) and Kim and Barros [2002b] to characterize soil moisture at large spatial scales. Here, we apply EOF analysis to soil moisture data from the Tarrawarra catchment, which has been widely studied [*Western et al.*, 1998; *Western and Grayson*, 1998; *Western et al.*, 1999a; *Western et al.*, 1999b; *Western et al.*, 2001; *Wilson et al.*, 2005] and found not to be time-stable in the traditional sense [*Vachaud et al.*, 1985;

Grayson and Western, 1998], that is a single temporally persistent pattern is not observed. The EOF approach allows identification of persistent underlying patterns of soil moisture variation, which has been a difficult problem in the past. We also develop a soil moisture estimation method based on EOF analysis. This method is data-based and relies on the availability of historical soil moisture observations to forecast soil moisture patterns for unobserved times. It does not require one to identify predictor variables such as WI or PSRI. The method is an extension of a technique developed by Rao and Hsieh [1991] for forecasting streamflow hydrographs.

In the following section (Section 2.2), we describe the Tarrawarra catchment and available soil moisture data. In Section 2.3, we describe the EOF analyses that are performed on the spatial and temporal anomalies of soil moisture, and in Section 2.4, we use the EOF approach to estimate soil moisture patterns for times when we know only the spatial average soil moisture. In Section 2.5, we identify locations where the spatial average soil moisture and important EOF variables can be monitored. Finally, in Section 2.6, we close the chapter with our main conclusions.

2.2 Field Site and Data

The soil moisture dataset used here is from the 10.5 ha Tarrawarra catchment in southern Victoria, Australia (37°39' south, 145°26' east). The data were originally collected and analyzed by Andrew Western and Roger Grayson [*Western and Grayson*, 1998] and are available online from the Global Soil Moisture Data Bank [*Robock et al.*, 2000]. The climate of Tarrawarra is temperate with an average annual rainfall of 820 mm and a potential evapotranspiration of 830 mm. The summer is relatively dry, and the

winter is relatively wet. The vegetation is planted pasture. Elevations from a total station survey are available for the site at approximately a 10 meter spacing. Figure 2.1 shows the catchment topography, which includes two subcatchments: an east subcatchment that comprises about 60% of the total area, and a west subcatchment that accounts for the remaining 40%. The catchment also contains two large hillslopes: a south-facing hillslope in the northern portion of the catchment and a northwest-facing hillslope in the southeastern portion of the catchment (labeled in Figure 1). The hillsides have slopes of approximately 11-14%, the valley slopes are about 4%, and the maximum relief is 27 m. Note that all figures herein show the Tarrawarra catchment such that UTM north is oriented 14° clockwise from vertical.

Soil moisture data were collected across the catchment on 13 dates during a period of about 14 months. The dates were selected to capture a normal range of moisture conditions during the year [*Western and Grayson*, 1998]. Data were obtained using a time domain reflectometry (TDR) device that was mounted on an all terrain vehicle. Each TDR measurement represents an average soil moisture over the top 30 cm of the soil profile. Throughout this paper, the soil moisture data are expressed as the volume of water per volume of soil. Soil moisture data points are spaced at approximately 10 meters along transects with approximately 20 meters between transects. Although the exact locations of the TDR measurements varied between observation dates, the circles in Figure 2.1 show a typical pattern of soil moisture sampling locations. The average soil moisture is highest during winter and lowest during summer, which is expected from the seasonal cycle of precipitation. The soil's A horizon is 20 to 35 cm

deep and is underlain by a low permeability B horizon. According to Western et al. [1999a], this results in perched water tables in the A horizon during the winter months.

The soil moisture data from Tarrawarra contain measurement errors that can be significant [*Western and Grayson*, 1998; *Western et al.*, 1999a]. Western et al. [1999a] cited an advertised standard error of +/-2% for the TDR equipment used, but their own calibration efforts showed a standard error of +/-1.7%. They note that this error is a significant part of the total soil moisture variance observed on dry sampling dates. For example, on 23 February 1996, the error variance is 58% of the observed soil moisture variance. On wet dates the instrument error is only a small part of the total soil moisture variance.

Because EOF analysis requires observations to be available at the same locations for each sampling time, the TDR data were interpolated onto a consistent grid with a similar spacing as the original data (20m column spacing by 10m row spacing) using the cubic spline method. The interpolation preserves the mean and has little effect on the soil moisture patterns, because the original observations are nearly on a grid (see Figure 2.1). However, the variance is slightly reduced. The loss in variance is expected for any smooth interpolation method and is associated with the loss of extreme soil moisture values. The spatial extent of the dataset is limited to the region that was consistently sampled on all 13 dates. For this study, 459 measurement locations are used for each of the 13 dates.

2.3 EOF Analysis of Soil Moisture

2.3.1 Method

EOF analysis is a widely applied statistical method for analyzing large multidimensional datasets. When applied to a space-time dataset, EOF analysis can be used to decompose the observed variability into a set of orthogonal spatial patterns (EOFs), which are invariant in time, and a set of time series called expansion coefficients (ECs), which are invariant in space. Together the EOFs and ECs can be used to reconstruct the variability in the original dataset. EOF analysis can be used to study modes of variability in a dataset, to identify variables that control observed patterns, and to approximate or compress a dataset [*Preisendorfer*, 1988]. In this paper, EOF analysis will be used to study both spatial and temporal variation of soil moisture. In this section, we first describe the EOF methodology for spatial variation in detail and then describe the differences when temporal variation is considered.

For the analysis of spatial variation, soil moisture data are converted to spatial anomalies. Spatial anomalies are computed by subtracting the average soil moisture for a given observation time from all observations collected at that time. If $s_i(t)$ is the soil moisture observation at location i and time t, then the spatial anomaly at the same location and time $z_i(t)$ can be calculated:

$$z_i(t) = s_i(t) - \frac{1}{m}\sum_{j=1}^{m} s_j(t) \qquad (1)$$

where j is an index of observation locations and m is the number of observation locations. By considering the spatial anomalies instead of the soil moisture, much of the temporal variation is excluded from consideration. Figure 2.2a shows the spatial anomalies of the

Tarrawarra soil moisture data. In this figure, the vertical axis is the temporal dimension and the horizontal axis is the spatial dimension. Thus, spatial patterns correspond to horizontal rows in the plot. Because we have removed the average soil moisture for each date, each row has the same average shading (i.e. the average of the spatial anomalies for each date is zero).

The first step in the EOF analysis is to calculate the covariance matrix. When considering the spatial anomalies, the relevant covariance is the spatial covariance. Mathematically, the spatial covariance $\sigma_{t,\tau}$ at two times t and τ can be written:

$$\sigma_{t,\tau} = \frac{1}{m}\sum_{i=1}^{m} z_i(t) z_i(\tau). \tag{2}$$

Equation (2) averages the products of the spatial anomalies at times t and τ across all locations. If t and τ are chosen to be the same time, the equation simplifies to the spatial variance of the anomalies at the selected time, which is why we refer to this covariance as a spatial covariance. To consider all times, Equation (2) is generalized to produce the covariance matrix. We first construct a matrix Z of the spatial anomalies. Note that capital letters are used here to denote matrices. If we have m sampling locations and n sampling times, then Z can be constructed:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ z_{m1} & \cdots & \cdots & z_{mn} \end{bmatrix}. \tag{3}$$

The spatial covariance matrix Σ can then be found:

$$\Sigma = \frac{1}{m} Z^T Z \tag{4}$$

where the superscript T indicates the matrix transpose. The dimensions of Σ are determined by the number of sampling times (it is an n by n matrix). The spatial covariance examines how the spatial anomalies are correlated between times. This characteristic is visible in Figure 2.2a. The strength of the covariance is indicated by the extent to which a shading pattern is repeated through time (vertically in the figure). Figure 2.2a makes it clear that the covariance of the spatial anomalies is quite complicated. Similar patterns can be seen for certain observation times, but these patterns disappear entirely at other times.

The next step of the EOF procedure is to diagonalize the covariance matrix. Mathematically, this is accomplished by finding its eigenvectors and eigenvalues [*Preisendorfer*, 1988]. The eigenvectors and eigenvalues of Σ satisfy the equation:

$$\Sigma E = LE \tag{5}$$

where E is an n by n matrix that contains the eigenvectors as columns:

$$E = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nn} \end{bmatrix} \tag{6}$$

and L is an n by n matrix that contains the associated eigenvalues along the diagonal:

$$L = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ 0 & l_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & l_{nn} \end{bmatrix}. \tag{7}$$

The eigenvectors in E are the weights applied to each component in Σ to diagonalize Σ. This transformation represents a rotation of the spatial anomalies in multi-dimensional space, where there is a dimension for each sampling time. The first axis identified by the

diagonalization is the direction in the multidimensional space that explains the most covariance in the spatial anomaly dataset. Each new axis is constrained to be orthogonal to the other axes and explains the most remaining covariance. These axes are new independent variables that essentially replace the sampling dates. When the covariance of the spatial anomalies is examined in terms of these new variables, the off-diagonal values are all zero. The eigenvalues contained in L represent the variance that occurs in the direction of each newly identified axis. The total variance of the spatial anomaly data is invariant with respect to the rotation, so the trace of L (i.e. the sum of the diagonal values) is equal to the total variance of the original spatial anomalies. Thus, the portion of the variance that lies along the j^{th} new axis (i.e. along the unit vector defined by the j^{th} column in E) is:

$$p_j = \frac{l_{jj}}{\sum_{k=1}^{n} l_{kk}}. \tag{8}$$

Using Equation (8) we can judge the importance of each new axis in describing the variability of the spatial anomalies by the magnitude of the associated eigenvalue.

Every spatial anomaly can be described in terms of the new variable axes. A matrix F containing the coordinates of the spatial anomalies on the new axes can be calculated by projecting the anomalies onto the rotated axes. Mathematically, this operation is simply:

$$F = ZE. \tag{9}$$

Note that the dimensions of F match the dimensions of Z (it is an m by n matrix). Each row in F corresponds to an original spatial location (same as in Z). However, the observation dates in the columns of Z have been replaced in F. While terminology is

inconsistent in the literature, each column of E can be called an EC and each column of F can be called an EOF. Because the soil moisture dataset has 13 observation times, the EOF analysis produces 13 EOF/EC pairs. Typically, the EOFs and ECs are rearranged so that the first column explains the most variance, the second column explains the second most variance, and so on. This sorting can be done by examining the associated eigenvalues in L. After this rearrangement, we refer to the first columns of F and E as EOF1 and EC1 respectively, the second columns of F and E as EOF2 and EC2, etc. Each EOF is a spatial pattern, whose values are the values of the new variable for each location. The expansion coefficients are time series that indicate how aligned the variability of each observation time is with the new variable. For example, a large positive EC value for a particular date indicates that the associated axis captures a large portion of the variance on that date. If the EC value is large on some second date, then the spatial anomalies of the first and second dates must be positively correlated. From this perspective, the associated EOF map also identifies points whose spatial anomalies are correlated between times. If we already know E and F from solving Equations (5) and (9), we can reconstruct the original spatial anomaly dataset from the following equation:

$$Z = FE^T . \tag{10}$$

One can also use EOF analysis to study temporal variation using the temporal anomalies of soil moisture. If the temporal anomaly at location i and time t is written $z'_i(t)$, then $z'_i(t)$ can be found from:

$$z'_i(t) = s_i(t) - \frac{1}{n}\sum_{\tau=1}^{n} s_i(\tau) \tag{11}$$

where τ is an index of sample dates. Here, we have removed the long-term average soil moisture at each location. Thus, by considering the temporal anomalies we are analyzing how soil moisture varies in time rather than space. Figure 2.2b shows temporal anomalies for the Tarrawarra dataset. Now every column has the same average shading because the temporal average has been removed. The strong banding observed in Figure 2.2b suggests that the temporal anomalies are highly correlated through time at different locations. When considering the temporal anomalies, the relevant covariance matrix is the temporal covariance whose dimensions are determined by the number of sampling locations (it is an m by m matrix). By considering the temporal covariance matrix, we are effectively exchanging the space and time dimensions in the EOF analysis. In this case, the eigenvectors are weights on different locations and are therefore spatial patterns. For simplicity, we refer to these spatial patterns as EOFs. When the temporal anomalies are projected onto the EOFs, the matrix F is produced which preserves the observation times but replaces the locations. We refer to the time series contained in F as ECs. Because we have 459 observation locations, the EOF analysis of the temporal anomalies produces 459 EOF/EC pairs.

2.3.2 Results for Spatial Anomalies

EOF analysis was performed on the spatial anomalies of soil moisture at Tarrawarra as described in the previous section. Figures 2.3a and 2.3b show the two EOFs that explain the most variation (EOF1 and EOF2), and Figures 2.3c and 2.3d show the ECs associated with these EOFs. For reference, Figures 2.3c and 2.3d also show the mean and variance of the soil moisture at each observation time. Using Equation (8) we

find that EOF1 explains about 55% of the dataset's spatial variance and EOF2 explains approximately 9%. Note that the percent variation explained here refers only to the spatial variation. Temporal variation is not accounted for in the spatial covariance matrix and is therefore not included in the trace of matrix *L*.

EOF1 has a strong resemblance to the catchment topography. We see large positive EOF values along the valleys and zero to negative valleys on the hillslopes. Large EOF1 values correspond to locations that have consistently large positive anomalies on dates with large positive EC1 weights. Looking at the plot of EC1 weights (Figure 2.3c), we conclude that the EOF1 spatial pattern is correlated on moderately wet days, especially 27 September 1995, 2 May 1996, and 25 October 1996. There is a clear decrease in the EC1 weights during the dry season, especially February, and during the very wet season, including July and September 1996. These results are consistent with the findings of Western et al. [1999a] that soil moisture patterns on moderately wet dates show a strong topographic influence, one that seems to diminish on both dry and very wet days. Western et al. [1999a] proposed that lateral redistribution is the dominant hydrologic flux on moderately wet days. Water is primarily moving through the shallow and relatively high permeability A horizon, where the driving hydraulic gradient is controlled by topography. The correlations between EOF1 and various topographic attributes were calculated, and EOF1 was found to be strongly correlated (correlation coefficient of 0.67) with the natural log of the land area drained by each grid cell. This suggests that EOF1 is primarily related to landscape position and the associated redistribution of water through surface and subsurface drainage. However, some dependence of EOF1 on the aspect of the hillslopes is observed. In particular, large

magnitude negative values are evident on the hillslope in the lower-right portion of the catchment in Figure 2.3a. Taken with the positive EC1 weights, these negative values suggest that these locations are dry (i.e. exhibit negative anomalies) when EOF1 is important. A statistically significant correlation between EOF1 and PSRI was also observed (correlation coefficient of -0.29). This implies that EOF1 is also related, if subordinately, to evapotranspiration. Because the EOF analysis determines these patterns by identifying points whose spatial anomalies are correlated in time, this result implies that some of the soil moisture variability induced by evapotranspiration occurs concurrently with the variability induced by lateral drainage. It is also worth noting that a large part of the catchment has EOF1 values near zero, especially the south-facing hillslope. Values near zero can occur if these locations have persistent zero anomalies on the dates involved with EOF1 or if the anomalies at these locations are uncorrelated on the dates involved with EOF1. The south-facing hillslope seems to fit the latter case as it is important in EOF2.

EOF2 is shown in Figure 2.3b and exhibits a pattern that appears to be predominantly related to the aspect of the topography. The south-facing hillslope generally has negative EOF2 values and the northwest-facing hillslope generally has positive EOF2 values. The correlations between EOF2 and various topographic attributes were calculated, and EOF2 was found to be correlated with PSRI (correlation coefficient of 0.56). Thus, the EOF2 pattern appears to be associated with the spatial distribution of evapotranspiration, with north-facing slopes receiving more incident solar radiation than south-facing slopes in the southern hemisphere. Looking at the ECs in Figures 2.3c and 2.3d, there are some days when either EOF1 or EOF2 is dominant (e.g., 27 September, 14

February, and 2 May). On other days, both EOFs are important, and EOF2 can be seen as modifying EOF1. For example, on 20 September 1996, EOF2 modifies EOF1 to make the south-facing hillslope drier and to shift the dry area on the northwest-facing slope to the southwest (negative EOF2 values multiplied by positive EC2 values result in decreased soil moisture and visa versa). It seems that EOF2 serves to modify EOF1 as the pattern of evapotranspiration changes with season and the role of evapotranspiration changes relative to lateral redistribution.

To understand the role of EOF2 in more detail, Figure 2.4 shows the spatial anomalies on 14 February and 20 September, which have the largest negative and positive EC2 values, respectively. For the south-facing hillslope, we observe positive anomalies or relative wetness on 14 February and negative anomalies or relative dryness on 20 September. This tendency is consistent with the contribution of the EOF/EC2 pair. On 14 February, the EC2 value is negative, so the negative EOF2 values on this hillslope result in positive anomalies. On 20 September, the EC2 value is positive, so the negative EOF2 values result in negative anomalies. Note that EOF1 has little influence because its values are near zero for this hillslope. The influence of EOF2 on the south-facing hillslope is also consistent with the seasonal variations in the strength of evapotranspiration. In February (during the dry season), evapotranspiration is dominant. Because the south-facing hillslope receives less solar radiation than the northwest-facing hillslope, it is expected to be relatively wet. Likewise during the wet season when evapotranspiration is less important, EOF2 acts to dry the south-facing hillslope, minimizing aspect related differences between the northwest and south-facing slopes. For the northwest-facing hillslope, the role of the EOF/EC2 pair is more complex because

EOF1 also plays a significant role. While EOF1 primarily identifies the pattern of soil moisture associated with lateral redistribution, it also identifies a dry region on the northwest-facing hillslope that occurs concurrent to the lateral distribution (during moderately wet periods). On the northwest-facing hillslope, EOF2 primarily acts to shift the location of the dry region between the wet and dry seasons. This shift may be associated with the seasonal shift in the angle of solar radiation incidence.

To help visualize the rotation that is produced by the EOF analysis, Figure 2.5a plots the spatial anomalies of soil moisture on 14 February against the spatial anomalies on 20 September (the same dates shown in the previous figure). The axes associated with EOF1 and EOF2 are shown and labeled. Overall, the anomalies for these two dates do not appear to be well correlated. However, the open circles in the plot identify points from the northwest-facing hillslope where the shifting dry patch occurs and EOF/EC2 is important. For these points, a negative correlation between the two dates can be seen, the direction of which is identified by EOF/EC2. It is important to note that the EOF axes do not necessarily appear orthogonal when viewed in this two dimensional manner. While these axes are orthogonal in the multidimensional space defined by all the sampling dates, they are being projected into a two dimensional space defined by only two sampling dates. This point is emphasized by Figure 2.5b, which plots the spatial anomalies of 13 April against those of 28 March along with the EOF1 and EOF2 axes. Here, the EOF axes are almost aligned, which indicates that the variation captured by EOF1 and EOF2 is the same and that the orthogonality exists in some other plane. As we have described above, the orthogonality exists where evapotranspiration operates

independently of lateral redistribution, for example in the plane created by very wet and very dry dates shown in Figure 2.5a.

An important question is whether the correlated spatial patterns identified by the EOF analysis are statistically significant. EOF analysis deconstructs any spatiotemporal dataset into EOFs and ECs irrespective of whether the observed correlations are significant. For example, a dataset comprised of independent identically distributed values for each sampling time would produce EOFs that each explain roughly the same amount of variation. In one statistical significance test [*North et al.*, 1982], an individual EOF can be shown to be significant if the lower confidence limit of its eigenvalue is larger than the upper confidence limit of the eigenvalue of the next most important EOF. If the confidence limits do not overlap, it indicates that the spatial covariance oriented along the respective EOF axis is significantly larger than that along other axes. North et al. [1982] developed confidence limits for the eigenvalues by assuming Gaussian random variables and a sampling standard error of order $(2/m)^{1/2}$, where m is the number of observations used to calculate the values in the covariance matrix (the number of sampling locations in the case considered here). Given these assumptions, they determined confidence limits such that the true eigenvalue falls within the confidence limits 68% of the time. For convenience, these confidence limits can also be written for the portion of the total variance explained by the j^{th} EOF (i.e. p_j). In this case, the confidence interval spans a range of p_j +/- Δp_j, where Δp_j can be calculated:

$$\Delta p_j = \frac{l_{jj}}{\sum_{k=1}^{n} l_{kk}} (2/m)^{1/2}. \qquad (12)$$

Applying Equation (12), the upper and lower confidence limits on the percent variance explained by EOF1 are 58.3% and 51.1%. For EOF2, EOF3, and EOF4, the upper and lower confidence limits are 9.7% and 8.5%, 6.3% and 5.6%, and 5.7% and 5.0%, respectively. Thus by the North et al. [1982] test, only EOF1 and EOF2 are statistically significant. Note that higher order EOFs may still be meaningful in some applications. The issue of statistical significance is considered in more detail in Chapter 3.

2.3.3 Results for Temporal Anomalies

As discussed in Section 2.3.1, the temporal anomaly EOF analysis involves removing the long-term soil moisture average at each location and then performing the EOF analysis using the covariance through time at all combinations of spatial locations. The results of this analysis are shown in Figure 2.6. Figures 2.6a shows EOF1 and Figure 2.6b shows EC1. Only one EOF/EC pair is shown in the figure because Equation (12) indicates that only the first EOF/EC pair is significant. For reference, Figure 2.6b also shows the spatial average soil moisture. Using Equation (8), the portion of the total temporal variance that is explained by each EOF was computed. EOF1 explains about 94% of the variation. The second temporal anomaly EOF explains a mere 1.3% of the variation.

EOF analysis of the temporal anomalies identifies spatial locations that tend to have wet or dry soil moisture anomalies at the same times. In particular, we know that wet and dry seasons occur during the year, but it is possible that there are other significant temporal patterns that affect portions of the catchment. However, the large percentage of the variation explained by EOF1 indicates that there is one dominant mode

of temporal variability for this dataset. If one compares the EC1 values in Figure 2.6b with the spatial average soil moisture, it is evident that this dominant mode of variation is associated with the wet and dry seasons. It is possible that a dataset with higher temporal resolution would identify other modes of temporal variability associated with individual storms. EOF1 in Figure 2.6a has positive values for all locations, which indicates that the soil moisture throughout the catchment responds in the same direction to the seasonal cycle. This is to be expected at this small spatial scale. Western and Grayson [1998] found no significant spatial variation in precipitation across the catchment that might introduce a more complex pattern. At a larger scale, Jawson and Niemann (preprint, 2005) found differences in the direction of movement of the temporal anomalies because precipitation events occurred over subsections of the study region. Although all locations in Tarrawarra respond in the same direction to the wet and dry seasons, the higher EOF1 values in valleys indicate that these areas vary more with the seasonal cycle than the hillslopes.

2.4 Estimation of Soil Moisture Patterns

2.4.1 Method

The basis for EOF estimation methods is the fact that the original dataset can be reconstructed from the EOFs and ECs. In particular, if the spatial anomalies were used in the EOF analysis, then these anomalies can be reconstructed as a product of the EOFs and ECs as shown in Equation (10). Then the reconstructed anomalies can be added to the observed spatial average for each date to reconstruct the original soil moisture data.

EOF analysis also provides an efficient method for approximating the original data. The

amount of spatial variance associated with each EOF/EC pair is known, so the less important EOF/EC pairs can be discarded by removing them in the matrices E and F in Equation (10). If the less important EOF/EC pairs are discarded, then the reconstructed spatial anomaly patterns include only the effects of the more important correlated patterns identified by the EOF analysis. As an example, Figure 2.7 shows different approximations of the spatial anomalies for four dates. The dates were chosen to span the range of spatial patterns observed at Tarrawarra. The patterns in Figure 2.7a show the reconstructed spatial anomalies when only EC1 and EOF1 are used for the four dates. In this case, the reconstructed patterns appear identical on all dates because they look exactly like EOF1 (compare Figure 2.7a to Figure 2.3a). In reality, the pattern magnitude varies between dates in response to the different EC1 weights, but this variation is not visible due to the relative shading scale used in the figure. Figure 2.7b shows the reconstructed spatial anomalies when the first two EOF/EC pairs are used. In contrast to Figure 2.7a, the spatial anomaly patterns now look very different on the different dates due to the distinct variations of EC1 and EC2 in time. Comparing the ECs in Figures 2.3c and 2.3d with the patterns in Figure 2.7b, we see that the anomaly pattern on 22 September 1995 is dominated by EOF1 whereas the anomaly pattern on 14 February 1996 is dominated by EOF2. The other two dates include significant contributions from both EOFs. Finally Figure 2.7c shows the reconstructed spatial anomalies when all of the observed EOF/EC pairs are used in Equation (10). In this case, the original data are reproduced exactly. Comparing Figures 2.7b and 2.7c, one sees that the final 11 EOF/EC pairs produce relatively minor modifications to the patterns produced by the first 2 EOF/EC pairs.

An estimation method can be devised using the principles of EOF approximation. This method applies to the circumstance where a detailed survey of soil moisture patterns has been made in the past for the catchment of interest. For the dates on which the soil moisture pattern is to be estimated, only the spatial average of soil moisture is assumed known. The objective of the method is to use the historical soil moisture data to estimate the soil moisture patterns from the spatial average. Kachanoski and De Jong [1988] distinguish between time-series models, which attempt to estimate the magnitude of temporal changes of soil moisture, and models of soil moisture patterns, which attempt to estimate how a given temporal change is distributed spatially. This estimation method is best classified with the latter group. The basic premise of the proposed EOF estimation method is that the most important ECs can be estimated at an arbitrary time and applied to the EOFs determined from historical data to develop estimates of the soil moisture anomalies $\hat{Z}$. Specifically, we are using the equation:

$$\hat{Z} = \hat{F}\hat{E}^T \tag{13}$$

where ^ indicates an estimation of the true value. Preisendorfer [1988] calls Equation (13) the synthesis equation. $\hat{F}$ consists of observed EOFs, but it is still considered an estimate because some higher order EOFs are neglected. $\hat{E}$ must be estimated because the true EC values are not available for the date when the soil moisture pattern is desired. In order to use Equation (13) to estimate soil moisture, three main issues must be resolved. First, because $\hat{Z}$ consists only of spatial anomalies, the spatial average soil moisture for the date of interest must be known in advance. If the spatial average is known, it can be added to the spatial anomalies to produce estimates of the soil moisture

at every location. Methods to determine the spatial average are described in Section 2.6. For now, we simply assume that the spatial average is known. Second, the number of EOFs to include in $\hat{F}$ must be decided, and third, a method to estimate the ECs in $\hat{E}$ must be devised.

A reasonable way to choose the number of EOFs to include in the estimation method is to select those that represent statistically significant correlation patterns. In Section 2.3.2, EOF1 and EOF2 were found to be statistically significant, so those EOFs are used in the estimation method. It should be noted that larger datasets might reveal more than two significant EOF/EC pairs. Other approaches could be used to select the appropriate EOFs, depending on the purpose of the estimation method. Rao and Hsieh [*Rao and Hsieh*, 1991] referenced several methods for deciding which EOFs to include in an estimation procedure. The critical factor here is that the temporal behavior of the significant EOFs be predictable from the limited number of sampling dates.

Figure 2.8 demonstrates that it is reasonable to include only EOF1 and EOF2 by plotting the Nash-Sutcliffe Coefficient of Efficiency (NSCE) as successive EOF/EC pairs are neglected in reconstructing the observed spatial anomalies from the Tarrawarra dataset. The NSCE indicates the portion of the observed variation explained by the estimates and is computed:

$$NSCE = \frac{\sigma_{obs}^2 - \sigma_{\varepsilon}^2}{\sigma_{obs}^2} \tag{14}$$

where σ^2_{obs} is the variance of the observations, and σ_{ε}^2 is the mean squared estimation error. Like Figure 2.7, the observed EOFs and ECs are used for this example. We are simply incrementally removing observed EOF/EC pairs when reconstructing the data.

When all EOF/EC pairs are used in Equation (13), the NSCE is one for all dates because the data are reconstructed exactly. If no EOF/EC pairs are included, then the NSCE is zero for all dates because zero is being used as an estimate for the spatial anomaly at every location. Figure 2.8 shows that EOF1 and EOF2 are important on nearly all dates. However, as one considers the less important EOFs, the incremental increase in variance explained tends to be associated with fewer dates. This tendency confirms that the higher order EOFs are not explaining persistent correlated patterns. Instead, they are composed of variability that occurs on isolated dates. From an estimation standpoint, such patterns are not very useful because they likely provide little information about future soil moisture anomalies. Of course, by reducing the number of EOF/EC pairs used in estimating the spatial anomalies, we are immediately conceding that some spatial variation of soil moisture cannot be reproduced by the estimation procedure. Even if the ECs could be estimated exactly in Equation (13), the observations would not be matched exactly. Thus, the line for EOF/ECs 1,2 in Figure 2.8 shows an upper bound for the performance of the EOF estimation method.

A method must also be determined to estimate the EC values for arbitrary times. Because we have already assumed that the spatial average soil moisture is known for the date of interest, it would be convenient if EC1 and EC2 could be determined from the spatial average. Figure 2.9 plots the observed EC1 and EC2 values against the observed spatial average soil moisture. While significant scatter is observed in both cases, nonlinear relationships appear to hold between the EC values and the average soil moisture. We approximate both of these relationships with a cosine function with the following form:

$$\hat{e}(t) = a + b\cos\left(\frac{2\pi}{c}\bar{s}(t) - d\right) \tag{15}$$

where $\hat{e}(t)$ is the estimated EC value for a given time t and $\bar{s}(t)$ is the spatial average soil moisture at that time. a, b, c, and d are parameters where a shifts the cosine function vertically, b controls the magnitude of the wave, c is the period, and d is the phase shift. Equation (15) allows the relationship between $\bar{s}$ and EC1 to peak at intermediate values of average soil moisture, and it allows the relationship between $\bar{s}$ and EC2 to peak (in terms of absolute value) at low values of average soil moisture and again at high values. Fitted versions of Equation (15) capture 72% of the variability in the EC1 values and 76% of the variability of the EC2 values.

There is a sound basis for estimating EC weights from either the spatial average or variance. EOF analysis of the spatial anomalies essentially plots the anomalies in multidimensional space where each axis represents a time and each data point represents a location. EC1, for example, is a unit vector in this multidimensional space that is oriented in the direction of the most variance. The component of EC1 on a given day is linearly related to the variance explained on that day by the EC axis. While the variance explained by the EC is not known for the estimation application, it would be reasonable to assume that it is related to the total spatial variance on that day if the spatial variance were known. Similarly, for a time-stable pattern as defined by Vachaud et al. [1985], the spatial variance is proportional to the spatial average. Thus, it is also reasonable to estimate the EC value using the spatial average. Given that time-stability is not satisfied at the Tarrawarra catchment [*Grayson and Western*, 1998], one expects the EC weights to be more correlated to the spatial variance on each date than to the spatial average, and

this expectation is confirmed by a simple correlation analysis. However, the spatial variance will likely prove more difficult to estimate than the spatial average for a date when the spatial distribution of soil moisture is unknown. Consequently, we have opted to estimate the EC weights as a non-linear function of the spatial average (Figure 2.9). The ECs are non-linearly related to the spatial average because the spatial variance at Tarrawarra is not proportional to the spatial average. The spatial variance is low when the spatial average is very low, then increases with the average, but decreases again when the spatial average is very high.

Finally, it should be noted that an alternative EOF/EC estimation procedure could be developed using the temporal anomaly EOF analysis described in Section 2.3.3. Based on some preliminary investigations, we believe that this method might provide superior estimates if one has more data than 13 sampling dates. Using the 13 sampling dates in the Tarrawarra dataset, however, this method was found to be unreliable.

2.4.2 Results

In this section, the EOF estimation method is tested using a jackknife verification procedure. In the jackknife procedure, 12 out of the 13 dates are used to determine the EOFs and to calibrate the relationship between the spatial average soil moisture and the EC1 and EC2 values (Equation (15)). Then the calibrated estimation method is used along with the observed spatial average from the 13^{th} date to estimate the soil moisture on that 13^{th} date. The procedure is repeated so that an estimate of the soil moisture pattern is eventually made for each date using the other 12 dates to calibrate the method. Table 2.1 shows the parameters used to estimate each date's soil moisture pattern. The parameters

for EC1 are fairly stable, with coefficients of variation from 0.14 to 0.22. On the other hand, the *a* parameter for estimating EC2 varies significantly, with a coefficient of variation of almost 1. Additional sampling dates would likely produce a more stable estimate of EC2.

Figure 2.10 shows results from the EOF/EC estimation method for 28 March and 20 September along with the associated estimation errors. The two soil moisture patterns look significantly different and reflect the different processes that influence the spatial distribution of soil moisture on those dates. The estimated 28 March pattern reflects the fact that evapotranspiration is the dominant hydrologic process controlling soil moisture in the austral summer and early austral fall. It also reflects the fact that lateral redistribution has yet to develop along the drainage paths defined by the topography at this point in the year [*Western et al.*, 1999a]. On the other hand, the estimated 20 September pattern exhibits the importance of lateral drainage during the wet part of the year, shows connected drainage paths, and still includes some aspect-related effects of evapotranspiration. The estimates also correctly show that the spatial variance is much lower on 28 March than 20 September. Looking at the patterns of the estimation errors in Figure 2.10, we see mostly high frequency variation and no obvious spatial organization or relationship to the topography. Some dependence on topographic characteristics is observed for certain other dates, but when the time-averaged estimation error pattern is calculated, one can see that the majority of topographic dependence has been accounted for by the estimation method.

The performance of the estimation method was quantified by calculating the NSCE for the soil moisture pattern estimated on each date, which is shown by the heavy

line in Figure 2.11. The performance of the EOF estimation method varies substantially between days, but it generally does better on moderately wet days (see Table 2.1 for the average soil moisture on each date). This is an interesting result. If soil moisture on very dry and very wet days is associated with soil properties like permanent wilting point or porosity, as has been previously supposed [*Western et al.*, 1999a; *Wilson et al.*, 2005], then such patterns should be locally-controlled and therefore well reproduced by EOF analysis. One possible explanation for the observed behavior is that the relevant soil properties do not vary enough across this small site to control the soil moisture patterns. Soil in the top 30 cm is a relatively homogeneous silty or clayey loam, and land use is uniform across the site [*Western and Grayson*, 1998]. Thus, the small variation in soil moisture caused by soil properties may be overwhelmed by uncorrelated variation and instrument error during very wet or very dry periods. Overall, the EOF estimation method was able to explain about 36% of the total observed spatial variation of soil moisture. Recall that by neglecting all but two EOFs, we immediately conceded that much of the variation could not be reproduced. In fact, the maximum variance that could have been reproduced is 64%. Thus, loosely speaking, half of the estimation error arises from the portion of the spatial variation that is not included in EOF1 and EOF2, and half of the estimation error arises from errors in the estimates of EC1 and EC2.

2.4.3 Comparison to Other Methods

The performance of the estimation method can be put into more context by comparing it to other possible methods. In this section, we compare its performance to that of three other methods.

Linear Regression. One of the simplest approaches to estimate soil moisture patterns at unobserved times would be a linear regression approach. In this method, one would assume that the soil moisture at a particular location in space and time is a linear function of the average soil moisture at that time, which we have already assumed is known. Thus, the soil moisture at location i and time t can be found from:

$$s_i(t) = a_i + b_i \overline{s}(t) \tag{16}$$

where a_i and b_i are the regression parameters for location i and $\overline{s}(t)$ is the average soil moisture at time t. This method is based on a perspective similar to time-stability [*Vachaud et al.*, 1985] because the soil moisture at each location has a local but time-invariant linear relationship to the spatial mean. Although this approach is conceptually very simple, it is cumbersome to implement. In order to estimate the soil moisture at the 459 locations in the Tarrawarra dataset, 918 parameters must be determined to implement Equation (16).

Wilson et al. Method. The second method we consider was introduced by Wilson et al. [2005]. They proposed a dynamic multiple linear regression of soil moisture against topographic attributes, which can be written:

$$s_i(t) = \overline{s}(t) + \sum_{j=1}^{k} \hat{\sigma}_s[\overline{s}(t)] \cdot p_{Tj}[\overline{s}(t)] \cdot \frac{\left(T_{ji} - \overline{T}_j\right)}{\sigma_{Tj}} \tag{17}$$

where j is an index of topographic attributes, k is the total number of topographic attributes considered, T_{ji} is the value of a topographic attribute at location i, $\overline{T}_j$ is the spatial average of the topographic attribute, and σ_{Tj} is the spatial standard deviation of the topographic attribute. $\hat{\sigma}_s$ is the spatial standard deviation of the soil moisture, and

p_{Tj} is the partial regression coefficient of the soil moisture with respect to T_j. Both $\hat{\sigma}_s$ and p_{Tj} are assumed to be a function of the spatial average soil moisture, which is consistent with observations [*Western et al.*, 1999a]. Like the other methods considered here, this method assumes that the spatial average soil moisture is known for the time when the soil moisture pattern is required. Wilson et al. [2005] include a predictor-corrector component in their model where they utilize the time-averaged error from a first estimate of soil moisture as a new predictor variable (T_j) in a second estimate. They refer to this version of their method as Case 2. The rational for this approach is that the error pattern includes persistent sources of variation such as soil properties or vegetation, which are not captured by the topographic attributes. Notice that Case 2 of the Wilson et al. model assumes that a high resolution soil moisture dataset is available, from which to compute the errors of estimation. In this way, Case 2 is comparable to the other estimation methods described here.

Rao and Hsieh Method. Rao and Hsieh [1991] used an EOF method to estimate discharge time series at ungaged locations. The primary differences between the EOF estimation method we propose in Section 2.4.1 and their method are: (1) our EOF analysis is applied to soil moisture instead of stream discharge, (2) our EOF analysis uses spatial anomalies instead of temporal anomalies, and (3) we estimate the ECs as a function of average soil moisture whereas they interpolate the EOFs between known values. The last distinction is the most important because estimation of ECs from an independent variable allows our method to be extended beyond the time constraints of the original dataset. For the purposes of comparison, we applied a method similar to Rao and Hsieh's to the Tarrawarra soil moisture data. In this application, the EOFs are still

developed from the spatial anomalies of soil moisture, but we determine the EC1 and EC2 values by linearly interpolating between the values on the preceding and following dates instead of determining their values from the average soil moisture.

Figure 2.11 shows the NSCE for the results of these alternative methods along with the NCSE for the proposed EOF method. All the methods perform similarly, as one might expect given their underlying similarities. In general, the methods perform better during moderately wet days and worse during the dry summer season when the spatial variation is small. One would like the estimation methods to consistently produce NSCE values above zero because all of the methods assume that the spatial average soil moisture is known, and the spatial average itself would produce an NSCE value of zero if it was used as an estimate of the soil moisture patterns. The lowest NSCE values for the new EOF estimation method and the Rao and Hsieh estimation method are close to zero, while the lowest values for the other two methods are a little below zero. The average NSCE values are 0.36 for the new EOF estimation method, 0.33 for the Rao and Hsieh method (note that the first and last dates cannot be estimated with this method), 0.32 for the linear regression method, and 0.30 for the Wilson et al. method.

2.5 Determining the Spatial Average and ECs

In all of the analyses above, we have assumed that the spatial average soil moisture is known. In practice, the spatial average could be determined in several ways. For example, remote sensing could be used. One weakness of this approach is that remote sensing methods observe the soil moisture near the ground surface rather than over the entire root zone, which is often of interest. Furthermore, the spatial average

estimated from remote sensing may span a much larger region than the catchment of interest. Alternatively, a water balance model could be used to estimate the average soil moisture for the catchment. This approach should also be used with caution because the effective average soil moisture given by a model may not correspond to the actual average soil moisture in the catchment. Another method to estimate the average soil moisture is to use in-situ soil moisture observations. Grayson and Western [1998] showed that it is possible to estimate an aerial average soil moisture by monitoring soil moisture at a small number of carefully selected locations, which they called catchment average soil moisture monitoring (CASMM) sites. The soil moisture at these points has a linear relationship with the average soil moisture that has a slope of one and an intercept of zero. Grayson and Western [1998] identified catchment properties that are associated with CASMM points and suggested using these properties to select monitoring locations. However, Grayson and Western [1998] also argued that if soil moisture sampling has been done for the entire catchment, a measurement of soil moisture at almost any point can provide a good estimate of the spatial average soil moisture. This statement is true because the soil moisture at nearly all locations exhibits an approximate linear relationship with the spatial average soil moisture. Thus, one can use a regression to determine this relationship and then use the relationship to estimate the average for any time based on the measured value. Because we have already assumed that soil moisture has been sampled throughout the catchment, the latter approach is most appropriate for our purposes. To implement this approach, we are interested in identifying locations with the highest linear correlation through time between the local soil moisture and the spatial average. Figure 2.12a shades each location according to the correlation between the local

soil moisture and the spatial average. It confirms that the correlation is high at nearly all locations. The highest correlation is 0.996 and much of the catchment has a correlation above 0.95. It confirms that nearly any point could be monitored to reliably determine the spatial average soil moisture, which validates our assumption in Section 2.4 that the spatial average can be estimated at an unobserved time.

It is also worth considering whether certain locations could be monitored to indicate the values of the spatial variance or EC1 and EC2, which could be used instead of relying on relationships between the ECs and the spatial average. Figure 2.12b shows the correlation between point soil moisture values and the spatial variance of soil moisture for the entire spatial domain. The highest correlation for all locations is 0.88. This supports our earlier decision to use the estimated spatial average instead of the estimated spatial variance in the EOF estimation procedure. Figure 2.12c and 2.12d show the temporal correlations between the soil moisture at each location and EC1 and EC2, respectively. As expected, locations where the variation is closely related to EOF1 tend to track EC1 through time. The most highly correlated locations are along the main drainage paths. The highest correlation between the soil moisture at a location and EC1 is 0.84. Likewise the soil moisture on the northwest-facing hillslope tends to track EC2 through time. The highest correlation between the soil moisture at a location and EC2 is 0.91. These results suggest that one might be able to use monitoring at selected sites to determine the EC values directly, although the estimates of the ECs would be less reliable than the estimates of the spatial mean. To test this approach, we repeated the EOF estimation procedure using point measurements of soil moisture to estimate the ECs instead of the relationships between the ECs and the spatial average. The average NSCE

of this approach is slightly higher (0.38 versus 0.36 in Section 2.4). However, this approach is more restrictive because the ECs must be determined from in-situ monitoring whereas the spatial average can be determined from more methods.

2.6 Conclusions

The primary conclusions of this paper are as follows:

1. For the Tarrawarra dataset, EOF analysis reveals that the spatial anomalies of soil moisture are determined in part by two specific spatial patterns whose individual contributions vary differently through time. The first pattern (EOF1) mainly produces positive anomalies in the valleys and zero or negative anomalies in the hillslopes. Such a tendency suggests that this pattern is primarily associated with lateral redistribution of soil water. Overall, this pattern explains 55% of the spatial variance in the dataset. The second spatial pattern (EOF2) modifies the first and, during the dry season, tends to produce negative anomalies for hillslopes that receive more solar radiation and positive anomalies for hillslopes that receive less solar radiation. During the very wet season the effect of EOF2 switches, reflecting the seasonally varying role of evapotranspiration. Overall, this spatial pattern explains 9% of the spatial variation.
2. A similar analysis of the temporal anomalies suggests that temporal variation in the Tarrawarra dataset is dominated by a single mode of variability which is associated with the wet and dry seasons. A dataset with greater temporal detail and extent may reveal more modes of variation associated with individual storm events.

3. A new method was proposed to estimate soil moisture patterns for times when only the spatial average soil moisture is known using the two spatial patterns identified by the EOF analysis of the spatial anomalies. For any given time, this method weights the two spatial patterns depending on the spatial average soil moisture for that time. The method outperforms other available methods for the Tarrawarra dataset, explaining 36% of the variance on average but as much as 75% of the variance on one date.
4. It was also shown that the appropriate weights for the two spatial patterns at any given time could also be inferred directly from soil moisture observations at carefully selected locations. In particular, the weight for the primary pattern can be determined by monitoring the soil moisture at a point in a main valley, and the weight for the secondary pattern can be determined by monitoring the soil moisture at a point on the northwest-facing hillslope (i.e. a point that is well aligned with the incident solar radiation).

Table 2.1 Parameters used in the EOF estimation method to estimate ECs from the spatial average soil moisture.

Estimation Date	Spatial Average Soil Moisture	EC1 parameters				EC2 parameters			
		a	*b*	*c*	*d*	*a*	*b*	*c*	*d*
27 Sept 95	0.38	0.22	0.16	45	5.1	0.04	0.36	54	6.3
14 Feb 96	0.26	0.09	0.27	71	3.5	0.18	0.50	58	6.3
23 Feb 96	0.21	0.23	0.14	40	6.3	0.11	0.42	56	6.3
28 March 96	0.24	0.23	0.14	40	6.2	0.11	0.42	56	6.3
13 April 96	0.35	0.24	0.16	39	6.3	0.07	0.40	55	6.3
22 April 96	0.41	0.24	0.15	40	6.3	0.32	0.61	78	5.4
2 May 96	0.42	0.23	0.15	43	5.8	0.08	0.39	55	6.3
3 July 96	0.45	0.24	0.15	41	6.1	0.08	0.39	54	6.3
2 Sept96	0.49	0.23	0.15	39	6.3	0.46	0.76	78	5.4
20 Sept 96	0.48	0.24	0.14	40	6.2	0.05	0.36	53	6.3
25 Oct 96	0.35	0.24	0.14	40	6.3	0.08	0.41	56	6.3
10 Nov 96	0.30	0.23	0.14	40	6.3	0.07	0.42	56	6.3
29 Nov 96	0.24	0.23	0.15	40	6.3	0.03	0.38	54	6.3

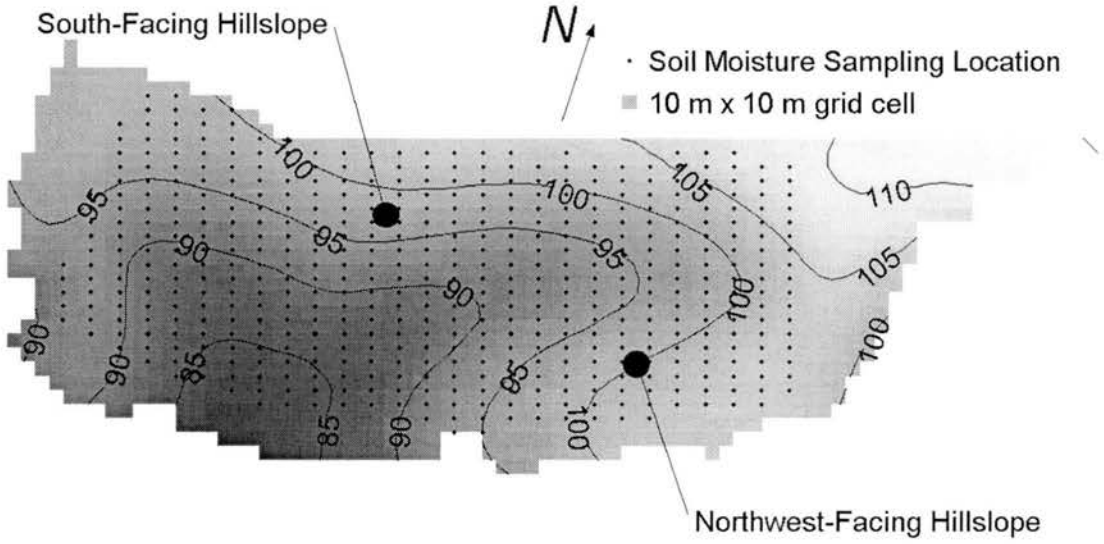

Figure 2.1 Tarrawarra catchment topography. Shading and contours indicate topographic elevation (in meters). Dots show a typical pattern of soil moisture sampling.

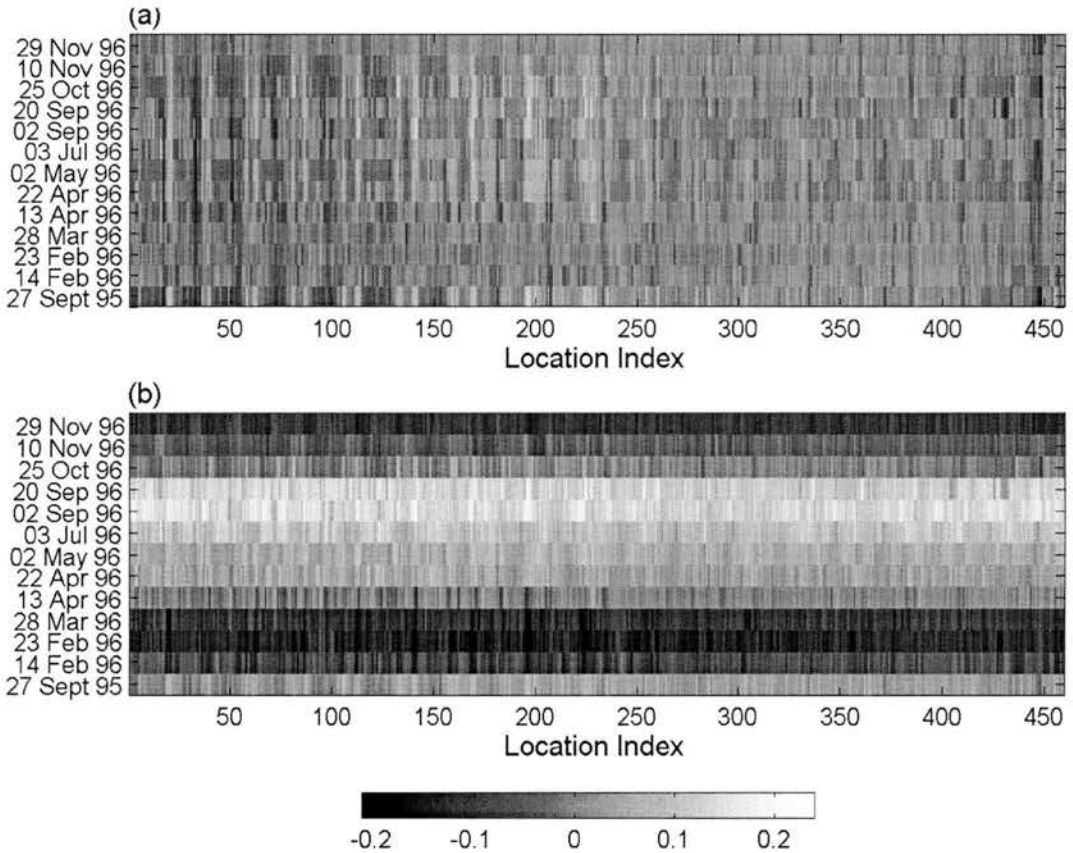

Figure 2.2 (a) Spatial and (b) temporal anomalies of soil moisture plotted by date and location. The location index is an arbitrary number assigned to each grid cell with soil moisture data.

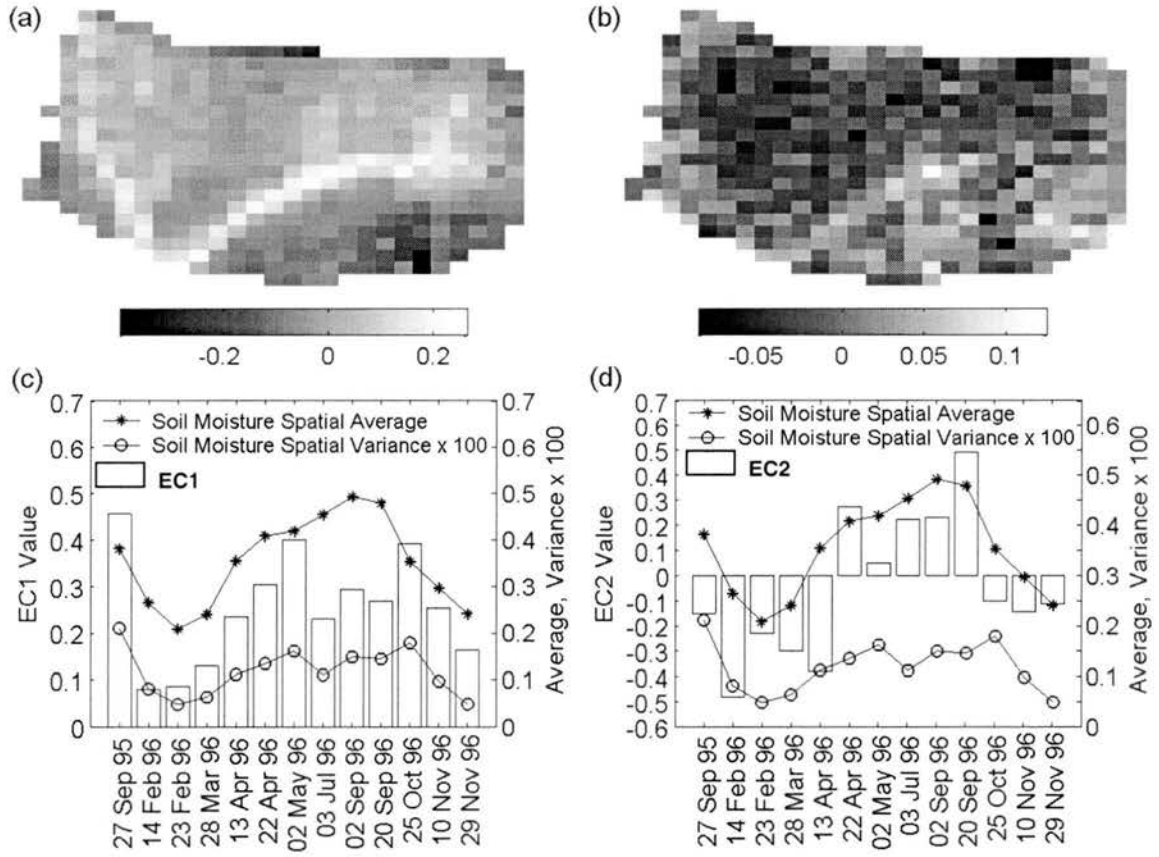

Figure 2.3 (a) EOF1 and (b) EOF2 computed from the spatial anomalies of soil moisture. The bars in (c) and (d) are EC1 and EC2, respectively. The spatial average and variance of the soil moisture on each sampling date are also shown in (c) and (d) for comparison.

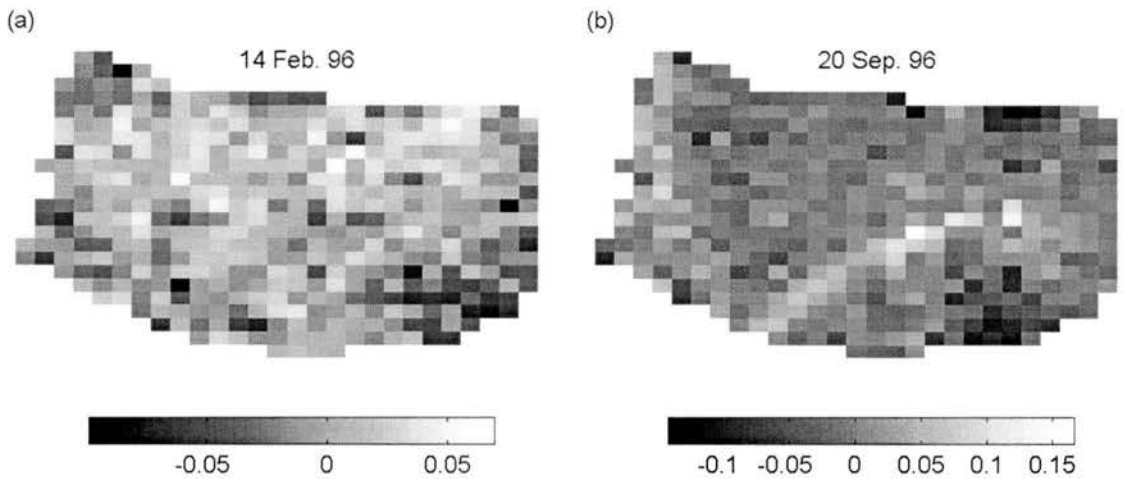

Figure 2.4 The spatial anomalies of soil moisture for (a) a very dry day (14 February 1996) and (b) a very wet day (20 September 1996). The former has the largest negative value for EC2, and the latter has the largest positive value for EC2.

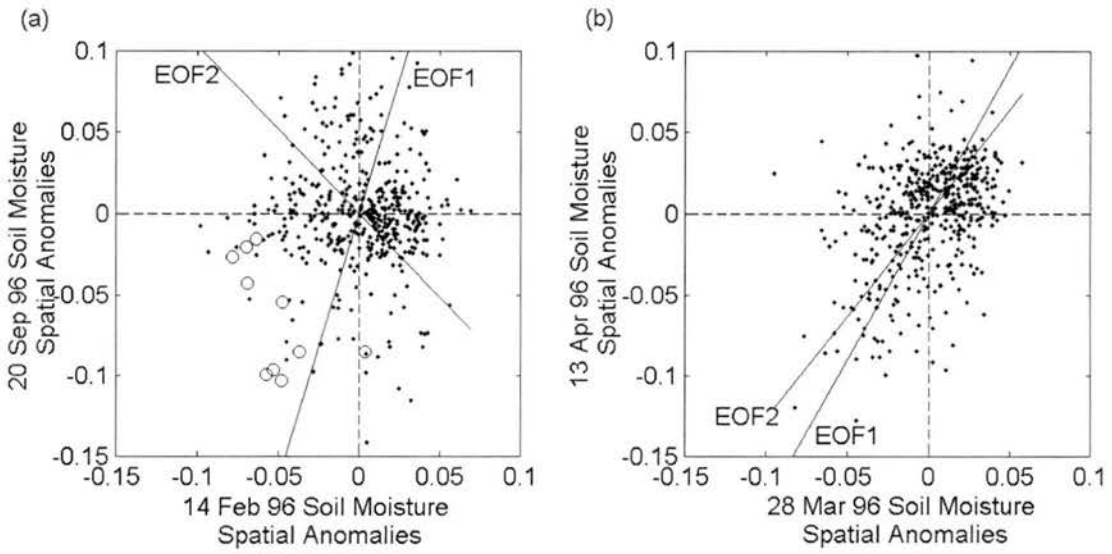

Figure 2.5 EOF1 and EOF2 overlaid on a plot comparing the spatial anomalies of soil moisture on two dates. (a) compares 20 September 1996 and 14 February 1996 and (b) compares 13 April 1996 and 28 March 1996. Each dot corresponds to a location in Tarrawarra. Open circles show locations on the northwest-facing slope that exhibit negative correlation between wet (20 Sep) and dry (14 Feb) days.

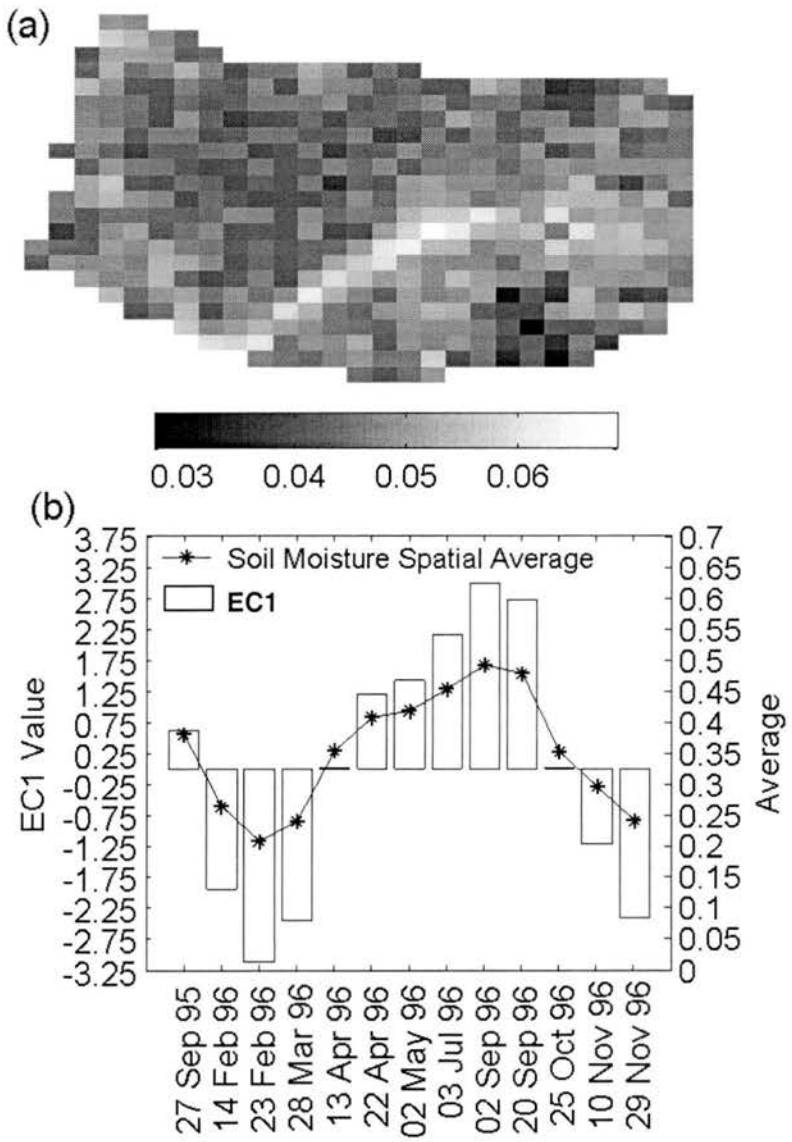

Figure 2.6 (a) EOF1 and (b) EC1 computed from the temporal anomalies of soil moisture. The bars in (b) are the EC1 values and the superimposed line is the spatial average of soil moisture.

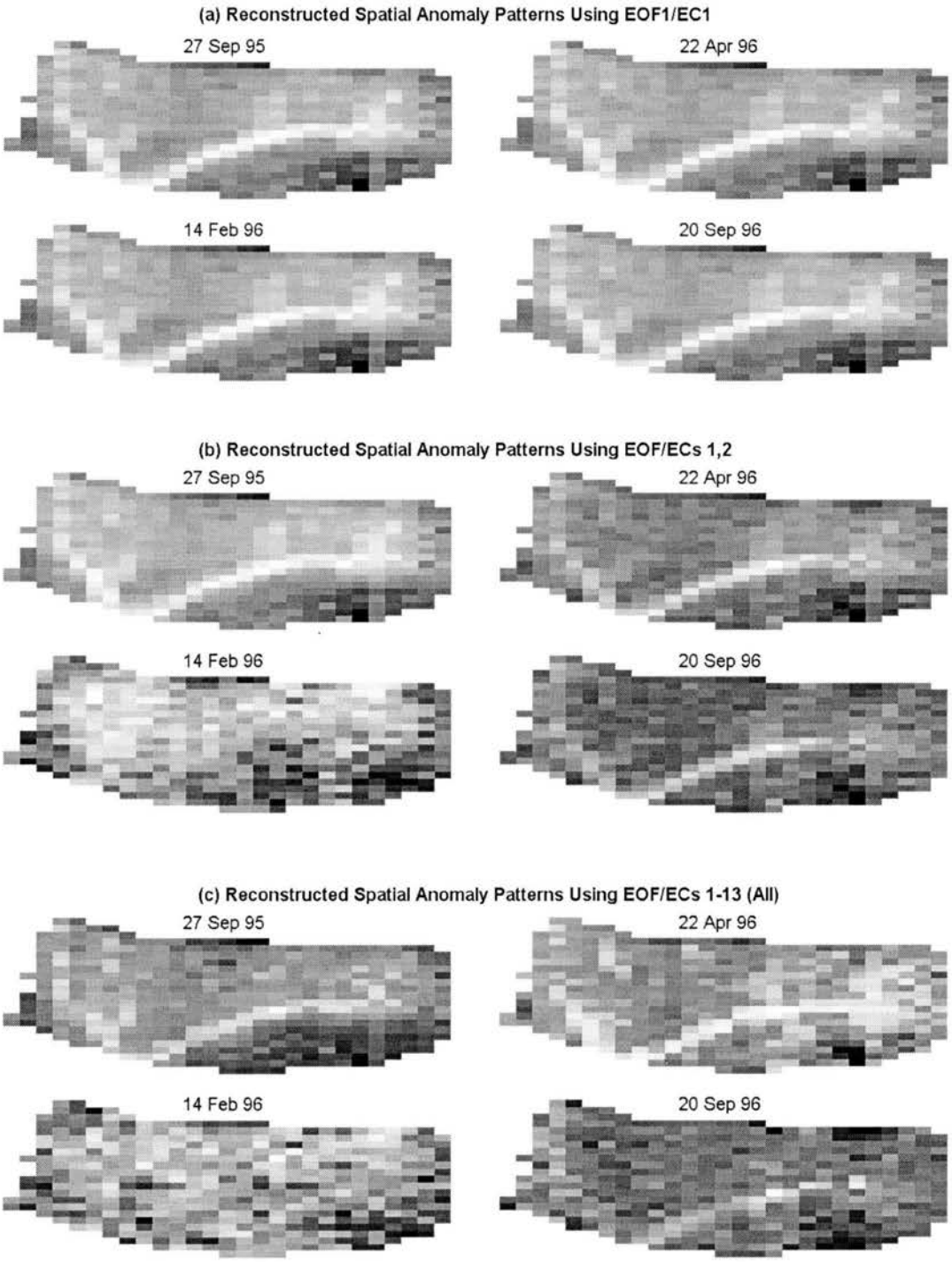

Figure 2.7 Reconstructed soil moisture spatial anomalies on four dates using (a) the first observed EOF/EC pair only, (b) the first two observed EOF/EC pairs, and (c) all 13 observed EOF/EC pairs. When all observed EOF/EC pairs are included, the original data are reproduced exactly.

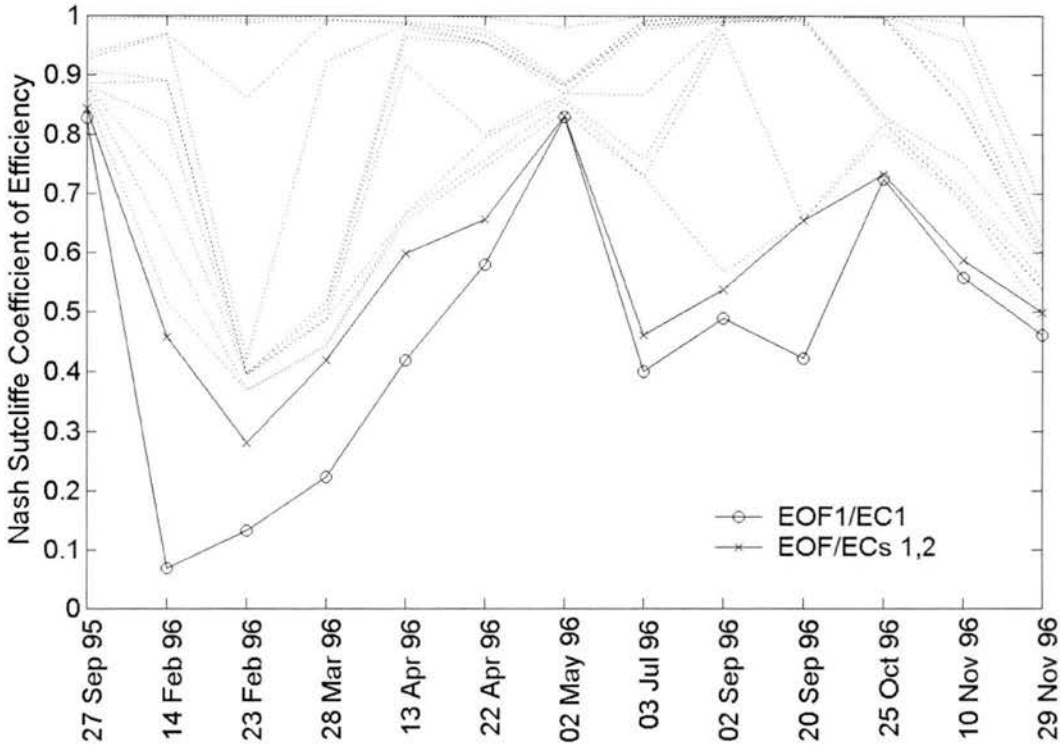

Figure 2.8 Nash Sutcliffe Coefficient of Efficiency (NSCE) when subsets of the 13 observed EOF/EC pairs are used to produced an estimate of the spatial anomalies of soil moisture. The lowest line uses only the first EOF/EC pair, the second lowest line uses the first two EOF/EC pairs. Each higher line (dotted) includes one additional EOF/EC pair. When all 13 observed EOF/EC pairs are included, the original data are reproduced exactly (NSCE is 1).

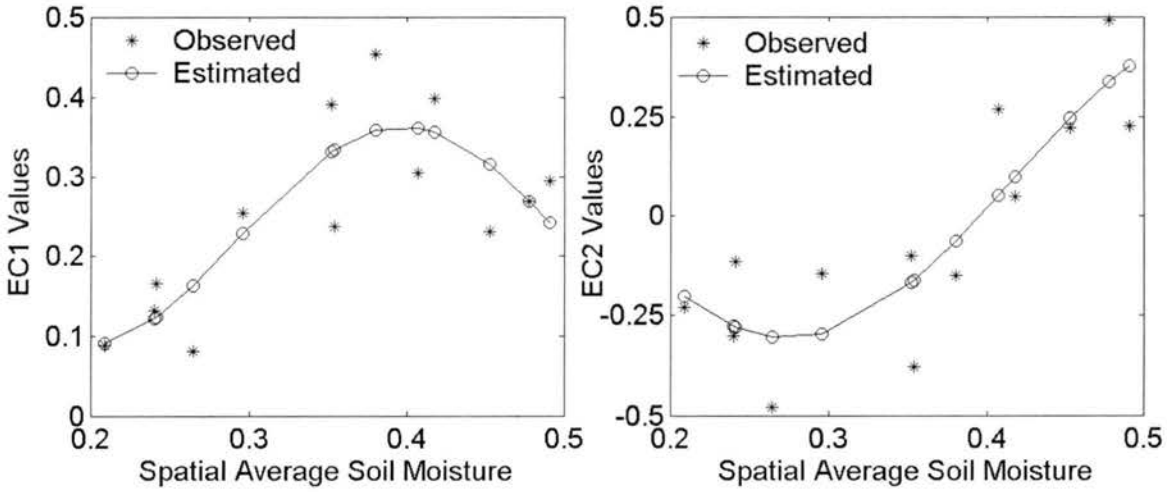

Figure 2.9 Relationships between EC1 and EC2 and the spatial average soil moisture. Observed data points are shown as well as the best-fit cosine functions.

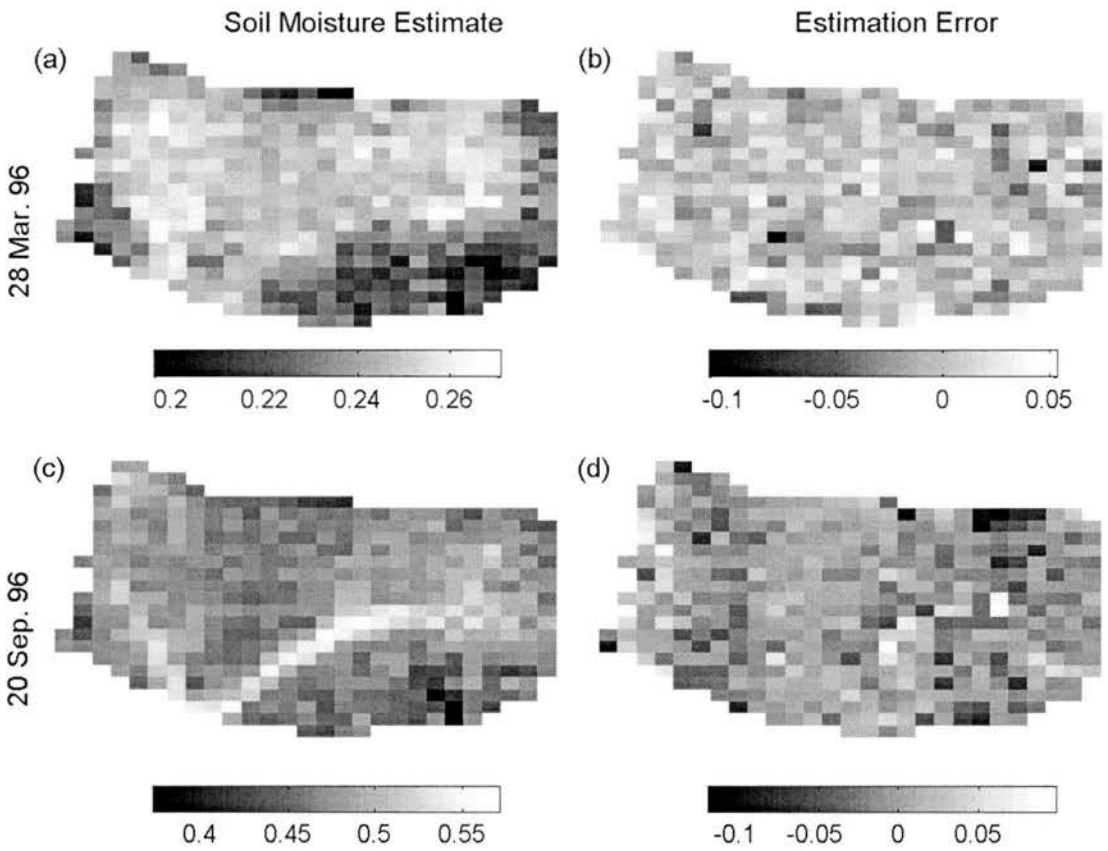

Figure 2.10 Results of the EOF/EC soil moisture estimation method. (a) shows the estimated soil moisture for 28 March 1996 and (b) shows the estimation error. (c) shows the estimated soil moisture for 20 September 1996 and (d) shows the estimation error.

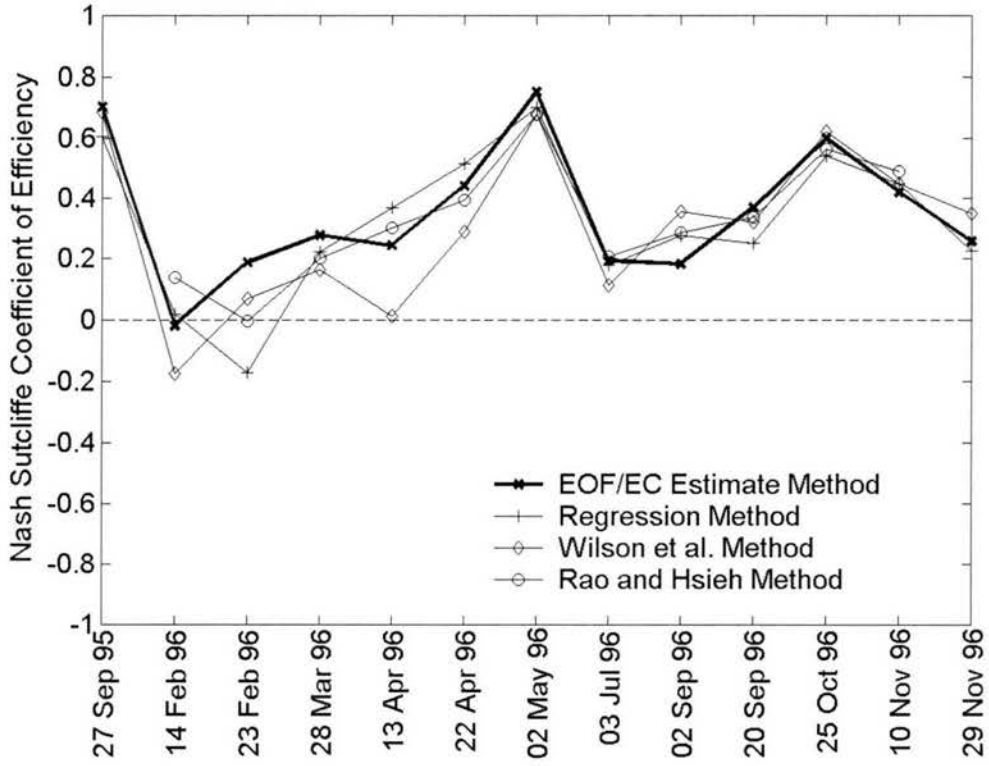

Figure 2.11 NSCE of the soil moisture estimates produced by four methods.

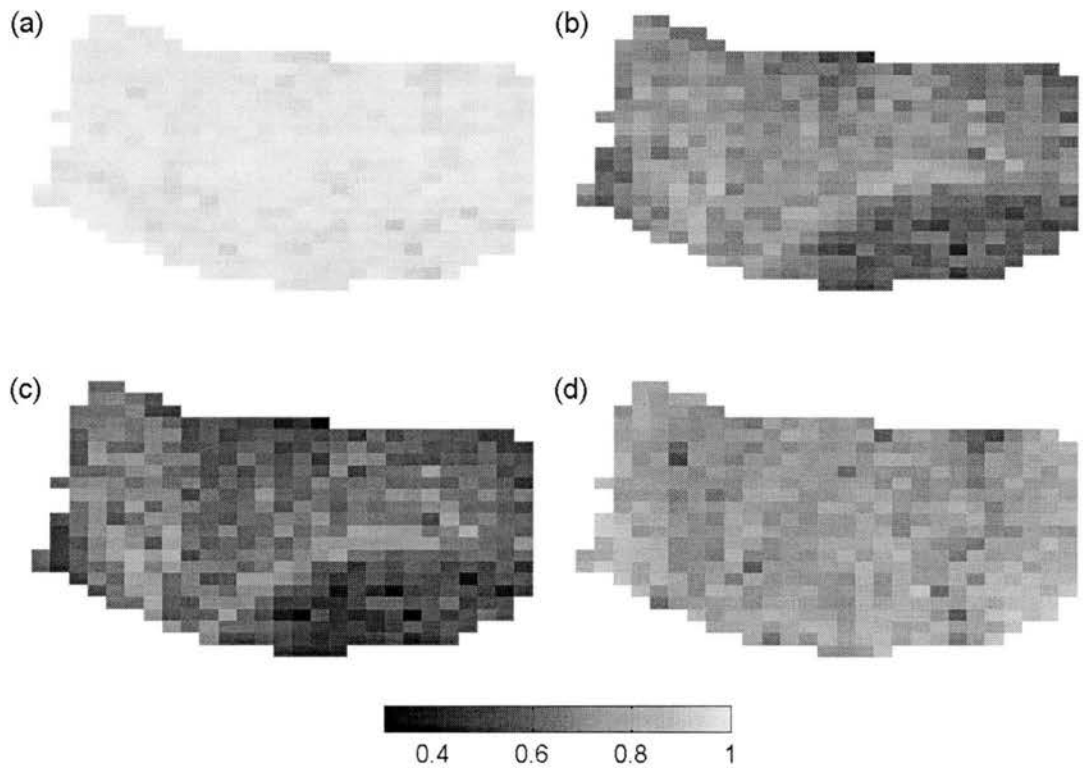

Figure 2.12 Maps of the observed temporal correlation coefficient between the soil moisture value and (a) the spatial average soil moisture, (b) the spatial variance of the soil moisture, (c) the EC1 value, and (d) the EC2 value.

3 Improved Spatial Interpolation of Soil Moisture at the Catchment Scale Using EOFs

Abstract

Spatial patterns of soil moisture at the catchment scale are important to many applications in hydrology, ecology, agriculture, and climatology. Because soil moisture exhibits complex variability in space and time, it is usually too costly to fully characterize soil moisture patterns by direct measurement. As a result, soil moisture patterns usually must be estimated from limited data. Previous studies at the catchment scale have shown that two largely independent processes control much of the variation in soil moisture. One process is lateral drainage of soil water and the other is evapotranspiration. Based on these findings, we hypothesize that soil moisture can be effectively decomposed using Empirical Orthogonal Function (EOF) analysis, which separates space-time variation into time-invariant patterns of orthogonal covariation (EOFs) and time-varying weights called expansion coefficients (ECs). Here, EOF analysis is applied to the Tarrawarra catchment in southeast Australia. It is found that EOF analysis decomposes soil moisture roughly according to the controlling processes and that the most important EOFs exhibit distinct but more consistent spatial structures than soil moisture itself. It is also shown that the less predictable variation occurs in the higher order EOFs. We show that spatial interpolation of soil moisture can be done more effectively by first using EOF analysis to

separate the system variation from the random variation, next interpolating the most important EOFs separately, and finally using the interpolated EOFs to construct the fine-scale soil moisture pattern. This EOF-based interpolation method is shown to outperform equivalent methods applied directly to soil moisture.

3.1 Introduction

Spatial variability of soil moisture is important because it affects agriculture [*Jaynes et al.*, 2003; *Green and Erskine*, 2004], climate [*Delworth and Manabe*, 1988; *Liu*, 2003], ecology [*Moore et al.*, 1993], and hydrology [*Bárdossy and Lehmann*, 1998; *Western et al.*, 1999b]. Jaynes et al. [2003] documented the complex response of crop yields to the spatial distribution of soil moisture. They found that low topographic areas tend to have higher soil moisture and thus have higher than average yields during dry years and lower than average yields during wet years. Western et al. [1999b; 2001] showed that the spatial pattern of soil moisture affects the discharge produced by a catchment in response to a rainfall event. If wet areas are well connected to the outlet, then there is less opportunity for run-on infiltration, resulting in a faster time to peak and a larger peak discharge. The spatial distribution of soil moisture also affects the energy balance of the land surface [*Bras*, 1999]. In particular, soil moisture controls the partitioning of energy between latent and sensible heat fluxes.

Spatial patterns of soil moisture are complicated because they depend on several distinct hydrologic processes. At the catchment-scale, Grayson et al. [1997] argued that two largely independent processes control much of the variation in soil moisture in seasonal climates. One process is the lateral redistribution of surface and subsurface

water, and the other is evapotranspiration. Lateral redistribution occurs during relatively wet periods. When soil is wetter, the hydraulic conductivity is larger. The increased hydraulic conductivity allows for the redistribution of soil moisture roughly in the downhill direction. The resulting spatial variation of soil moisture tends to have wetter locations in the valleys and drier locations on the hillslopes. During dry periods, the hydraulic conductivity decreases non-linearly, so lateral redistribution essentially stops. In this state, the spatial pattern is controlled by evapotranspiration or vertical soil water fluxes. This pattern resembles topographic slope and aspect, which affect the evapotranspiration process. Moore et al. [1993] also considered the controls on soil moisture patterns with an analysis in the Brindabella Range in Australia. They identified the controls as water divergence/convergence, infiltration, and evapotranspiration. They related the divergence and convergence of soil water to hillslope position, and they suggested that infiltration and evapotranspiration are related to topographic slope and aspect. Zaslavsky and Sinai [1981] observed that soil moisture is higher in areas of concave curvature for a site in the desert of Israel, where soils are unsaturated and the water table is very deep. They showed that water accumulation is caused by the decrease in lateral soil water flux resulting from a decrease in the topographic slope at such locations. Zaslavsky and Sinai [1981] also described how soil water can be redistributed laterally in the unsaturated zone due to anisotropic vertical and horizontal hydraulic conductivities caused by soil layering.

Numerous topographic attributes have been proposed to describe the influence of the hydrologic processes that control soil moisture. The well-known wetness index (*WI*) [*Beven and Kirkby*, 1979] has been used to describe patterns of lateral drainage of soil

water. *WI* increases with increasing upslope drainage area and decreasing local slope. Likewise, various forms of ground curvature have been proposed to identify regions of flow convergence [*Mitasova and Hofierka*, 1993; *Schorghofer and Rothman*, 2001]. Topographic slope, aspect, and the potential solar radiation index (*PSRI*) have been used to describe the distribution of solar insolation [*Moore et al.*, 1993; *Dingman*, 2002]. Burt and Butcher [1985] studied a 1.4 ha hillslope and documented the redistribution of soil water in the downhill direction and the accumulation of water in lower slope and more convergent areas. In this case, the *WI* explained the spatial distribution of soil moisture at some times but not at others. Western et al. [1999a] observed linear correlations with both *WI* and *PSRI* at the 10.5 ha Tarrawarra catchment in southeast Australia. On one date, these two attributes together explained 61% of the observed variation in soil moisture. They noted that *WI* and *PSRI* are highly independent, explaining different parts of observed variation and that their relative importance changes through time. On the Canadian prairie, Florinsky et al. [2002] found soil moisture to be linearly correlated with *WI* and curvature.

As with topography, other physical properties of a site can influence soil moisture patterns. Seyfried and Wilcox [1995] showed that soil moisture can vary at scales as small as plant spacing and as large as the scale of topographic features and geological formations. It is well-known that soil properties like hydraulic conductivity are highly variable at small scales [*Skoien et al.*, 2003], while soil texture also varies at large scales and can produce large scale variations in soil moisture [*Cosh et al.*, 2004]. In natural landscapes, vegetation density and type also vary significantly and play an important role

in creating soil moisture variation. A discussion of this topic is provided by Moore et al. [1993].

It must also be acknowledged that a large part of observed soil moisture variation is simply due to measurement error [*Verhoest et al.*, 1998; *Western and Grayson*, 1998; *Wilson et al.*, 2004]. Wilson et al. [2004] described the role of measurement error when using ground-based soil moisture measurements. In their case, measurement error made up between 10% and 60% of the observed spatial variation of soil moisture, depending on the amount of actual variation on the sampling date. Li and Islam [1999] used microwave soil moisture measurements to update a soil moisture profile model through a data assimilation procedure. To evaluate the effect of measurement error, they introduced uniform random variation in the soil moisture measurements. They concluded that assimilation of soil moisture measurements containing errors of realistic size would not improve the predictions from their model. Verhoest et al. [1998] used principal component analysis to process synthetic aperture radar signals. They decomposed the signals into separate components that are related to topography, soil moisture, and land use. They attributed the remaining 11% of the variation to noise.

Several other studies have decomposed soil moisture patterns into underlying patterns in order to determine the roles played by the controlling processes and the associated physical characteristics. Using remotely-sensed soil moisture data from the Washita '92 experiment [*Jackson et al.*, 1995], Hu et al. [1998] decomposed soil moisture into component variables using wavelet transforms. They decomposed soil moisture by frequency based on the view that soil moisture patterns result from the interaction of large-scale atmospheric processes and various smaller scale land surface

processes [*Hu et al.*, 1998]. They found that the variability over a range of relatively small scales was approximately self-similar, while the soil moisture as a whole exhibits a multifractal tendency. Jawson and Niemann (preprint, 2005) decomposed remotely-sensed soil moisture data from the Southern Great Plains 1997 field campaign [*Jackson et al.*, 1999] using empirical orthogonal function (EOF) analysis. Rather than identifying component patterns by frequency, EOF analysis decomposes space-time data based on the assumption that the underlying patterns are orthogonal modes of variation called EOFs. Jawson and Niemann (preprint, 2005) found that the two most important underlying patterns (EOF1 and EOF2) explained 79% of the total spatial variance of the dataset. EOF1 was most highly correlated with soil texture, while EOF2 was most highly correlated to elevation. Their interpretation was that EOF1 described the spatial variation induced by drainage and water retention processes, and EOF2 was related to the distribution of precipitation, which was correlated with elevation in that case. The remaining EOF patterns were considered to be statistically insignificant or indistinguishable from random variation. At the scale of agricultural fields, Yoo and Kim [2004] used EOF analysis to decompose space-time soil moisture patterns for two sites in the Little Washita experimental watershed in Oklahoma, USA. Their decomposed patterns were related to topography and soil properties. Wilson et al. [2004] also decomposed soil moisture variation at the catchment scale using an *a priori* assessment of what the sources of variation might be. Based on previous studies, they expected that topographic attributes are important controls, so they performed a type of multiple linear regression of soil moisture patterns against topographic attributes for various sampling dates. They also computed the time-averaged residual, which is the average pattern of

variation not explained by topography. They reasoned that the average residual pattern might contain variation caused by persistent non-topographic controls such as soil texture and vegetation. The remaining residual was considered measurement error and unexplained variation.

All of these studies about soil moisture patterns are valuable because soil moisture is not usually observed at spatial scales that are adequate for most practical applications. Satellite remote sensing can characterize the spatial and temporal variability of soil moisture [*Bras*, 1999; *Entekhabi et al.*, 2004], but currently it observes the average soil moisture over very large land areas. As a result, methods are required to downscale or disaggregate remotely sensed soil moisture. Numerous studies have focused on the application of statistical methods to simulate sub-grid variation [*Charpentier and Groffman*, 1992; *Hu et al.*, 1998; *Kim and Barros*, 2002a]. These methods aim to produce realistic variations of soil moisture at fine resolutions; they do not attempt to produce estimates of soil moisture at particular locations within a grid cell. Consequently, they are aimed at applications like climate modeling and weather prediction. For example, Kim and Barros [2002a] used a modified fractal interpolation technique with contraction mapping. Their method used spatially and temporally varying scaling functions derived from spatial distributions of ancillary data (e.g., soil and vegetation). They demonstrated downscaling from 10 km to 825 m where the final resolution is limited by the resolution of the ancillary data.

A second approach to obtain spatial patterns of soil moisture is to estimate them using either process-based or statistically-based models. The objective of this class of methods is to estimate the soil moisture values at all locations within a catchment or a

remote sensing grid cell. Pellenq et al. [2003] coupled a soil vegetation atmospheric transfer (SVAT) model with Topmodel [*Beven and Kirkby*, 1979] to estimate soil moisture patterns at a 20 m x 20 m pixel resolution. In this approach, the evapotranspiration process is modeled with the SVAT component, while lateral redistribution is modeled with Topmodel. Because potential evapotranspiration is considered to be spatially uniform, the spatial variation of soil moisture is controlled by lateral drainage. Downer and Ogden [2003] used a process-based distributed hydrologic model called GSSHA on a 21 km^2 watershed using 125 m x 125 m pixel resolution. They calibrated the model using discharge at the watershed outlet and demonstrated that the model reproduced observed trends in soil moisture at monitored locations within the catchment. Soil moisture spatial variation in GSSHA is controlled by variation in evapotranspiration and infiltration. Western et al. [1999b] used a conceptual hydrologic model called Thales to estimate the soil moisture pattern in a 10.5 ha catchment at an average element size of 140 m^2. In this model, potential evapotranspiration is spatially uniform while lateral redistribution of soil water can occur if the soil moisture exceeds the field capacity. Thus, the spatial variation of soil moisture is controlled by lateral redistribution in this model. Wilson et al. [2005] used a statistical model to estimate soil moisture patterns at various sites ranging from hillslopes (5 ha) to small catchments (60 ha) with grid resolutions varying from 100 m^2 to 1600 m^2. Their model is a dynamic multiple linear regression that links the spatial distribution of soil moisture to topographic attributes. Finally, in Chapter 2 herein a statistical method based on EOF analysis was used to estimate soil moisture patterns at unobserved times. Underlying persistent

patterns were related to different controlling hydrologic processes such as lateral drainage and evapotranspiration.

A third approach for obtaining soil moisture patterns from limited data is spatial interpolation of sparse measurements. In this approach, soil moisture values are estimated between sparse observations using spatial relationships to the observation points and/or correlations to other properties observed at a finer resolution. Bárdossy and Lehmann [1998] interpolated soil moisture to a 12.5 m x 12.5 m resolution using 110 irregularly spaced time domain reflectometry (TDR) gages over a 6.3 km^2 catchment. They demonstrated various permutations of kriging including ordinary kriging and external drift kriging. Ordinary kriging performs poorly because it ignores the physical processes that control soil moisture patterns such as the tendency of water to move downhill from a point. External drift kriging partially overcomes this limitation by introducing a linear relationship with some independent variable, which Bardossy and Lehmann [1998] chose to be *WI*. This method accounts for the role of lateral variation, but it does not account for role of evapotranspiration, which was unimportant for that particular dataset. Thattai and Islam [2000] used kriging to determine whether remotely-sensed soil moisture from widely-spaced flight paths could be interpolated to produce a full soil moisture pattern at 200 m x 200 m resolution. They showed that flight lines could be spaced four times greater (i.e., observing only 25% of the original rows) and then filled-in by interpolation.

While less research has focused on interpolation than the other two approaches, interpolation has several benefits. Unlike the modeling methods described earlier, interpolation methods do not require large amounts of data. Process-based models like GSSHA and Thales require knowledge of a large number of spatially-distributed

properties to realize their full potential and they require calibration based on past streamflow or soil moisture measurements. Similarly, the statistical methods proposed by Wilson et al. [2005] and in Chapter 2 herein require past high-resolution soil moisture data for training. And unlike downscaling methods, interpolation can be applied to ground-based, point measurements, which are valuable for agriculture [*Green and Erskine*, 2004] and catchment scale hydrology [*Western and Grayson*, 1998]. Point measurements are also valuable for research efforts to understand subgrid variation in climate models [*Charpentier and Groffman*, 1992; *Bras*, 1999] and for ground-truthing remote sensing data [*Mohanty et al.*, 2000].

In this paper, we focus on interpolation of soil moisture data at the catchment scale. Our main objective is to develop a method of soil moisture interpolation that accounts for multiple underlying processes and other sources of variation. Our hypothesis is that soil moisture patterns can be decomposed into component patterns by EOF analysis that have more predictable spatial variation than soil moisture itself. If this hypothesis is confirmed, then interpolation of soil moisture may be better accomplished by decomposing the variability into the component patterns, interpolating those patterns separately, and then reassembling the components to produce an interpolated soil moisture pattern. Our analysis considers soil moisture data from the Tarrawarra catchment in Australia, which is described in Section 3.2. Grayson et al. [1997] have shown that this dataset exhibits the effects of both lateral flow and evapotranspiration, and it was shown in Chapter 2 herein that different EOFs roughly correspond to these distinct hydrologic processes. The decomposition of this dataset by EOF analysis is reviewed in Section 3.3. In Sections 3.4 and 3.5, the spatial properties of the EOFs are

explored. Specifically Section 3.4 compares the EOFs to various topographic attributes and Section 3.5 examines the spatial structures of the EOFs. In Section 3.6, these results are used to develop and test a new technique for spatial interpolation of soil moisture data. Section 3.7 offers conclusions.

3.2 Field Site and Data

This study utilizes the soil moisture dataset from the Tarrawarra catchment in southeast Australia [*Western and Grayson*, 1998], which is publicly available at the Global Soil Moisture Databank [*Robock et al.*, 2000]. The site consists of two small valleys and surrounding hillslopes as shown in Figure 3.1. There are no channels. Land use is pasture, and vegetation is grass with a line of Cyprus trees along the northern boundary of the site. Annual precipitation is approximately 820 mm and annual potential evapotranspiration (PET) is about 830 mm. A clear wet season occurs between April and September (austral winter) when precipitation exceeds PET, and a dry season occurs between October and March when PET exceeds precipitation. Soils are duplex with a silty loam A horizon overlying a clay B horizon. The A horizon is approximately 15-40 cm deep. Saturated conditions form in the A horizon during wet periods [*Grayson et al.*, 1997; *Western and Grayson*, 1998; *Western et al.*, 2004].

The Tarrawarra soil moisture dataset was collect using time domain reflectometry (TDR). Data were collected on 13 dates spanning approximately 14 months, and they capture the normal range of seasonal soil moisture conditions [*Western and Grayson*, 1998]. The TDR readings were taken on a 10 m x 20 m grid, where the 10 m spacing is roughly in the north-south direction. The exact locations of the observations varied

between sampling dates, but generally about 500 measurements were collected per sampling date. A typical set of observation locations are shown in Figure 3.1. The TDR measurements report the average soil moisture in volume of water per volume of soil for a cylinder of soil approximately 30 cm deep and 5 cm in diameter. Thus, in comparison to the catchment, the data are essentially point measurements. The TDR data are reported to have an error variance of approximately 2.9×10^{-4} [*Western et al.*, 1999a]. Western et al. [1999a] note that this error variance becomes a significant part of the observed soil moisture variance on dry sampling dates. For example, it is 55% of the total observed variation on 23 February.

For the purposes of our study, the TDR data were interpolated using the cubic spline method onto a regular grid with a similar spacing as the original data (20m x 10m). Figure 3.2 shows the resulting soil moisture patterns for two sample dates. The interpolation is necessary because the EOF analysis (Section 3.3) requires observations at the same locations through time. Generally measurements were taken within 0.5 m of the grid point, so the interpolation caused little change in the patterns. The mean is preserved but the variance is slightly reduced, as expected for any smooth interpolation. In addition, the dataset for each sampling date was clipped to include only the aerial extent common to all sampling dates (also shown in Figure 3.2).

A detailed topographic survey is available for the catchment with point elevations available at approximately a 10 m spacing. In order to facilitate comparisons with soil moisture, the elevation data were also interpolated using the cubic spline method to a 10 m x 10 m grid where every other column is coincident with the soil moisture grid.

3.3 EOF Decomposition

3.3.1 Method

EOF analysis is a form of principal component analysis that is applied to a space-time dataset. One can conceptualize EOF analysis as follows. A space-time dataset like soil moisture is plotted in a space where each coordinate axis is soil moisture at a particular sampling time and each data point in this space represents a particular location. Instead of describing the data in terms of these coordinates, one can also define an alternative coordinate system or basis space. In EOF analysis, this basis space is uniquely defined by the fact that the first axis explains the most variation in the original dataset, and the next axis is orthogonal to the first and explains the most residual variation. Axes are defined in this manner until all the variation is explained. One can show that the original data can be decomposed in this fashion into a set of statistically independent variables equal in number to the rank of the original dataset's covariance matrix [*Preisendorfer*, 1988]. The original data can then be described as a weighted sum of the new statistically independent variables [*Matalas and Reiher*, 1967]. EOF analysis can be used to separate distinct modes of system variation [*Jackson*, 2003], where the new variables are used to identify the modes of system variation. EOF analysis can also be used for data compression or approximation because the first d variables account for more of the variance than any other d variables [*Preisendorfer*, 1988]. Thus, one can retain only the first d variables and still retain the most important modes of variation in the dataset.

A detailed mathematical treatment of EOF analysis is given in texts on the topic [*Preisendorfer*, 1988; *Jolliffe*, 2002; *Jackson*, 2003], but a brief overview of relevant details follows. The analysis begins with a matrix of discrete space-time data:

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ s_{m1} & \cdots & \cdots & s_{mn} \end{bmatrix} \tag{18}$$

where s_{ij} corresponds to measured property s (in our case, soil moisture) at location i and time j. m is the number of sample locations and n is the number of sample times. For the clipped and interpolated Tarrawarra soil moisture data m is 459 and n is 13. Each row i in matrix S corresponds to a particular location (x_i, y_i). Note that we use capital letters to denote matrices and lowercase letters to denote scalars.

Next, a matrix of spatial anomalies Z is computed from the original data by subtracting the spatial average for each time from all measurements at that time. Specifically,

$$z_{ij} = s_{ij} - \frac{1}{m}\sum_{k=1}^{m} s_{kj} \tag{19}$$

where z_{ij} is the spatial anomaly at location i and time j. The covariance matrix Σ (n x n) is then computed as:

$$\Sigma = \frac{1}{m} Z^T Z \tag{20}$$

where T indicates the matrix transpose. The next step is to identify the characteristic roots and vectors of the covariance matrix. This is accomplished by solving:

$$\Sigma E = LE \tag{21}$$

for E and L, where E and L are related as the unit vectors and magnitudes of ΣE. E is an n x n matrix of the characteristic vectors or eigenvectors of the covariance matrix Z, and L is an n x n diagonal matrix of the corresponding characteristics roots or eigenvalues. The eigenvectors are the orthonormal axes that define the new basis space relative to the original space. The eigenvalues correspond to the variance that lies along each of the new axes. Orthonormal means that E has the following properties [*Jackson*, 2003]:

$$\sum_{i=1}^{n} e_{ij}^2 = 1 \quad \text{for } j = 1,...,n, \tag{22}$$

$$\sum_{i=1}^{n} e_{ij} e_{ik} = 0 \quad \text{for } j \neq k, \tag{23}$$

$$E^T = E^{-1} \tag{24}$$

where E^{-1} is the matrix inverse of E. Equation (22) states that each eigenvector is a unit vector, Equation (23) says that each eigenvector is orthogonal to the other eigenvectors, and Equation (24) say that the inverse of E is simply its transpose. The spatial anomaly dataset Z can be projected onto the orthonormal axes to create a new set of statistically independent variables, F. Mathematically, this is:

$$F = ZE \tag{25}$$

where F has dimensions of space-by-time (the same as Z). The terminology in the literature is inconsistent, but we define the k^{th} expansion coefficient (ECk) as the unit vector of the k^{th} axis, which is found in the k^{th} column of E in Equation (25). We define EOFk as the values of the data projected onto the k^{th} axis. EOFk is found in the k^{th} column of F in Equation (25). Each row i in F corresponds to the same location (x_i, y_i) as

in the original data matrix S. The spatial coordinates have been preserved through the data rotation, but the temporal coordinates have been replaced by the EOFs.

Preisendorfer [1988] calls Equation (25) the analysis equation because it determines the EOFs from the observations. Using Equation (24), we can rearrange Equation (25) to reconstruct the spatial anomalies as:

$$Z = FE^T , \tag{26}$$

which Preisendorfer [1988] calls the synthesis equation. This equation implies that each original spatial anomaly can be written as a linear combination of the new orthogonal variables with the linear weights provided by the eigenvectors. Stated in EOF terminology, the original data can be reconstructed as a linear combination of the EOFs weighted by the corresponding ECs. Expanding Equation (26), we obtain:

$$z_{ij} = \sum_{k=1}^{n} f_{ik} e_{kj}^T \quad \text{for} \quad i = 1,...,m, \quad j = 1,...,n \tag{27}$$

or

$$z_{ij} = \sum_{k=1}^{n} EOFk_i ECk_j \quad \text{for} \quad i = 1,...,m, \quad j = 1,...,n \tag{28}$$

where $EOFk_i$ is the i^{th} element of the k^{th} order EOF, and ECk_j is the j^{th} element of k^{th} order EC.

The eigenvalue l_{kk} in L is the variance of EOFk. The variance of the original dataset is invariant with rotation [*Preisendorfer*, 1988], so the trace of L is equal to the trace of the original covariance matrix Σ. Thus, the portion of the original variation accounted for by the new variable EOFk can be computed from:

$$p_k = \frac{l_{kk}}{\sum_{j=1}^{n} l_{jj}}. \qquad (29)$$

By convention, the EOFs are sorted in descending order by the portion of the variance explained (i.e. $p_k > p_{k+1}$). It is typical that the first d EOFs (where $d < n$) together explain a large percentage of the total dataset variance. Thus, it is possible to drop the remaining n-d EOFs and modify Equation (27) to obtain an approximation of the original data as [*Jackson*, 2003]:

$$\hat{s}_{ij} = \overline{s}_j + \sum_{k=1}^{d} f_{ik} e_{kj}^T + \varepsilon_{ij} \quad \text{for } i = 1,...,m \; j = 1,...,n \qquad (30)$$

where $\hat{s}_{ij}$ is an estimate of soil moisture at location i and time j, $\overline{s}_j$ is the spatial average soil moisture on date j, and ε_{ij} is the error.

3.3.2 Results of EOF Decomposition

The EOF analysis described above (and in Chapter 2) was applied to the Tarrawarra catchment TDR soil moisture dataset. Figures 3.3a and 3.3b show the two most important EOFs (EOF1 and EOF2), which together explain 64% of the observed soil moisture variation. EOF1 alone explains 55%. The EOFs are time invariant and occur on every sample date in accordance with the associated EC weights shown in Figure 3.3c and 3.3d. The physical significance of EOFs 1 and 2 was discussed in Chapter 2. It was noted that EOF1 has a clear topographic dependence, where the valleys have higher values than the hillslopes. This pattern is consistent with the process of lateral redistribution in which water moves from high to low elevations and accumulates in convergent zones. EC1 values are all positive, meaning that the EOF1 pattern as

shown in Figure 3.3a occurs to some extent on all dates. However, it is most important in May and October when Western et al. [1999a] found that lateral flow was most important. EOF1 also has some apparent dependence on topographic aspect, which suggests that it is also partly associated with evapotranspiration. This tendency occurs because some of the spatial variation induced by evapotranspiration occurs concurrent with the variation from lateral distribution in the original dataset. As a result, EOF analysis cannot fully distinguish these two processes. EOF2 also exhibits an aspect dependence, but the interpretation of this EOF is more complex because the EC2 values are both positive and negative. In the dry austral summer (December to February), the EC2 values are negative, which causes EOF2 to enhance the dry anomalies produced by EOF1 on the northwest-facing hillslope. In the austral winter, the EC2 values are positive which causes EOF2 to erase the aspect dependence in EOF1. In short, EOF2 serves to adjust EOF1 in response to the changing role of evapotranspiration throughout the year. Overall, EOF analysis is successful at distinguishing the roles of distinctly orthogonal spatial processes that control soil moisture.

EOF3 (Figure 3.3e) was not considered in Chapter 2, but it appears to be a predictable pattern. In particular, large values seem to be located in the lower elevations of the catchment. EC3 indicates that EOF3 is most important during the transitions between the wet and dry seasons. In the fall, the EC3 values are positive, which causes EOF3 to enhance the large soil moisture values from EOF1 at lower elevations as the catchment dries. In the spring, the EC3 values are negative, which causes EOF3 to enhance the soil moisture values from EOF1 at higher elevations as the wet season begins. This transitional behavior was previously observed by Western et al. [1999a].

EOF3 accounts for only about 6% of the total observed variation, so it has a small effect. However, EOF3 appears to have a spatial structure that is distinct from EOF1 and EOF2, and it is distinct from random noise (see Section 3.5). For comparison, Figure 3.3f shows EOF13, which is the least important EOF. This EOF explains 1% of the total variation and appears to be temporally uncorrelated noise. Thus, EOF analysis of soil moisture also seems to distinguish meaningful system variation, which mainly appears in the lower order EOFs, from random noise, which mainly appears in the higher order EOFs.

3.4 Topographic Analysis

3.4.1 Method

One of the benefits of EOF analysis is that it produces static spatial patterns that can be meaningfully compared with topographic attributes. Previous authors have compared individual soil moisture patterns with topographic attributes but have recognized the conceptual problems associated with this approach [*Western et al.*, 1999a; *Florinsky et al.*, 2002]. In this section, we examine whether strong linear relationships are present between the EOF patterns and topographic attributes. If such relationships are present, they can be exploited for interpolation (recall from Chapter 1 that the term *interpolation* will be employed here to refer to methods for filling in high resolution patterns from sparse observations).

The comparison between the EOFs and the topographic attributes is made by calculating simple and multiple correlation coefficients [*Dowdy and Wearden*, 1991]. The simple correlation coefficient measures the linear covariation between two variables. Here, it is calculated for each EOF pattern and each topographic attribute on a cell by cell

basis over the entire spatial domain. The multiple correlation coefficient compares an EOF pattern with two or more topographic attributes, measuring their combined covariance with the EOF. Standard tests for statistical significance were used to test the hypothesis that the correlations are different than zero at the 95% confidence level. Statistical significance for the simple and multiple correlation coefficients was tested using a T statistic [*Kleinbaum et al.*, 1988] and an F statistic [*Graybill*, 1976], respectively, which both assume normally distributed errors about a best-fit linear relationship between variables.

A number of commonly-used topographic attributes are considered in this analysis (Figure 3.4). These attributes are described briefly below, but the reader is referred to Florinsky et al.[2002], Western et al. [1999a], and additional references cited below for mathematical descriptions.

- *Elevation* is used as a topographic attribute because previous studies have found that it is correlated with soil moisture patterns [*Yoo and Kim*, 2004]. Elevation may be important due to its association with subtle variations in soil properties, or because it identifies points with more or less connection to stable sources of water (streams and aquifers).
- *Slope* is the gradient of the topographic surface. In practice, it is determined by finding the maximum slope between the point and its adjacent neighbors with lower elevations (i.e. the direction of steepest descent). Slope is a surrogate for hydraulic gradient [*Western et al.*, 1999a].
- *Curvature* is related to the second derivative of the function defining the topographic surface. *Vertical curvature* (k_v) is the rate of change of slope in the direction of

steepest local slope. Negative vertical curvature represents a concave-up surface. *Horizontal curvature* (k_h) can be thought of as the rate of change of topographic aspect. Negative horizontal curvature represents convergence of the topographic surface. Florinsky et. al. [2002] described zones with negative horizontal and vertical curvature as areas of flow accumulation. Alternatively, Zaslavsky and Sinai [1981] used total curvature, which is not used here.

- *Specific Contributing Area* (a) is the upslope area that drains through a unit length of contour. Here, values of a are derived using flow directions determined by routing flow from each DEM pixel to one of eight adjacent pixels in the direction of steepest descent, i.e., the D8 method. Costa-Cabral et al. [1994] and Erskine [2005] described some limitations of this method; it is possible that stronger correlations could be found by using a more refined algorithm. Once the flow directions are assigned, the specific contributing area is calculated by summing the upslope area and dividing by the length of the grid cell (10 m). The natural log of contributing area $ln(a)$ is also used herein.
- *Wetness index* (*WI*) is defined as the natural logarithm of the ratio of the specific contributing area and the local slope [*Beven and Kirkby*, 1979]. *WI* was initially proposed to identify areas of saturation by subsurface flow.
- *Potential solar radiation index* (*PSRI*) is the ratio of the solar radiation received by a sloping surface to that received by a horizontal surface at the same time of year and location on the Earth's surface [*Moore et al.*, 1993]. It is intended to identify locations that receive more or less solar radiation due to their slope and aspect. We

only consider the PSRI for 20 June, which is around the time of maximum spatial variation of solar radiation in the southern hemisphere.

- *Lowness* (L) is a measure of the elevation at a point relative to that of the local landscape [*Roberts et al.*, 1997; *Wilson et al.*, 2005]. Values of L are computed here as the average elevation for a 13 cell by 13 cell area (0.169 ha) minus the elevation of the cell at the center of the area. Assuming the water table conforms to surface topography, Roberts et al. [1997] used lowness to indicate areas where soil moisture may be higher because the water table is closer to the surface.

3.4.2 Results of Topographic Analysis

Table 3.1 shows the simple correlation coefficients between the first six EOFs and the topographic attributes. For reference, it also shows the simple correlation coefficients between the individual soil moisture patterns and the topographic attributes [*Western et al.*, 1999a]. Blank entries in Table 3.1 indicate correlations where the null hypothesis (correlation equal to zero) was not rejected at the 95% confidence level.

Several observations can be made from Table 3.1. First, the correlations confirm the interpretations of the EOFs given in Chapter 2. EOF1 has particularly high correlations with k_h, $ln(a)$, WI, and L. This collection of attributes confirms the interpretation that this EOF predominantly reflects lateral redistribution. EOF1 also has a significant negative correlation with *PSRI*, which suggests that dry locations occur where solar insolation is large. This correlation confirms the subdominant role of evapotranspiration in this pattern. EOF2 has a large positive correlation with *PSRI*, which confirms its association with evapotranspiration as well. The positive correlation

occurs because EOF2 is multiplied by negative EC2 values during the dry season. Positive EOF2 values identify locations that are driest during the dry season, and thus one expects positive EOF2 values to be associated with locations with large *PSRI* values. EOF3 exhibits a large negative correlation with elevation. A negative correlation with elevation is consistent with the previous interpretation of EOF3, specifically that it modifies EOF1 to describe transition times between the wet and dry seasons. In the late autumn months, the catchment becomes wetter, and higher soil moisture forms in the drainage headwaters (higher elevations). In contrast, the same areas have low soil moisture in the late spring as the catchment dries. This behavior is consistent with the observations of Barling et al. [1994] and Grayson et al. [1997] that the *WI* does not accurately reflect soil moisture distribution at transition times, when its assumption of steady-state flow is not satisfied. It suggests that elevation may be a useful attribute for describing the modification induced by the transient conditions.

Another important observation from Table 3.1 is the tendency of the most important EOFs to exhibit the strongest correlations with topographic attributes. The multiple correlation coefficients show that the EOFs are almost perfectly sorted according to their correlations with the topographic attributes. EOFs 1-3 have fairly high multiple correlation coefficients after which there is a sharp decline. EOF1 is significantly correlated to every topographic attribute considered. EOFs 2 and 3 are significantly correlated with several attributes, while the remaining EOFs have less topographic dependence. This result suggests that topography plays a strong role in determining the most important patterns of spatial variation in the Tarrawarra catchment.

Table 3.1 also shows that the most important EOFs have stronger correlations with the topographic attributes than the soil moisture patterns themselves. Although EOF1 is present to some extent on every sampling date, soil moisture itself is significantly correlated with all topographic attributes like EOF1 only on one date (27 September). Furthermore, EOF1 has higher correlations with 5 of the 10 attributes than soil moisture on any date. To illustrate this point, Figure 3.5 plots soil moisture on two sampling dates (13 April and 20 September) versus *WI* and *PSRI*, and it plots EOFs 1 and 2 versus *WI* and *PSRI*. The dates were chosen because EOF1 and EOF2 both play an important role on these dates. Together, EOF1 and EOF2 explain about 60% of soil moisture variation on 13 April and 65% on 20 September. This figure clearly shows that the decomposed variables EOF1 and EOF2 have stronger correlations with these two topographic attributes than the soil moisture patterns from which the EOFs were derived. Stronger correlations are obtained for the EOFs in part because the EOF decomposition sorts persistent spatial variations into lower order EOFs and noise into higher order EOFs. Because the persistent patterns underlying the soil moisture data are determined largely by topography, the lower order EOFs have particularly strong relationships with topographic attributes.

3.5 Spatial Analysis

3.5.1 Method

It has been observed that the processes controlling the variation of soil moisture can occur at distinct scales [*Kachanoski and De Jong*, 1988; *Seyfried and Wilcox*, 1995; *Hu et al.*, 1998; *Western and Grayson*, 1998]. If the EOFs distinguish different

processes, it is possible that the EOFs will exhibit different spatial structures. Furthermore, just as the EOFs showed stronger relationships with topography than the soil moisture patterns, it is also possible that they will also show more consistent spatial structures than the soil moisture patterns. To examine these points, the sample semivariogram $\gamma(h)$ is calculated for each EOF. For a dataset, the sample semivariogram is computed as [*Cressie*, 1991]:

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{m(h)} \left[z(x_i) - z(x_j) \right]^2 \tag{31}$$

where $z(x_i)$ is the observation of z at location x_i, and x_i and x_j belong to the set of $m(h)$ pairs separated by horizontal distance h. h is defined as a radial distance herein, resulting in the omni-directional semivariogram. The semivariogram measures variation of a spatially distributed property as a function of separation distance. The underlying assumption of the semivariogram is that properties at locations closer in space are more similar than those farther apart. Thus, the semivariogram function generally has low values for close distances and increases at larger distances in proportion to the decrease in spatial autocorrelation. It is important to note that the omni-directional semivariogram implicitly assumes isotropy in the field being characterized. For variables that depend on the drainage structure of topography, this assumption is clearly violated in a complex manner. In our case, some EOFs and soil moisture patterns reflect the drainage pattern while others do not, so different tools could be used for each variable in order to precisely study their structures. However, in order to produce a consistent comparison, an omni-directional semivariogram is used here in all cases. If a more refined analysis were required, one could use a directional variogram, which allows for distinct correlation

structures in pre-specified directions [*Green and Erskine*, 2004] or one could consider points spaced along flow paths [*Gottschalk*, 1993].

The sample semivariogram can be fit with a curve typically called the theoretical semivariogram. Numerous theoretical models are available [*Bárdossy and Lehmann*, 1998; *Biau et al.*, 1999; *Thattai and Islam*, 2000], but in a previous geostatistical analysis of the Tarrawarra soil moisture data, Western and Grayson [1998] found that the exponential model worked well. The exponential semivariogram γ_e can be written as:

$$\gamma_e(h) = a + (b-a)[1-\exp(-h/c)] \tag{32}$$

where a, b, and c are model parameters. The parameter a is the so-called nugget and is interpreted as the variance at zero separation distance. The nugget can be caused by instrument error or may be due to variation that occurs at a smaller scale than the sample spacing. The parameter b is the sill, which corresponds to the semivariogram's asymptotic upper limit, and is related to the sample variance. The parameter c determines the rate of increase from a to b. Western et al. [1998] define the range, i.e. the distance to the sill, as $3c$ for the exponential model. Note that the exponential model assumes that a sill exists. Some geostatistical analyses of soil moisture have found that a sill may not exist [*Green and Erskine*, 2004].

3.5.2 Results of Spatial Analysis

The sample semivariogram was computed for the EOFs and the soil moisture on each date. To facilitate comparison, all data were standardized by subtracting the mean and dividing by the standard deviation prior to computing the semivariograms. Figure 3.6 shows the sample variograms for the first three EOFs and the soil moisture patterns

on 3 July and 20 September. These two dates were selected because multiple EOFs are important on these dates. EOFs 1-3 together account for about 75% of variance on 3 July and more than 65% on 20 September. The exponential model was fit to each sample semivariogram, and Table 3.2 shows the estimated parameters. The parameters for the *WI* and *PSRI* semivariograms are also provided for comparison. Western et al. [1998] previously published semivariogram parameters for soil moisture patterns on all dates.

Figure 3.6 shows that the first three EOFs explain spatial variation over different ranges of spatial scales. EOF1 rapidly rises to a relatively constant value for large separation distances (its sill), which suggests that the correlations in EOF1 occur at relatively small spatial scales. EOF2 rises a bit more slowly indicating longer correlation distances, while EOF3 rises very slowly and does not reach the model sill of 1.2. These results are quantified by the parameters in Table 3.2. In particular, the range for EOF1 is 94 m, the range for EOF2 is 116 m, and the range for EOF3 is 370 m. These ranges are also related to those of the related topographic attributes. For example, *WI* has a range of 101 m, which is similar to EOF1, and *PSRI* has a range of 214 m, which influences the somewhat larger range of EOF2. In both cases, the EOF ranges are smaller because the EOFs depend on other topographic attributes with shorter ranges (for example, k_h which has a range of 75 m). It is interesting to note that the semivariograms of soil moisture resemble the semivariograms of the EOFs that are important on those dates. For example, the semivariograms for 3 July and EOF3 are similar because EOF3 explains nearly 30% of the variation on this date.

Figure 3.6 also suggests that much of the noise in the soil moisture data has been removed from the most important EOFs. The semivariogram value for EOF1 is quite

small at small separation distances, which results in a nugget of zero. This result hints that much of the measurement error might have been removed from this EOF. The nugget tends to increase as one considers EOFs 2 through 7. It is possible that this small scale variation is due to soil variations. For example, Wilson et al. [2004] attributed small scale soil moisture variations that were not explained by topography to soil properties. However, as one considers high order EOFs, the amount of variation explained by each EOF is becoming smaller. Thus, it becomes increasingly difficult to distinguish meaningful system variation from random noise. Thus, we expect that the EOFs are increasingly mingled with noise. In fact, the semivariogram parameters estimated from EOFs 8 and higher are consistent with parameters we estimated from simulated fields of uncorrelated noise, while the parameters for EOFs 6 and 7 are marginally similar. These results suggest that as many as 7 EOFs may describe system variation, even though only the first few are significantly correlated with topography.

The implicit assumption of a semivariogram is that the analyzed pattern can be well characterized as a function of separation distance. The better this assumption holds, the better distance-based interpolation methods will perform. To test this assumption, we note that the semivariogram plots the mean of the squared difference of data values as a function of separation distance. One can also consider the deviation of the squared differences about the mean. Specifically, we calculate:

$$RMSE = \sqrt{\frac{1}{m_{pairs}} \sum_{m_{pairs}} \{[z(x_i) - z(x_j)]^2 - \hat{\gamma}(h)\}^2} \qquad (33)$$

where z, x_i, x_j, h, and $\hat{\gamma}$ were defined in Equation (31) above and m_{pairs} is the total number of pairs (x_i, x_j) used to compute the sample variogram. *RMSE* is the root mean

square error of the differences about the sample semivariogram curve. If the spatial structure is well described by a function of separation distance, one expects a lower *RMSE* about the sample semivariogram. Note that *RMSE* here is measured about the sample semivariogram in order to measure consistent spatial behavior, *not* about any particular model semivariogram to measure goodness of fit. Table 3.2 shows that lower *RMSE* values are observed for the most important EOFs than for the soil moisture patterns themselves. Table 3.2 shows that the *RMSE* for EOF1 is higher than the *RMSE* for higher order EOFs. This is likely due to the fact that the omni-directional variogram does not do a good job of describing the drainage related pattern of EOF1, as was discussed in Section 3.5.1. The point here is that the EOFs exhibit lower *RMSE* values than the soil moisture patterns from which they were derived. The simpler structure of the EOFs suggests that they may be easier to interpolate than the soil moisture patterns.

3.6 Spatial Interpolation Using EOFs

3.6.1 Method

The results of the previous sections all hint that interpolating in the EOF domain may produce better estimates of soil moisture than interpolating soil moisture itself. Section 3.3 demonstrated that soil moisture patterns could be decomposed roughly according to the controlling processes using EOF analysis. Section 3.4 found that the most important EOFs have stronger correlations with topography than the soil moisture patterns themselves, and Section 3.5 found that the spatial structures of the most important EOFs are better described using variograms than the soil moisture patterns. Next, we turn our attention to interpolating soil moisture with EOFs.

Interpolation and estimation using EOFs has been done successfully in other applications. Rao and Hseih [1991] decomposed temporal variation of streamflow data using EOF analysis and used interpolated EOFs to construct streamflow time series for nearby ungaged basins. Liu [2003] used EOF analysis to decompose soil moisture and precipitation variation and found that the resulting ECs exhibited higher cross-correlations than the original data. He proposed monitoring soil moisture and using the lower order ECs to forecast precipitation. Biau et al. [1999] created a three-dimensional space from precipitation and the first two ECs for sea level pressure. For a given location, they interpolated precipitation in time using ECs estimated from modeled variations in the sea level pressure. An EOF method of soil moisture estimation was developed in Chapter 2 herein. In that method, EOFs of soil moisture were determined from a space-time dataset with observations available at every location. Then the ECs were estimated for an unobserved time using empirical relationships to the spatial average soil moisture. Observed EOFs, estimated ECs and the spatial average were used to estimate a soil moisture pattern at an unobserved time.

In this section, we propose a new EOF-based method to estimate soil moisture at unobserved locations using sparse observations from the current time and at least one previously observed time. The first step of the proposed interpolation technique is to perform an EOF analysis on the sparse soil moisture data using the method described in Section 3.3. The ECs are computed from Equation (21) and the EOFs are computed at the observed locations using Equation (25). In order for the EOF analysis to be non-trivial, two or more observation dates must be available. In part, the success of the EOF interpolation method depends on how well EOFs and ECs can be identified from sparse

data. Figure 3.7 shows EOFs 1 and 2 and ECs 1 and 2 computed from a sparse set observations, which were produced by sampling the original 10 m x 20 m data uniformly to produce data on a 30 m x 60 m grid. This sampling procedure keeps every third data point in each direction or 11% of the original observations (Table 3.3). The figure also shows the EOF and EC values at the same locations if they are generated from the entire dataset. The figure shows that the EOF1 values are reproduced almost exactly from the sparse data. EC1 is reproduced well, but some deviation is observed for certain dates. EOF2 and EC2 are reproduced fairly well, but apparently the covariance pattern identified by EOF2 is more difficult to distinguish from other variation when the observations are sparse. In general, it is expected that higher order EOFs are increasingly difficult to recognize from limited data and perhaps should be excluded from the interpolation method (see below). However, the most important EOFs and ECs are identifiable from sparse observations.

The next step in the interpolation method is to interpolate the EOFs over the entire spatial domain. Recall that the ECs are spatially invariant and should therefore apply to all locations. However, the EOFs must be interpolated to gain values at unobserved locations. One can interpolate the EOFs with any method that could be used directly on soil moisture data, but we consider two distinct methods. The first is a regression-based method that exploits the observed correlations between topographic attributes and EOFs (Section 3.4). In this method, elevation data are assumed to be available for the entire domain at the final resolution of the interpolation method, and topographic attributes are computed as described in Section 3.4.1. A standard stepwise partial correlation analysis (Salas et al., manuscript in preparation, 2004) is then used to build a parsimonious

multiple linear regression (MLR) model. In the partial correlation analysis, the marginal increase in the multiple correlation is checked at each step and only variables that produce a statistically significant improvement are added to the regression model. The result is a regression model with the following form:

$$\hat{f}_{ik} = a_k + \sum_{j=1}^{l} T_{ij} b_{jk} \qquad \text{for } i=1,\ldots,m \tag{34}$$

where $\hat{f}_{ik}$ is the interpolated k^{th} order EOF at location i, a_k and b_{jk} are the parameters determined from the regression of the sparse k^{th} order EOF against topographic attributes T_j, l is the total number of topographic attributes included in the model, and m is the total number of locations. The second interpolation method used here is the inverse distance weighted (IDW) scheme, which exploits the simplified spatial behavior of the EOFs demonstrated in Section 3.5. The IDW method determines $\hat{f}_{ik}$ from:

$$\hat{f}_{ik} = \sum_{j=1}^{l} w_j f_{jk} \qquad \text{for } i=1,\ldots,m \tag{35}$$

where f_{jk} is the value of the k^{th} order EOF at observation location j and l is the number of neighboring locations used in the interpolation. In this paper, we use $l=5$, but the results shown below are insensitive to this choice. w_j is the weight applied to each neighbor. The weights are computed as:

$$w_j = \frac{h_j^{-2}}{\sum_{ll=1}^{l} h_{ll}^{-2}} \tag{36}$$

where h_j is the horizontal distance between locations i and j. More complex methods could be used such as kriging [*Biau et al.*, 1999] or co-kriging which considers both the topographic dependences and spatial structures of the EOFs.

After the EOFs have been interpolated, soil moisture can be estimated for the entire spatial domain using the ECs from the observations and the interpolated EOFs. The soil moisture is found from a variation of Equation (30). In particular,

$$\hat{s}_{ij} = \overline{s}_j + \sum_{k=1}^{d} \hat{f}_{ik} e_{kj}^T \quad \text{for} \quad i = 1,...,m \quad j = 1,...,n \tag{37}$$

where $\hat{s}_{ij}$ is an estimate of the soil moisture at location i on date j, $\overline{s}_j$ is the average of the sparse samples on date j, and e_{kj} is the k^{th} order EC computed from the sparse observations. Notice that the number of EOFs included d is allowed to be less than the total number of EOFs n. The intent here is to retain only the EOFs that contain mostly system variation while disregarding the higher order EOFs that contain mostly noise.

Many authors have studied the number of EOFs that should be considered statistically significant and thus retained in EOF analysis. Here, we consider three different methods. North et al. [1982] devised a method based on confidences limits for the eigenvalues associated with the EOFs. For Gaussian random variables, they determined +/- 1 standard deviation confidence limits Δl_j for the eigenvalue l_{jj} as:

$$\Delta l_j = l_{jj} \left(2/m\right)^{1/2} \tag{38}$$

North et al. [1982] suggested that if the confidence limits of an eigenvalue (l_{jj}-Δl_j, l_{jj}+Δl_j) do not overlap with those for the next higher order eigenvalue, then l_{jj} can be considered statistically significant. This argument is based on the observation that a dataset without significant covariation will exhibit the same amount of variation in all directions. As a result, the EOF coordinate transformations are not unique and the eigenvalues are expected to be nearly the same. The second approach was developed by Bartlett [1950],

who judged significance by considering the hypothesis that the eigenvalues of the last (*n-d*) EOFs are all equal. The relevant statistic is χ^2_{crit}, which is calculated [*Jackson*, 2003]:

$$\chi^2_{crit} = -\nu \sum_{j=d+1}^{n} \ln(l_j) + (\nu)(n-d)\ln\left[\frac{\sum_{j=d+1}^{n} l_j}{(n-d)}\right] \qquad (39)$$

where ν is the degrees of freedom of the covariance matrix Σ (Equation (20)), which is the number of sample locations minus one. Bartlett [1950] showed that χ^2_{crit} has a chi-squared distribution with (1/2)(*n-d*-1)(*n-d*+2) degrees of freedom. If χ^2_{crit} is less than the standard tabulated χ^2 variate with (1/2)(*n-d*-1)(*n-d*+2) degrees of freedom at the selected confidence level, then the hypothesis is accepted and the last (*n-d*) EOFs are discarded. The third method was proposed by Cattell [1966] and is a simple graphical test called the scree test. The eigenvalues or their logarithms are plotted sequentially in a graph. If a breakpoint is observed in the plot, all eigenvalues beyond this breakpoint are considered to be insignificant [*Jackson*, 2003].

Figure 3.8 demonstrates these three tests on the eigenvalues generated from the 30 m x 60 m grid of observations described earlier. According to the North et al. [1982] test, EOF1 is clearly significant at this spacing and EOF2 is marginally significant, whereas the Bartlett [1950] chi-squared test indicates that EOFs 1-5 are significant at the 95% confidence level. The Cattell [1966] scree test indicates that EOFs 1-3 are significant. Note that the Bartlett [1950] test uses the 95% confidence level, while the North et al. [1982] test uses +/- 1 standard deviation, which is the 68% confidence level. If both tests were applied at the same confidence level, the Bartlett et al. [1950] test would retain far more EOFs. For all of these methods, the number of significant EOFs depends on the

number of observations and thus on the spacing of the observations for a given catchment. Generally, smaller numbers of significant EOFs are identified as a smaller number of samples are used. Based on these tests, it seems that 3 to 4 EOFs should be retained for this spacing. Later, the sensitivity of the soil moisture interpolation results to the number of EOFs retained will be examined.

3.6.2 Results of Spatial Interpolation

Figure 3.9 shows examples of interpolated soil moisture patterns using the EOF-based method with observations spaced 30 m x 60 m. In this case, 52 observations from each of the 13 sampling dates were used for the EOF analysis, and 407 points were subsequently interpolated on each date. Figure 3.9a shows the 3 July soil moisture pattern computed from EOFs 1-3, which were interpolated using the IDW method. Comparing the interpolated pattern to the observed pattern in Figure 3.2a, the large-scale system variation seems to be generally reproduced. The upper end of the valley on the right side is wet, and the northwest-facing slope is dry. The gully on the left side is somewhat wet, and the northern boundary of the catchment along the stand of Cyprus trees is dry. The interpolated pattern is clearly smoother than the observed pattern, which is expected due to the loss of small scale variation associated with the higher order EOFs and the use of the IDW method. Figure 3.9b shows the 20 September soil moisture pattern computed from EOFs 1-3, which were interpolated from a MLR against topographic attributes. The interpolated pattern can be compared to the observed pattern in Figure 3.2b. Much of the observed dependence on topographic features is reproduced in the interpolated surface. In particular, the left and right gullies are much wetter than

the hillslopes. Again, the interpolated pattern is smoother than the observed pattern. Notice that the interpolated patterns in Figure 3.9a and 3.9b appear much different from each other. This difference arises mostly from the different weightings of the underlying EOF patterns. Figures 3.9c and 3.9d show the estimation errors for the 3 July and 20 September patterns, respectively. In both cases, the errors do not appear to depend on topography and exhibit mostly small scale variations.

Interpolation performance can be measured with the average Nash Sutcliffe Coefficient of Efficiency (*NSCE*) among the 13 days in the dataset, computed as:

$$NSCE = \frac{1}{n}\sum_{i=1}^{n}\frac{\sigma^2_{obs,i} - \sigma^2_{\varepsilon,i}}{\sigma^2_{obs,i}} \tag{40}$$

where $\sigma^2_{obs,i}$ is the variance of the observed soil moisture pattern on day i and $\sigma^2_{\varepsilon,i}$ is the mean squared estimation error on day i. The term in the summation is the usual definition of the *NSCE* [*Nash and Sutcliffe*, 1970]. $\sigma^2_{obs,i}$ and $\sigma^2_{\varepsilon,i}$ are computed only at locations that are not part of the observations supplied to the interpolation method. If the *NSCE* is one on a particular day, then the interpolated surface reproduces the unobserved values exactly. If the *NSCE* is zero, then the interpolated surface performs as well as the true spatial average of the soil moisture would.

Figure 3.10 evaluates the performance of the EOF interpolation method as a function of the number of EOFs included in the method. To generate Figure 3.10, observations were again supplied at a 30 m x 60 m spacing. However, notice that nine different datasets can be generated with this uniform spacing from the original 10 m x 20 m dataset depending on the particular points that are retained (Table 3.3). In Figure 3.10, the data points show the average *NSCE* (including all dates) among these nine

realizations. The error bars show the standard deviation of the *NSCE* for the nine realizations. Figure 3.10a evaluates the performance when the MLR approach is used to interpolate the EOFs. Here, the *NSCE* improves as the number of EOFs increases to 3 or 4. Then, as additional EOFs are added, the average *NSCE* among the nine realizations remains around 0.27. The dotted line in Figure 3.10a shows the multiple correlation coefficient between each added EOF and the topographic attributes from Section 3.4. Higher order EOFs generally have lower multiple correlation coefficients with topographic attributes. EOFs with insignificant correlations with topography are dropped from the regression model during the partial correlation analysis, so higher order EOFs are not necessarily used to estimate the soil moisture. Thus, there are essentially redundant significance tests if the MLR method is used, which is why retaining insignificant EOFs does not affect the *NSCE* in this case. This result suggests that it is better to retain too many EOFs in this case than to retain too few. Figure 3.10b shows the *NSCE* results if the IDW method is used for EOF interpolation. In this case, the average *NSCE* among the nine realizations peaks at about 0.25 when 3 to 4 EOFs are included. When more EOFs are added, a clear decline in performance is observed. For the IDW method, the only test for statistical significance is the decision to include or discard EOFs, so it is imperative that the EOFs containing noise are eliminated.

Figure 3.11 evaluates the performance of the EOF-based interpolation method as a function of the spacing of the observations and compares it to methods that directly interpolate soil moisture. For each uniform spacing shown in the plot, every possible set of observations was interpolated (see Table 3.3). The average *NSCE* (including all dates) was calculated among these datasets and is plotted as a function of the spacing in the

figure. The error bars show the standard deviation of the *NSCE* among the realizations at each spacing. The figure shows the results of the EOF-based method using both the MLR and IDW approaches to interpolate the EOFs, which we refer to as the EOF-MLR and EOF-IDW methods, respectively. For comparison the figure shows the results for the MLR and IDW interpolation methods applied directly to the soil moisture data. In all cases, the resolution of the interpolated surface is 10 m x 20 m. It should be noted that the data requirements differ between the various methods. The IDW method uses only sparse soil moisture data. The MLR uses both the sparse soil moisture data and fine-scale elevation data. Both EOF-based methods (IDW and MLR) use the sparse soil moisture observations on all 13 sampling dates to compute the EOFs, which are then interpolated to produce the fine-scale soil moisture patterns. Both direct interpolation methods use sparse observations only from the individual sampling date being interpolated. The effects of these varying data requirements are investigated below.

The closest spacing of observations considered by Figure 3.11 is 10 m x 20 m, which is equivalent to the resolution of the interpolated surface. In this case, the MLR and EOF-MLR methods simply estimate soil moisture at the same locations used to develop the regressions. Likewise, the IDW and EOF-IDW methods are employed to estimate soil moisture at each point assuming that all other points are known. The 10 m x 20 m case gives an assessment of the maximum ability of the interpolation methods to reproduce the unobserved observations. One can see from the *NSCE* in Figure 3.11 that a maximum of about 40% of the observed variation can be reproduced. As a check, this maximum *NSCE* is approximately the same as the squares of the multiple correlation coefficients from Table 3.1 averaged over the 13 sampling dates, which corresponds to

the amount of soil moisture variation that can be explained by topography. The unexplained variation likely includes the measurement error. Recall that the TDR data has an error variance of about 2.9×10^{-4} [*Western et al.*, 1999a], which is 25% of the average observed variance among the 13 sampling dates. The remaining 35% is system variation that is not captured by the interpolation methods used here. The IDW methods have the potential to reproduce slightly more of the observed variation than the MLR methods at the 10 m x 20 m spacing, because not all of the system variation is related to topography.

As the spacing increases, the performance of all the methods decreases, but the EOF-based interpolation methods always out-perform their respective direct interpolation methods. This behavior was explained in Sections 3.4 and 3.5. Lower order EOFs were shown to be sorted roughly by the strength of their topographic dependence and were shown to exhibit more consistent dependence on topographic attributes than the soil moisture patterns. When topographic-related variation is partitioned from other variation in the EOF analysis, more consistent regression relationships with topography can be identified, hence the MLR-EOF method outperforms the direct MLR method. Likewise, the EOFs were shown to have distinct and more consistent variogram functions than soil moisture. Therefore, a distance-based interpolation method like IDW is better able to interpolate the EOFs as a function of distance than soil moisture itself.

The relative performance of the different methods changes as the spacing increases. At the 20 m x 40 m sample spacing, the EOF-IDW interpolation method outperforms the EOF-MLR method. At this scale, neighboring soil moisture data contain more information about the local soil moisture than local topographic data do. However,

this behavior changes at the 30 m x 60 m spacing because the information content of neighboring soil moisture data decreases as the sample spacing increases. Furthermore, the assumption that topographic data is available at high resolution means that the MLR methods can use more information than the IDW methods. Thus, the EOF-IDW method is a better choice for closely-spaced observations, while the EOF-MLR is a better choice for widely-spaced observations, when topographic data is available at the target resolution of soil moisture.

Table 3.4 compares the estimation errors from the EOF-MLR method and those from the direct MLR method. In both cases, the interpolated soil moisture patterns were generated from a single 30 m x 60 m set of observations, and the estimation errors were computed for each date. Then, the correlations were calculated between the estimation errors at the locations that were not provided to the interpolation method and the topographic attributes. On 10 out of 13 dates, the errors from the EOF-MLR method have a lower multiple correlation coefficient with the topographic attributes than the errors from the direct MLR method. This result suggests that the EOF-MLR method accounts for the topographic influence better than the direct MLR method. This is related to Figure 3.4 where it was shown that the lower order EOFs have stronger correlations with topography due to the fact that noise has been partitioned into higher order EOFs.

It was noted previously that the EOF-based interpolation methods use more data than the direct methods. Specifically, the EOFs are generated based on the sparse observations from multiple sampling dates. In contrast, the direct MLR and IDW only consider the sparse data from the current date. To assess the role that the additional data plays, the *NSCE* is evaluated for one sampling date (28 March) as successive dates are

added to the dataset that is used for the MLR and EOF-MLR techniques (Figure 3.12). The successive dates are simply added in chronological order. Only the direct MLR and EOF-MLR methods are compared because the direct IDW method cannot easily include data from multiple dates. For the MLR method, data from each additional date is normalized and standardized to remove the different means and standard deviations for each day, which produces a better regression. Figure 3.12 shows that the MLR and EOF-MLR methods yield identical results when only one date is used. In this case, the only EOF pattern is the soil moisture pattern itself. However, when two sample dates are used in each method, the *NSCE* for the EOF-MLR method immediately doubles from 0.075 to about 0.15, while the *NSCE* for the direct MLR method actually decreases. The *NSCE* in both cases does not change significantly with the addition of three or more observation dates. The analysis was repeated numerous times by adding the successive dates in different orders with the same general results. This analysis implies that the difference in the estimation efficiencies of the MLR and EOF-MLR methods is not simply due to a difference in the amount of data required. Instead, the difference is due to the method by which that data is interpreted. The analysis also shows that the EOF-based method is beneficial as soon as two sampling dates are used because two sampling dates immediately allow the method to separate some system variation from noise.

3.7 Conclusions

The main conclusions of this study are as follows:

(1) EOF analysis is effective at partitioning soil moisture variation roughly along the lines of different hydrologic processes and partitioning system variation from random

noise. The most important EOF (EOF1) mainly reflects the spatial pattern induced by lateral flow, the second most important EOF (EOF2) reflects the modification to EOF1 due to the changing role of evapotranspiration during the year, and the third most important EOF (EOF3) modifies EOF1 at transition times when drainage conditions are far from steady-state. Higher order EOFs (EOF8 and higher) are consistent with random noise.

(2) The most important EOFs generated from the soil moisture data have higher correlations with topographic attributes than individual soil moisture patterns. The higher correlations occur because the most important EOFs identify persistent patterns of covariation, which are expected to be associated with static variables like topography.

(3) The spatial structures of the most important EOFs are better described as a function of separation distance than the soil moisture patterns. This result occurs because each of the most important EOFs have distinct relationships with separation distance whereas the soil moisture patterns blend together these distinct relations. It also occurs because random noise is partitioned into the higher order EOFs.

(4) If data are available from more than one observation date, then interpolation of soil moisture data is better done by conducting an EOF analysis and interpolating the EOFs than by direct interpolation of the soil moisture observations. The improved efficiency of the EOF-based method occurs because the EOFs have stronger relationships with topographic variables and simpler spatial structures than the soil moisture patterns.

(5) If the available soil moisture observations are closely-spaced, then the EOFs are better interpolated with a distance-based method. If the observations are widely-spaced and fine-scale topographic data is available, then the EOFs are better interpolated with a multiple linear regression against fine-scale topographic attributes.

Table 3.1 Comparison of EOF and soil moisture correlations with topographic attributes. Blank entries correspond to statistically insignificant correlations at the 95% confidence level.

	Simple Correlation Coefficient										Multiple Correlation Coefficient
	Elev.	slope	slope^{-1}	k_h	k_v	a	$ln(a)$	WI	PSRI	L	
EOF1	-0.39	-0.26	0.39	-0.58	-0.36	0.40	0.67	0.67	-0.29	0.67	0.82
EOF2		-0.40	0.40	-0.13		0.20		0.12	0.56		0.63
EOF3	-0.33	-0.19	0.22			0.16	0.15	0.19	-0.17		0.50
EOF4		0.11		0.13				-0.10			0.15
EOF5	0.15				0.10						0.20
EOF6		-0.09		-0.11	-0.13						0.13
27 Sep	-0.41	-0.23	0.35	-0.52	-0.33	0.36	0.65	0.64	-0.38	0.62	0.82
14 Feb	-0.10	0.28	-0.23		-0.18		0.14		-0.40	0.18	0.48
23 Feb		0.20	-0.12				0.11		-0.31	0.13	0.36
28 Mar	-0.21			-0.23	-0.23		0.27	0.21	-0.38	0.31	0.50
13 Apr	-0.16			-0.36	-0.31	0.12	0.41	0.35	-0.38	0.43	0.58
22 Apr	-0.22	-0.29	0.37	-0.50	-0.26	0.33	0.52	0.54		0.51	0.62
2 May	-0.35	-0.30	0.40	-0.63	-0.34	0.37	0.63	0.64	-0.18	0.64	0.77
3 Jul		-0.14	0.21	-0.39		0.22	0.31	0.32		0.32	0.51
2 Sep	-0.30	-0.24	0.38	-0.35	-0.24	0.34	0.43	0.46	-0.13	0.46	0.57
20 Sep	-0.33	-0.34	0.40	-0.42	-0.25	0.35	0.43	0.47		0.45	0.56
25 Oct	-0.41	-0.24	0.35	-0.45	-0.30	0.40	0.63	0.63	-0.30	0.57	0.74
10 Nov	-0.35	-0.18	0.29	-0.40	-0.27	0.31	0.51	0.51	-0.46	0.49	0.75
29 Nov	-0.22		0.18	-0.36	-0.21	0.21	0.43	0.40	-0.32	0.46	0.60

Table 3.2 Estimated parameters of the exponential semivariogram model for the most important EOFs, two soil moisture patterns, and selected topographic attributes.

Spatial Pattern	Nugget, a $(V/V)^2$	Sill, b $(V/V)^2$	Range, $3c$ (m)	RMSE
EOFs				
1	0.00	1.0	94	1.61
2	0.25	1.0	116	1.35
3	0.50	1.2	370	1.54
4	0.58	1.4	536	1.43
5	0.71	1.1	264	1.51
6	0.53	1.0	64	1.49
7	0.68	1.0	71	1.46
Soil Moisture				
3 July	0.37	1.1	138	1.81
20 Sep	0.17	1.0	62	1.81
Topographic Attributes				
WI	0	1.1	101	--
PSRI	0	1.2	214	--
k_h	0	0.98	75	--

Table 3.3 Characteristics of the datasets used for interpolation.

Observation Spacing	Number of Observations	Percent of Original Observations Retained	Number of Realizations Available for Analysis
10 m x 20 m	459	100	1
20 m x 40 m	113	25	4
30 m x 60 m	52	11	9
40 m x 80 m	30	6.5	16
50 m x 100 m	19	4.1	25
60 m x 120 m	12	2.6	36

Table 3.4 Multiple correlation coefficients between soil moisture estimation errors and topographic attributes. The estimation errors are from interpolated soil moisture from 30 m x 60 m observation spacing. The topographic attributes used include: elevation, slope, slope^{-1}, k_h, k_v, a, $ln(a)$, WI, PSRI, and L. Only EOFs 1-3 were retained for the EOF-based estimate shown here.

Soil Moisture Sampling Date	EOF-MLR	Direct MLR of Soil Moisture
9/27	0.21	0.12
2/14	0.17	0.13
2/23	0.11	0.14
3/28	0.12	0.25
4/13	0.14	0.22
4/22	0.16	0.22
5/2	0.23	0.22
7/3	0.18	0.21
9/2	0.08	0.27
9/20	0.09	0.18
10/25	0.20	0.42
11/10	0.17	0.20
11/29	0.15	0.27

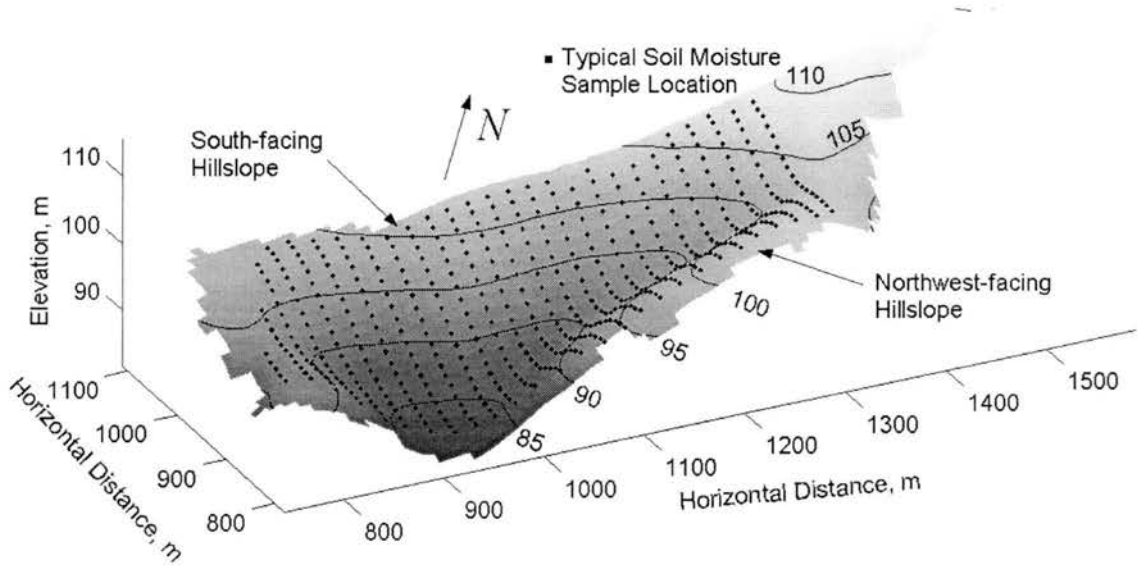

Figure 3.1 Tarrawarra catchment topography. The overlaid dots show the soil moisture sampling grid for a typical observation date. Note that the vertical scale is exaggerated (approximately 6x).

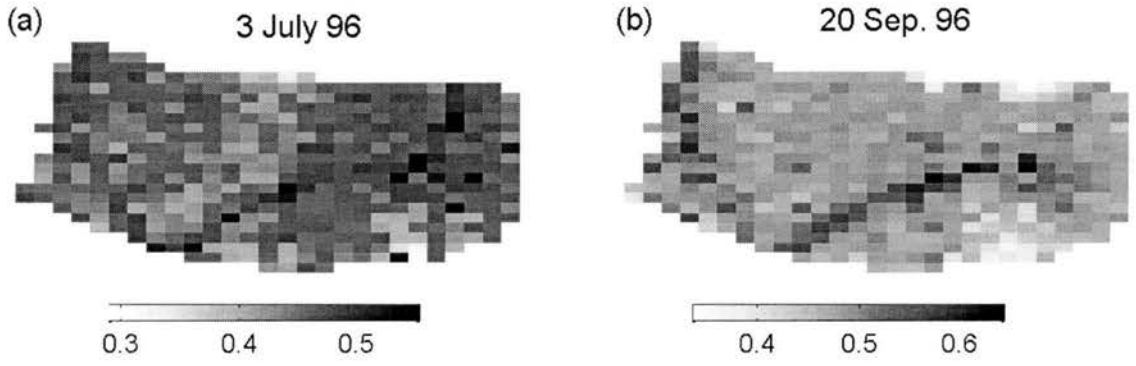

Figure 3.2 Spatial patterns of soil moisture on (a) 3 July 1996 and (b) 20 September 1996. The units are in volume of water per volume of soil.

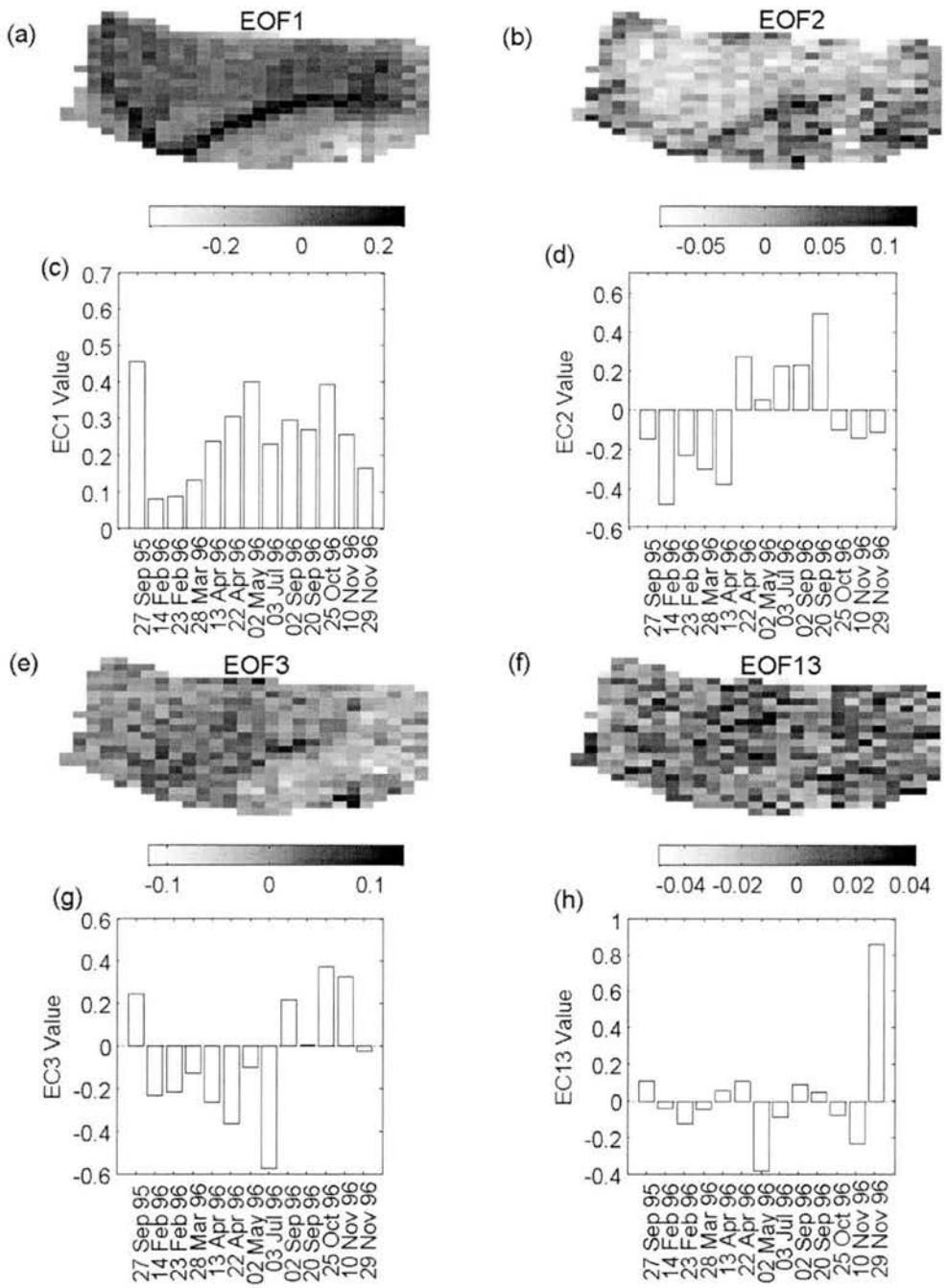

Figure 3.3 The spatial patterns of covariation (a) EOF1 and (b) EOF2 and the corresponding time-varying weights (c) EC1 and (d) EC2. Spatial patterns (e) EOF3 and (f) EOF13 and time-varying weights (g) EC3 and (h) EC13.

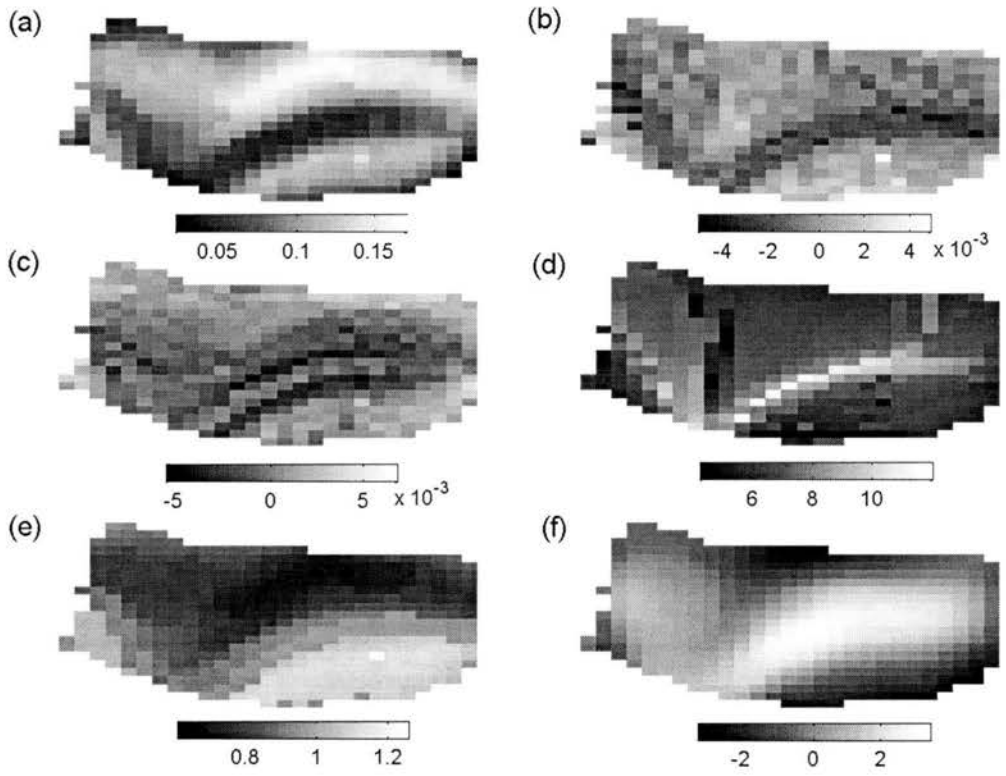

Figure 3.4 The topographic attributes: (a) slope, (b) horizontal curvature, (c) vertical curvature, (d) wetness index, (e) potential solar radiation index, and (f) lowness.

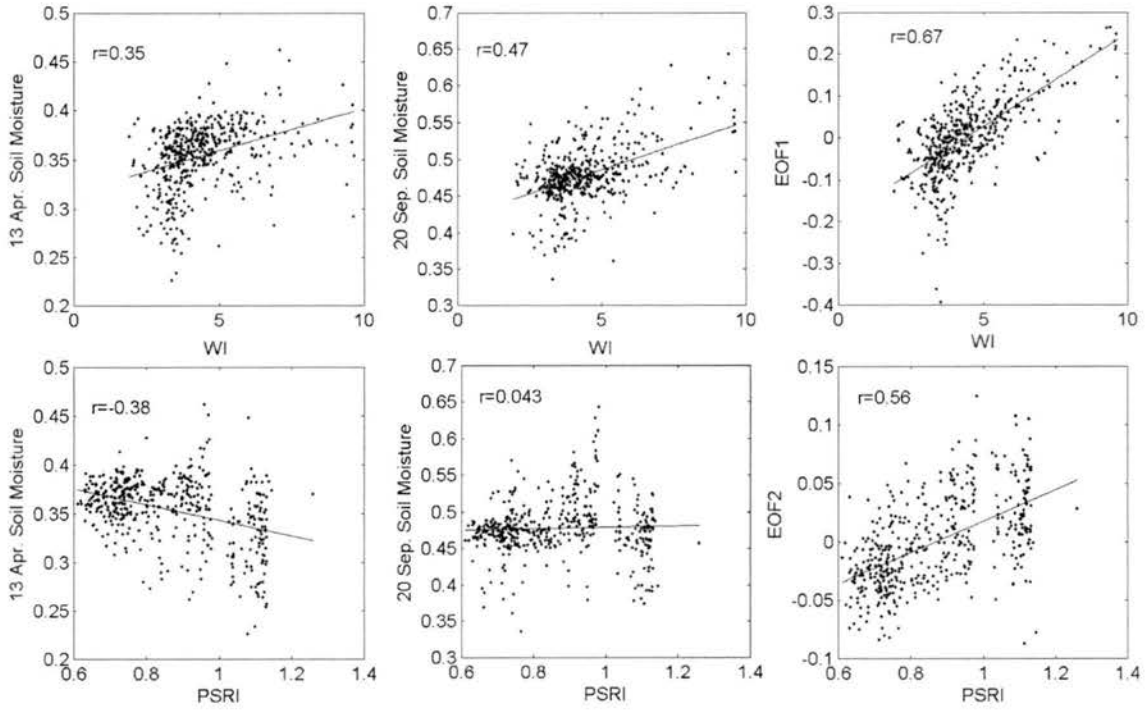

Figure 3.5 Comparison of the relationships between soil moisture on various sampling dates and two topographic attributes (PSRI and WI) to the relationships between EOFs 1-2 and the same topographic attributes.

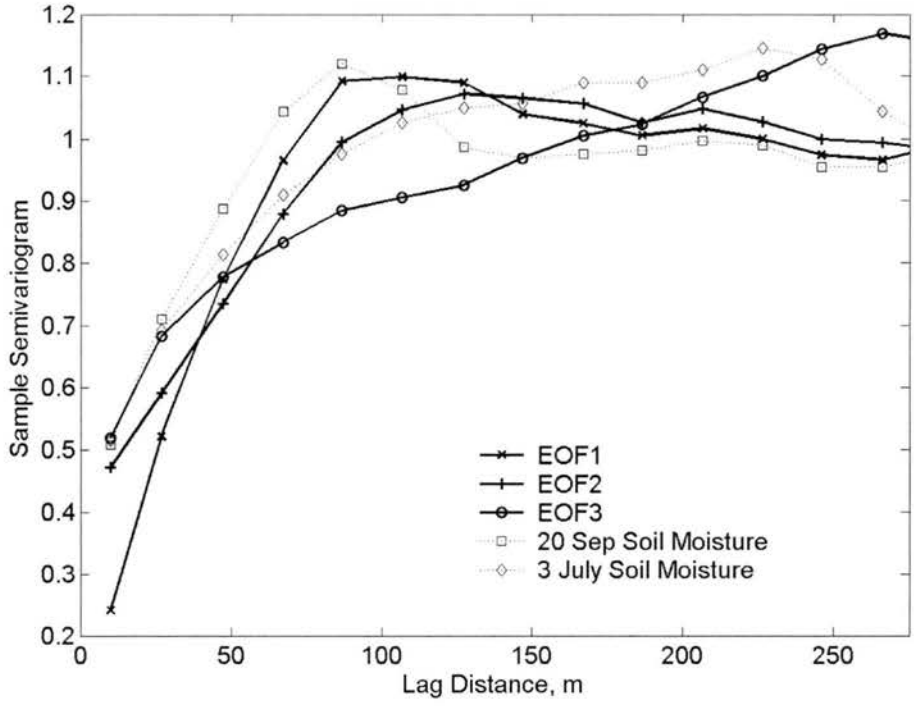

Figure 3.6 Comparison of the sample semivariograms for EOFs 1-3 and the soil moisture patterns on two observation dates.

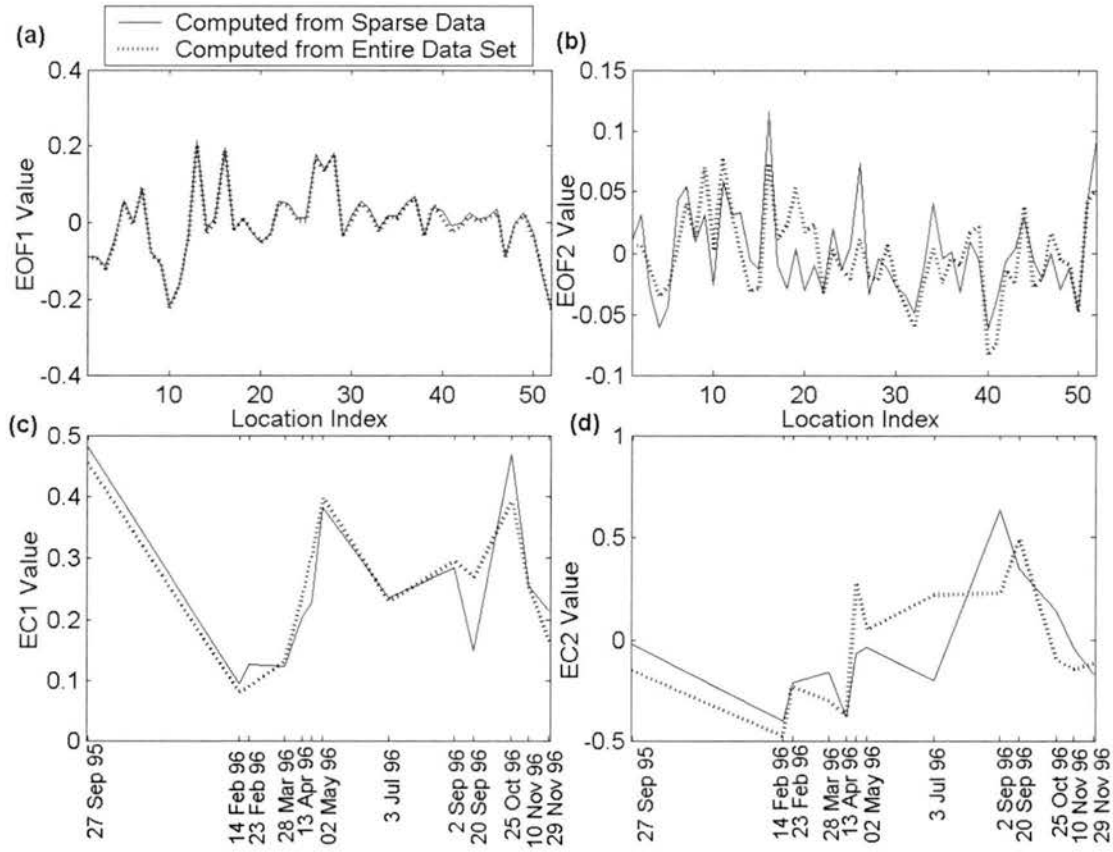

Figure 3.7 Comparison of the EOFs and ECs determined from the sparse sample data (30 m x 60 m spacing) and the corresponding EOFs and ECs determined from the entire dataset.

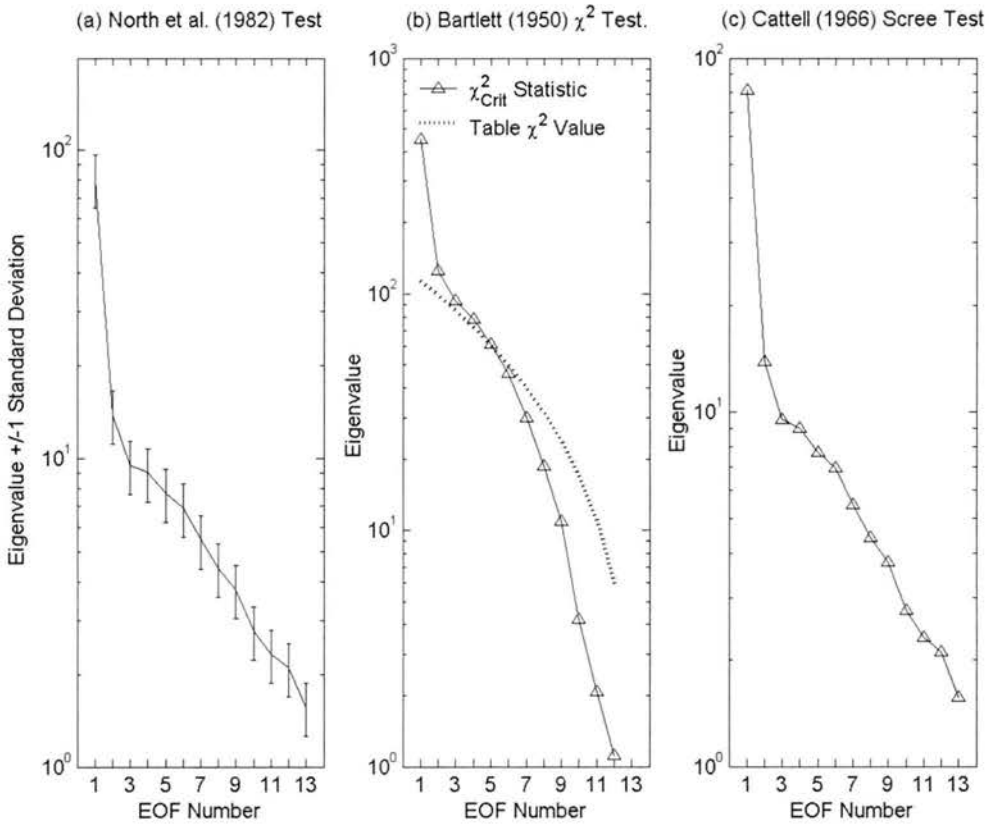

Figure 3.8 Comparison of three tests for statistical significance of EOFs. (a) North et al. [1982] test using the 68% confidence limits on the eigenvalues. Non-overlapping confidence limits indicate a statistically significant EOF. (b) Bartlett [1950] chi-squared test of the hypothesis that a given eigenvalue is equal to all higher order eigenvalues. (c) Cattell [1966] scree test, where a break point in the slope of plot is used to identify significant eigenvalues . All results shown here are for 30 m x 60 m sample spacing.

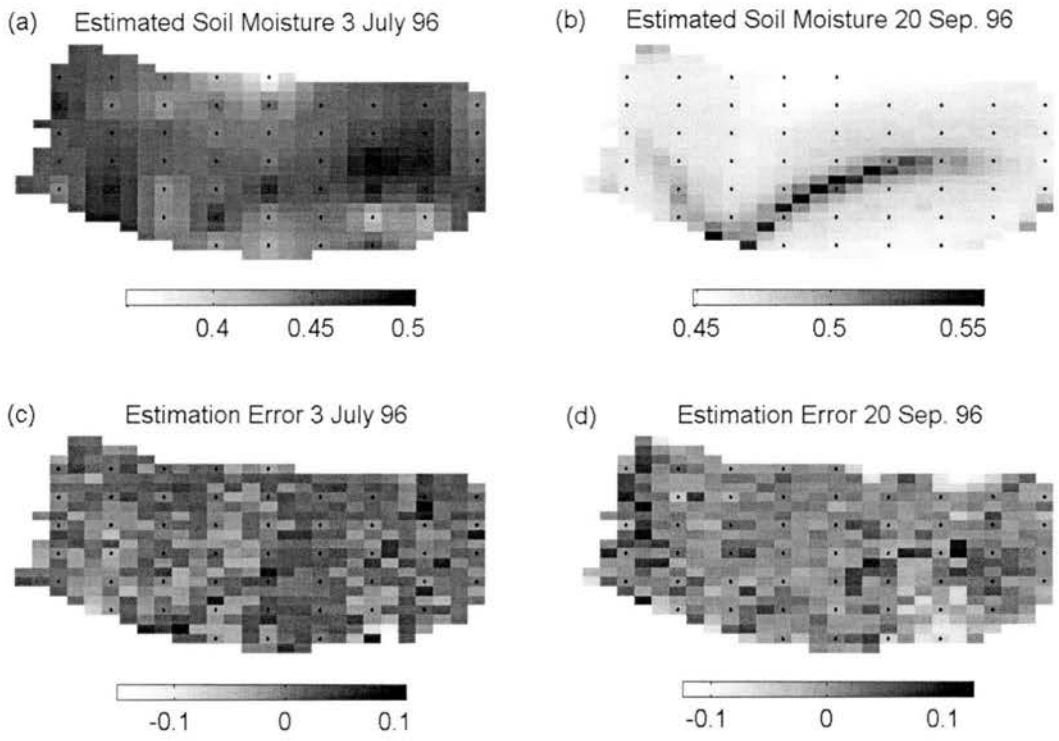

Figure 3.9 Soil moisture estimates and associated estimation errors using EOF-based interpolation methods. (a) 3 July estimate from EOFs 1-3 interpolated using the inverse distance weighting method. (b) 20 September estimate from EOFs 1-3 interpolated using a multiple linear regression against topographic attributes. (c) The associated pattern of estimation errors for 3 July. (d) The associated pattern of estimation errors for 20 September. The dots indicate the locations of the observations (30 m x 60 m spacing, 52 points total).

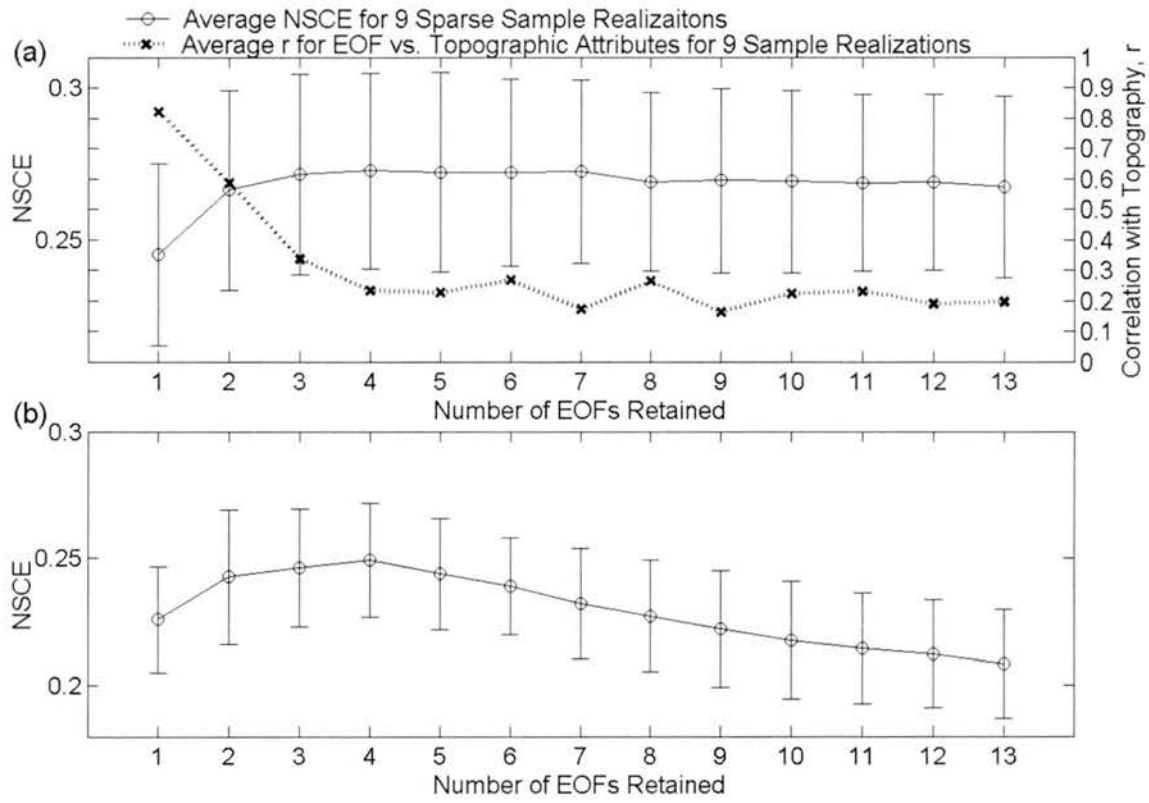

Figure 3.10 Evolution of *NSCE* for estimated soil moisture as additional EOFs are retained for interpolation by (a) multiple linear regression against topography and (b) the inverse distance weighted method. The x-axis represents the number of EOFs retained, and the EOFs are retained in order of the variance they explain. The multiple linear correlation coefficient r between each EOF and the topographic attributes considered is overlaid in part (a). Error bars show +/-1 standard deviation of *NSCE* computed for nine different 30 m x 60 m observation sets.

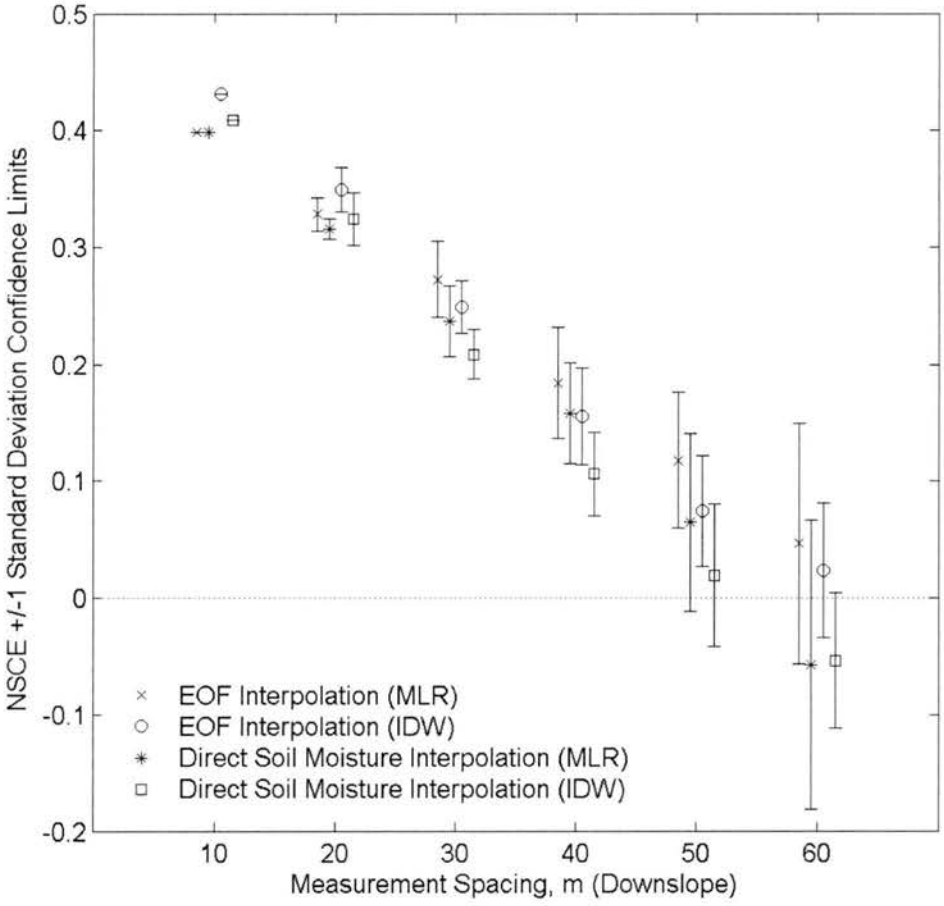

Figure 3.11 Comparison of NSCE for estimated soil moisture from EOF-based interpolation and direct soil moisture interpolation as a function of sample spacing. Recall that the sampling grid is anisotropic (e.g., 30 m x 60 m). The spacing shown here on the x-axis corresponds to the shorter dimension on the anisotropic grid. The plot shows the mean NSCE and +/- one standard deviation obtained from all sampling realizations.

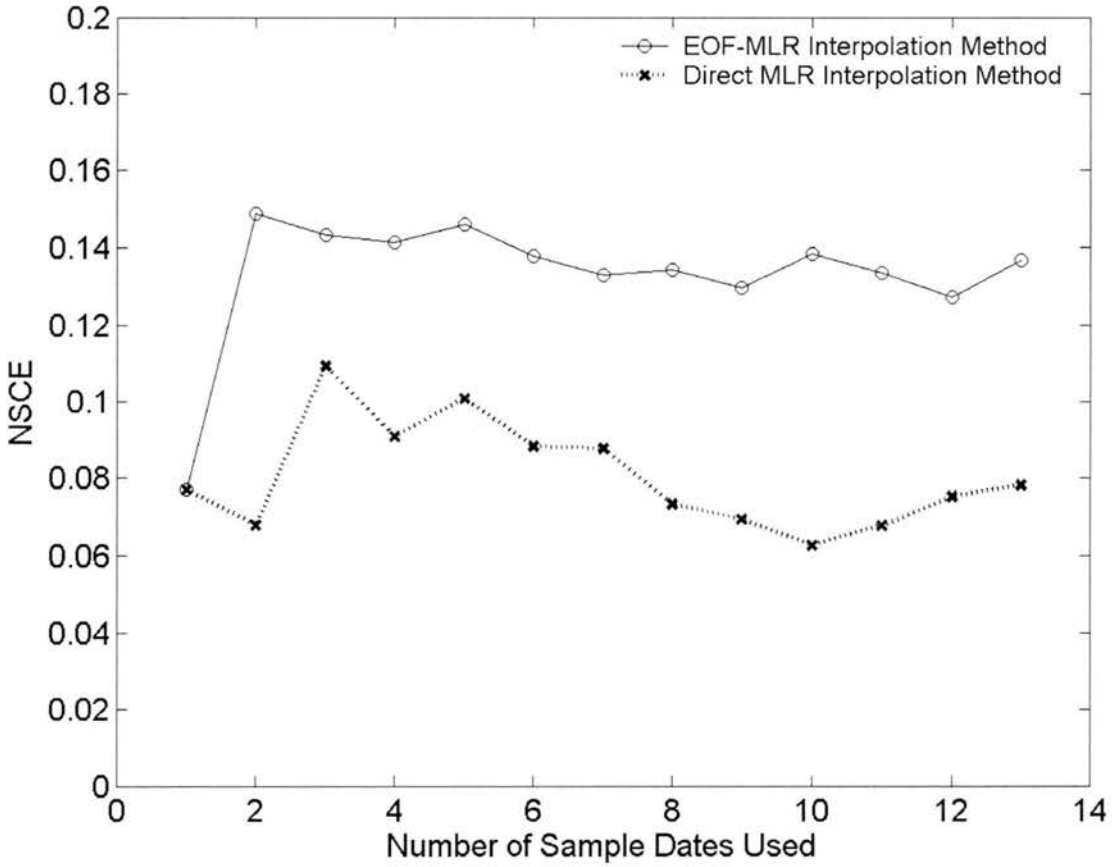

Figure 3.12 The *NSCE* for interpolation of 28 March soil moisture as a function of the number of sample dates used. The plot compares EOF interpolation and direct soil moisture interpolation using a multiple linear regression against topography in both cases.

4 Conclusions

The objective of the closing chapter is to assess the overall implications of this work and to suggest possibilities for future work. The main contribution of this work is showing that EOF analysis is effective at decomposing catchment scale soil moisture in a way that leads to better estimation of soil moisture at unobserved times and better interpolation in space when sparse measurements are available. EOF analysis decomposes the soil moisture data along orthogonal covariation. The analysis here has shown orthogonal covariation to be an effective, if imperfect, criterion for separating variation associated with individual processes that affect soil moisture.

Soil moisture variation results from the interaction of multiple processes. The importance of each process can change through time. EOF analysis is capable of modeling this behavior by identifying underlying persistent patterns (EOFs) and associated time-varying weights (ECs). Here, EOF1 was shown to be related mainly to the pattern of lateral drainage during steady-state conditions, EOF2, to the pattern of evapotranspiration, and EOF3, to the pattern of drainage during non-steady state conditions. Likewise EC1 was shown to be most important during moderately wet times, EC2, during very dry and very wet times, and EC3, when water is accumulating high in the catchment and draining from the catchment. The physical interpretation of EOFs was shown to be important as it allowed better characterization and prediction of their space-time behavior.

Previous studies have observed strong topographic dependence of soil moisture variation at the Tarrawarra catchment. Here, topographic dependence was confirmed. EOFs that explained the most soil moisture variation were shown to be highly correlated with topography. In addition, important EOFs showed a strong correlation to topographic attributes related to a particular process. It was shown that important EOFs exhibit distinct and stronger topographic dependence than measured soil moisture. This was demonstrated by higher linear correlations between EOFs and common topographic attributes than between soil moisture and the same topographic attributes. Stronger topographic dependence was also demonstrated by comparing the residual errors from an EOF-based interpolation method with errors from a comparable interpolation method applied directly to soil moisture. Analysis of the errors showed that the EOF method accounted for more of the topographic dependence of soil moisture.

If previous studies have identified the importance of topography at the Tarrawarra catchment, they have had difficulty quantifying the time-varying relationship between dynamic soil moisture patterns and static topography. This study has shown that persistent EOF patterns exhibit time-stable relationships to topography. The time-varying nature of the soil moisture-topography relationship can be quantified with the associated ECs. For example, EOF2 was observed to be related to topographic aspect. EC2 was shown to vary seasonally with the catchment spatial average soil moisture.

EOF analysis was shown to be effective at separating system variation from noise, e.g., measurement error. Higher order EOFs were shown to have variogram parameters similar to those for a field of random noise, while lower order EOFs were distinct from random noise. By quantifying the large amount of variation associated with noise, this

analysis has emphasized the continued need to minimize measurement errors. Variation caused by measurement error can subsume true soil moisture variation, especially during very wet and very dry times when overall soil moisture variation is low. Measurement error limits our ability to understand soil moisture variation.

Based on the results presented here, it is recommended that future study of soil moisture should involve decomposition. Other methods besides EOF analysis should be investigated. Related methods might include canonical component analysis, independent component analysis and cluster analysis. Soil moisture variation can also be decomposed according to frequency using spectral analysis or wavelet transforms. Regardless, decomposition should be performed to partition variation with uncertain spatial behavior into variation that exhibits consistent behavior. As demonstrated here, it should be possible to characterize, estimate and interpolate decomposed variables better than soil moisture itself.

It was shown here that EOF-based interpolation methods outperform comparable methods applied directly to soil moisture. However, EOF-based interpolation was demonstrated using commonly available interpolation techniques (e.g. inverse distance weighted method). Future work should be done to develop new spatial interpolation methods that specifically exploit the distinct spatial behavior of the decomposed variables. For example, it was acknowledged that the omni-directional variogram is not the best way to describe flow related variation. Variation along flow lines is clearly not the same in all directions. Better interpolation methods could be developed to account for the unique spatial structure of decomposed variables.

The analysis here was conducted for a site where soil type, landuse and vegetation were spatially uniform. The site emphasized topographic control of soil moisture. Future work should involve similar decomposition, estimation and interpolation at sites with non-uniform vegetation and soils. The space-time interaction of vegetation and soil moisture should be investigated. For example, is soil moisture variation related to variation in plant growth at certain times of year? And do different plant communities create different temporal patterns of variation? The role of soil properties in producing persistent soil moisture patterns should be investigated. Previous studies have attributed very dry and very wet soil moisture patterns to soil properties, namely permanent wilting point and water content at saturation, respectively. If this is true, EOF analysis should have been very effective at identifying persistent patterns during very dry and very wet periods. Surprisingly it was least effective at these times.

This analysis was performed for a site with a strong seasonal climate. The important modes of variation identified were generally related to the seasonal variation of hydrologic processes. Future work should investigate sites without seasonal variation or should investigate intra-season variation. Important non-seasonal variation likely occurs at smaller time scales, e.g., in response to daily weather.

Finally analysis of EC temporal behavior was limited here because the dataset consisted of only 13 sampling dates. Unfortunately, there is a scarcity of soil moisture datasets with large space and time dimensions. One possible solution is computer simulation of soil moisture data using a physically-based, distributed model that includes important soil moisture processes in its formulation. Simulation of large amounts of soil moisture data could allow better characterization of ECs using time-series analysis. ECs

associated with different processes should exhibit distinct time-structures. Based on results here, it is anticipated that ECs will exhibit more certain temporal behavior than soil moisture. This should allow better soil moisture forecasting when time-series modeling is done on ECs versus on soil moisture itself.

References

Bárdossy, A., and W. Lehmann (1998), Spatial distribution of soil moisture in a small catchment. Part 1: geostatistical analysis, *Journal of Hydrology*, *206*, 1-15.

Barling, R. D., I. D. Moore, and R. B. Grayson (1994), A quasi-dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content, *Water Resources Research*, *30*, 1029-1044.

Bartlett, M. S. (1950), Tests of Significance in Factor Analysis, *Br. J. Psych. Stat. Sec. 3*, 77-85.

Beven, K. J., and M. J. Kirkby (1979), A physically based variable contributing area model of basin hydrology, *Hydrological Sciences*, *24*, 43-69.

Biau, G., E. Zorita, H. Von Storch, and H. Wackernagel (1999), Estimation of precipitation by kriging in the EOF space of the sea level pressure field, *Journal of Climate*, *12*, 1070-1085.

Bras, R. L. (1999), A brief history of hydrology, The Robert E. Horton Lecture, *Bulletin of the American Meteorological Society*, *80*, 1151-1164.

Buell, C. E. (1978), The number of significant proper functions of two-dimensional fields, *Journal of Applied Meterology*, *17*, 717-722.

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

Cattell, R. B. (1966), The scree test for the number of factors, *Mult. Behav. Res.*, *1*, 245-276.

Charpentier, M. A., and P. M. Groffman (1992), Soil moisture variability within remote sensing pixels, *Journal of Geophysical Research*, *97*, 18987-18995.

Cosh, M. H., J. R. Stedinger, and W. Brutsaert (2004), Variability of surface soil moisture at the watershed scale, *Water Resources Research*, *40*, 1-9.

Costa-Cabral, M. C., and S. J. Burges (1994), Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas, *Water Resources Research*, *30*, 1681-1692.

Cressie, N. A. C. (1991), *Statistics for Spatial Data*, John Wiley & Sons, New York.

Delworth, T. L., and S. Manabe (1988), The influence of potential evaporation on the variabilities of simulated soil wetness and climate, *Journal of Climate*, *1*, 523-547.

Dingman, S. L. (2002), *Physical Hydrology*, 2nd ed., Prentice Hall, Upper Saddle River.

Dowdy, S., and S. Wearden (1991), *Statistics for Research*, 2nd ed., John Wiley & Sons, New York.

Downer, C. W., and F. L. Ogden (2003), Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment, *Water Resources Research*, *39*, 1-13.

Eltahir, E. A. B., and R. L. Bras (1994), Precipitation recylcing over the Amazon basin, *Quarterly Journal of the Royal Meteorological Society*, *120*, 861-880.

Entekhabi, D., E. G. Njoku, P. Houser, M. Spencer, T. Doiron, K. Y, J. Smith, R. Girard, S. Belair, W. J. Crow, T J, Y. H. Kerr, J. S. K. Kimball, R, K. C. McDonald, P. E. O'Neill, T. Pultz, S. W. Running, J. Shi, E. Wood, and J. van Zyl (2004), The Hydrosphere State (Hydros) Satellite Mission: An earth system pathfinder for

global mapping of soil moisture and land freeze/thaw., *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 2184-2195.

Erskine, R. (2005), Effects of DEM Accuracy, Grid Cell Size, and Alternative Flow Routing Algorithms on Computed Terrain Attributes, Colorado State University, Fort Collins.

Florinsky, I. V., R. G. Eilers, G. R. Manning, and L. G. Fuller (2002), Prediction of soil properties by digital terrain modelling, *Environmental Modelling & Software*, *17*, 295-311.

Gottschalk, L. (1993), Correlation and covariance of runoff, *Stochastic Hydrology and Hydraulics*, *7*, 85-101.

Graybill, F. W. (1976), *Theory and Application of the Linear Model.*, Duxbury Press, Belmont, California.

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, *Journal of Hydrology*, *207*, 68-82.

Grayson, R. B., A. W. Western, H. S. Chiew, and G. Blöschl (1997), Preferred states in spatial soil moisture patterns: Local and nonlocal controls, *Water Resources Research*, *33*, 2897-2908.

Green, T. R., and R. H. Erskine (2004), Measurement, scaling, and topographic analysis of spatial crop yield and soil water content, *Hydrological Processes*, *18*, 1447-1465.

Hu, Z., C. Yizong, and S. Islam (1998), Multiscaling properties of soil moisture images and decomposition of large- and small-scale features using wavelet transforms, *International Journal of Remote Sensing*, *19*, 2451-2467.

Jackson, J. E. (2003), *A User's Guide to Principal Components*, John Wiley & Sons, Hoboken, New Jersey.

Jackson, T. J., D. B. Le Vine, A. Y. Hsu, A. Oldak, P. J. Starks, C. T. Swift, J. D. Isham, and M. Haken (1999), Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains hydrology experiment, *IEEE Trans on Geoscience and Remote Sensing*, *37*, 2136-2151.

Jackson, T. J., D. B. Le Vine, C. T. Swift, T. J. Schmugge, and F. R. Schiebe (1995), Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita 92, *Remote Sensing of Environment*, *52*, 27-37.

Jaynes, D. B., T. C. Kaspar, T. S. Colvin, and D. E. James (2003), Cluster analysis of spatiotemporal corn yield patterns in an Iowa field, *Agronomy Journal*, *95*, 574-586.

Jolliffe, I. T. (2002), *Principal Component Analysis*, 2nd ed., Springer, New York.

Kachanoski, R. G., and E. De Jong (1988), Scale dependence and the temporal persistence of spatial patterns of soil water storage, *Water Resources Research*, *24*, 85-91.

Kim, G., and A. P. Barros (2002a), Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data, *Remote Sensing of Environment*, *83*, 400-413.

Kim, G., and A. P. Barros (2002b), Space-time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data, *Remote Sensing of Environment*, *81*, 393-403.

Kitanidis, P. K., and R. L. Bras (1980), Real time forecasting with a conceptual hydrologic model. 2. Application and results., *Water Resources Research, 16*, 1034-1044.

Kleinbaum, D. G., L. L. Kupper, and K. E. Muller (1988), *Applied Regression Analysis and Other Multivariate Methods*, PWS-Kent Publishing Co., Boston, Mass.

Li, J., and S. Islam (1999), On the estimation of soil moisture profile and surface fluxes partitioning from sequential assimilation of surface layer soil moisture, *Journal of Hydrology, 220*, 86-103.

Liu, Y. (2003), Spatial patterns of soil moisture connected to monthly-seasonal precipitation variability in a monsoon region, *Journal of Geophysical Research, 108*, 1-14.

Matalas, N. C., and B. J. Reiher (1967), Some Comments on the Use of Factor Analysis, *Water Resources Research, 3*, 213-223.

Mitasova, H., and J. Hofierka (1993), Interpolation by regularized spline with tension: II. Applications to terrain modeling and surface geometry analysis, *Mathematical Geology, 25*, 657-669.

Mohanty, B. P., T. H. Skaggs, and J. S. Famiglietti (2000), 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 Resources Research, 36*, 1023-1031.

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., *Transactions of the American Society of Agricultural Engineering, 31*, 1098-1107.

Moore, I. D., T. W. Norton, and J. E. Williams (1993), Modelling environmental heterogeneity in forested landscapes, *Journal of Hydrology, 150*, 717-747.

Nash, J. E., and J. V. Sutcliffe (1970), River forecasting through conceptual models, part I, a discussion of principle, *Journal of Hydrology, 10*, 282-290.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng (1982), Sampling errors in the estimation of empirical orthogonal functions, *Monthly Weather Review, 110*, 699-706.

Pellenq, J., J. Kalma, G. Boulet, G. M. Saulnier, S. Wooldridge, Y. Kerr, and A. Chehbouni (2003), A disaggregation scheme for soil moisture based on topography and soil depth, *Journal of Hydrology, 276*, 112-127.

Preisendorfer, R. W. (1988), *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, New York.

Rao, A. R., and C. H. Hsieh (1991), Estimation of variables at ungaged locations by empirical orthogonal functions, *Journal of Hydrology, 123*, 51-67.

Roberts, D. W., T. I. Dowling, and J. Walker (1997), FLAG: A fuzzy landscape analysis GIS method for dryland salinity assessment, Technical Report No. 8/97, 23 pp., CSIRO Land and Water, Canberra, Australia.

Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai (2000), The global soil moisture databank, *Bulletin of the American Meteorological Society, 81*, 1281-1299.

Schorghofer, N., and D. H. Rothman (2001), Basins of attraction on random topography., *Physical Review E, 63*, 1-7.

Seyfried, M. S., and B. P. Wilcox (1995), Scale and the nature of spatial variability: Field examples having implications for hydrologic modeling, *Water Resources Research, 31*, 173-184.

Skoien, J. O., G. Bloschl, and A. W. Western (2003), Characteristic space scales and timescales in hydrology, *Water Resources Research, 39*.

Thattai, D., and S. Islam (2000), Spatial analysis of remotely sensed soil moisture data, *Journal of Hydrologic Engineering, October*, 386-392.

Vachaud, G., A. Passerat De Silans, P. Balabanis, and M. Vauclin (1985), Temporal stability of spatially measured soil water probability density function, *Soil Science Society of America Journal, 49*, 822-828.

Verhoest, N. E., P. A. Troch, C. Paniconi, and F. P. De Troch (1998), Mapping basin scale variable source areas from multitemporal remotely sensed observations of soil moisture behavior, *Water Resources Research, 34*, 3235-3244.

Western, A. W., G. Bloschl, and R. B. Grayson (2001), Toward capturing hydrologically significant connectivity in spatial patterns, *Water Resources Research, 37*, 83-97.

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

Western, A. W., and R. B. Grayson (1998), The Tarrawarra data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements, *Water Resources Research, 34*, 2765-2768.

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

Western, A. W., R. B. Grayson, and T. R. Green (1999b), The Tarrawarra project: high resolution spatial measurement, modelling and analysis of soil mositure and hydrologic response, *Hydrological Processes, 13*, 633-652.

Western, A. W., S. L. Zhou, R. B. Grayson, T. A. McMahon, G. Blöschl, and D. J. Wilson (2004), Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, *Journal of Hydrology, 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 Resources Research, 40*, 1-10.

Wilson, D. J., A. W. Western, and R. B. Grayson (2005), A terrain and data-based method for generating the spatial distribution of soil moisture, *Advances in Water Resources, 28*, 43-54.

Yoo, C., and S. Kim (2004), EOF analysis of surface soil moisture field variability, *Advances in Water Resources, 27*, 831-842.

Zaslavsky, D., and G. Sinai (1981), Surface Hydrology, I. - Explanation of phenomena, *Journal of the Hydraulics Division, 107*, 1-16.