REMOTELY SENSED DATA ASSIMILATION TECHNIQUE TO DEVELOP MACHINE LEARNING MODELS FOR USE IN WATER MANAGEMENT

by

Bushra Zaman

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

of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

Approved:

Dr. Mac McKee Major Professor	Dr. Gilberto Urroz Committee Member
Dr. Adele Cutler Committee Member	Dr. Christopher M.U. Neale Committee Member
Dr. Kashif Gill Committee Member	Dr. Wynn R. Walker Committee Member

Dr. Byron R. Burnham
Dean of School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2010

ABSTRACT

Remotely Sensed Data Assimilation Technique to Develop

Machine Learning Models for Use in Water Management

by

Bushra Zaman, Doctor of Philosophy

Utah State University, 2010

Major Professor: Dr. Mac McKee
Department: Civil and Environmental Engineering

Increasing population and water conflicts are making water management one of the most important issues of the present world. It has become absolutely necessary to find ways to manage water more efficiently. Technological advancement has introduced various techniques for data acquisition and analysis, and these tools can be used to address some of the critical issues that challenge water resource management.

This research used learning machine techniques and information acquired through remote sensing, to solve problems related to soil moisture estimation and crop identification on large spatial scales. In this dissertation, solutions were proposed in three problem areas that can be important in the decision making process related to water management in irrigated systems. A data assimilation technique was used to build a learning machine model that generated soil moisture estimates commensurate with the scale of the data. The research was taken further by developing a multivariate machine learning algorithm to predict root zone soil moisture both in space and time. Further, a model was developed for supervised classification of multi-spectral reflectance data using

a multi-class machine learning algorithm. The procedure was designed for classifying crops but the model is data dependent and can be used with other datasets and hence can be applied to other landcover classification problems.

The dissertation compared the performance of relevance vector and the support vector machines in estimating soil moisture. A multivariate relevance vector machine algorithm was tested in the spatio-temporal prediction of soil moisture, and the multi-class relevance vector machine model was used for classifying different crop types. It was concluded that the classification scheme may uncover important data patterns contributing greatly to knowledge bases, and to scientific and medical research. The results for the soil moisture models would give a rough idea to farmers/irrigators about the moisture status of their fields and also about the productivity. The models are part of the framework which is devised in an attempt to provide tools to support irrigation system operational decisions. This information could help in the overall improvement of agricultural water management practices for large irrigation systems. Conclusions were reached based on the performance of these machines in estimating soil moisture using remotely sensed data, forecasting spatial and temporal variation of soil moisture and data classification.

These solutions provide a new perspective to problem-solving techniques by introducing new methods that have never been previously attempted.

(140 pages)

To my parents, Parween and Fasihuzzaman

ACKNOWLEDGMENTS

I express my deep gratitude and feel obligated to my advisor, Dr. Mac McKee, who provided me this wonderful opportunity to work with him. I can never forget my first day, when I walked to the lab in January, very unsure of what awaited me. I am thankful for that day and consider myself fortunate to have met Dr. Mac. Dr. Mac let me work in a free environment and gave me the liberty to establish my research areas. I got the best guidance, encouragement, help, and moral support to make this research possible. I appreciate and acknowledge his role in my life with all my being.

I am thankful to Dr. Adele Cutler for her guidance and Dr. Kashif Gill and Dr. Yasir Kaheil for answering my questions and being very patient with me. I thank Dr. Christopher Neale for his intelligent remarks that improved my research paper a lot. I sincerely appreciate Dr. Wynn Walker's remarks about the practical utility of my research, and I feel grateful to Dr. Gilberto Urroz for his support and for answering my MATLAB questions.

I extend my gratitude to my colleagues, Andres Ticlavilca and Alfonso Torres, and my UWRL support system, especially Jan Urroz, for helping me out in times of need. My friend, Dr. Brijesh Yadav, provided a lot of help with his useful suggestions whenever needed.

I am grateful for my parents' constant moral support and the unfailing support and encouragement of my friend and fiancé Nilesh. Above all, I think God has been kind to me.

Bushra Zaman

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

General Introduction

Competition for water, high pumping costs, complexities of water storage and delivery, and concerns for the environment are among the factors that drive an interest in improving the operation of large irrigation systems. Irrigation water management requires timely application of the right amount of water. For a clear picture of timely and efficient water application, agricultural managers have long relied on soil moisture measurements. A situation where soil moisture information can be obtained beforehand can be of value. This information gives an idea about low soil moisture levels that can reduce yields or of excessive water application that can result in water logging or leaching of nutrients below the root zone. Many methods of estimating or measuring soil moisture are available. The preferred method depends on a variety of factors such as accuracy, cost, and ease of use. Field measurements of soil moisture are very reliable but are also time- and resource-consuming. Several physical models have been devised for soil moisture estimation, but such models require detailed information about the physical parameters that is sometimes hard to obtain. This has encouraged scientists to look for alternative methods for soil moisture estimation, such as data mining and data-driven modeling.

Data mining has attracted much attention in the recent years due to wide availability of huge amounts of data, especially from remote sensing platforms, and the imminent need for turning such data into useful information and knowledge. There has been a widening gap between data and information that requires a systematic

development of data mining tools that can turn data tombs into “golden nuggets” of knowledge (Han and Kamber, 2006). One attractive alternative is the use of artificial intelligence techniques. They distinguish themselves from physically based models or conceptual models in that the set of functions they use for defining the mapping are neither physically based nor conceptually derived. Modeling based upon machine learning techniques provides more flexibility in choosing variables of interest for solving water resource management problems and does not require set parameters. Data-driven models have seen limited use in hydrology. In contrast to physically based models, they are easy to calibrate.

Another major concern in agricultural water management and soil moisture information is that of scale. Measurements of soil moisture are obtained generally from field measurements, but because of the time and resource requirements of these practices, it is often impossible to monitor large areas in detail. Remote sensing techniques are helpful in addressing some of these problems. Remotely sensed data required for hydrologic models come from different sources, e.g. remote sensing satellites, aerial image acquisition and field measurements. The analysis of these datasets requires a model that is flexible enough to accommodate different spatial and temporal resolutions and/or data availability and frequency. This research uses machine learning techniques to build models to estimate hydrologic state variables. These models have an additional advantage of providing estimates having resolution commensurate with remotely sensed data.

The work reported here is directed towards the use of data assimilation techniques and machine learning tools to develop models that are useful in addressing water

management problems. The models provide information to the farmer/irrigator about the crop status and status of soil moisture on the field. This has the potential to help on site specific management and improving on-farm irrigation and real-time canal operations.

The first study reported here deals with the estimation of soil moisture from readily available remotely sensed data, meteorological information, and crop physiological characteristics. Satellite data provides a huge resource for image retrieval in diverse fields including weather prediction, water resource management, and agriculture and environment sciences (Healy and Jain, 1996). Recent years have seen a rapid increase in the size of digital image collections. Every day, several giga-bytes of images are generated and a huge amount of information is available for use. This study uses reflectance in the optical and near-infrared wavebands. The idea of using solar irradiance at specific wavelengths to estimate surface soil moisture has been explored by many researchers, and the possibility of estimating surface soil moisture from visible and near-infrared (NIR) reflectance has also been demonstrated. This research is the first attempt to analyze the reflectance data in the optical and NIR bands with machine learning techniques to retrieve soil moisture. The relevance vector machine (RVM) was used to extract the information hidden in the data. Different data were assimilated which included precipitation, air temperature, soil temperature and texture, canopy temperature, vegetation indices (SAVI, NDWI) and leaf area index (LAI). The RVM model was able to recognize the input and output pattern between soil moisture and the assimilated data. The results proved that soil moisture up to a depth of 30 cm can be extracted using this procedure.

The results from this analysis were encouraging and provided a foundation for testing the ability of the learning machine to predict soil moisture in space and time. Hence the research was further extended where the multivariate relevance vector machine (MVRVM) was used to forecast soil moisture to a depth of two meters for several days in the future. With this goal in mind, a data assimilation technique was applied wherein, soil moisture at shallower depths, soil temperature, and precipitation were used as inputs to a MVRVM model. The model exhibited an excellent capability of forecasting soil moisture. The method has the potential for forecasting soil moisture profiles at lower cost and low complexity and should be well suited for routine use by horticulturists, soil researchers, and perhaps irrigation system operators.

Data assimilation was further explored for supervised classification of remotely sensed data. Multispectral radiometer reflectance data was combined with other ancillary data, and the multiclass relevance vector machine (MCRVM) was used to build a classification algorithm. The classification results obtained by using this procedure produced the best accuracies that have been reported in the literature. This classification algorithm was designed mainly for crop identification. The ancillary data that were sensitive to vegetation differences were used as inputs. This MCRVM supervised classification model is data dependent and can be used for classifying different types of data by defining a suitable dataset.

These approaches use a concrete paradigm that is mathematically sound with manageable computational complexity. These sparse learning machines are theoretically elegant and well-regularized, in general require few parameters, and are relatively easy to calibrate.

Purpose and Objectives

Purpose of the study

Due to changing environmental conditions and behavioral patterns of water demand, even over short time periods, a prediction tool is required to adapt incrementally and preferably in real time to predict hydrologic state variables such as soil moisture and perform environmental modeling and crop mapping which help in yield estimation and updating. Water management is often implemented on the ground by canal and reservoir operators. Most of them know how long it takes to irrigate fields and avoid crop stress during average conditions. Human experience, in this context, is valuable but there is always room for improvement. There are enormous amounts of available data, including data from remote sensing platforms, which could be used for developing machine learning models that can predict soil moisture and perform crop classification on large spatial and temporal scales. These modeling tools can be integrated in the decision support system which might help in solving water management problems.

Objectives

The objectives of the study are to develop a machine learning models using data assimilation technique:

- For estimating surface soil moisture by assimilating remotely sensed reflectance measurements in visible and near-infrared bands and other ancillary variables affecting soil moisture.
- For simultaneous spatio-temporal forecasting of soil moisture at different root zone depths.

— For performing supervised classification for landcover and crop identification by assimilating spectral reflectance data with other ancillary data.

Research Motivation

Many different types of modeling approaches have been used in hydrology, such as physical or scale models, mathematical models, lumped conceptual models, distributed physically based models, and empirical models. During the last decade the area of empirical modeling has received greater attention due to developments in the area of machine learning.

Hydrologists are confronted with problems of prediction and estimation that are characterized by physical processes that exhibit a high degree of spatial and temporal variability, issues of non-linearity, and uncertainty (Khalil, Mckee, and Kaluarachchi, 2005). Data driven modeling is relatively new to hydrologic applications. It is based on the analysis of all the data characterizing the system under study. A data driven model can then be inferred on the basis of the relationship between the system state variables (input and output variables) with only a limited number of assumptions about the physical behavior of the system. Physical models are well established but the difficulty associated with measurement of the physical parameters required by these models sometimes serves as an impediment. Data-driven approaches are characterized by their fundamental ability to deduce models of system behaviors from available data. Without sacrificing accuracy, they provide a potentially valuable method for reducing the cost of data collection to support the information needs of complex water management systems (Velickov and Solomantine, 2000). Data-driven modeling is called so because the model

"learns" the inferring function from the data. Recently, data-driven modeling has gained attention in remote sensing applications as a valuable inverse model that can retrieve physical characteristics of interest, such as soil moisture, from remote sensing measurements collected from radar or satellites. The spatial coverage of remote sensing measurements relative to ground-based measurements and their high resolution can improve the usefulness of hydrological modeling at both local and global scales. One of the greatest advantages of using remote sensing data for hydrological modeling and monitoring is its ability to generate information at the appropriate spatial and temporal scales, which is very crucial for successful model analysis, prediction, and validation. Remote sensing data helps in solving water management problems on a large scale. Also, ancillary data, either in addition to or derived from remotely sensed data, has the potential for increasing accuracy and precision of the model. Therefore this dissertation explores the use of the above-mentioned data, tools, and techniques to address water management problems.

Another important motivation behind the research was testing the capability of machine learning approaches. The methods used in this dissertation are directed at utilizing massive amounts of data in designing machine learning models which can be used as tools for decision support systems of water management. They adapt to time varying behavior and incrementally learn changes as they come across more data in near real time. Irrespective of the underlying physical relationship between predictors and the predicted, locally learned mapping effectively estimates a conditional expectation that is statistically consistent with the real data. Therefore such models do not just provide

predictions but also give an insight into the probabilistic nature of the underlying processes. Hence, RVMs, which have this ability, were used for this research.

Research Contributions

The work that has been done provides tools and techniques that could help in solving . variables) with only a limited number of assumptions about the physical behavior of the system.

Soil moisture serves as a substantial input to the soil water balance calculations and is one of the hydrological variables that play an important part in the energy budgets necessary for climate studies. An operational capability to predict the temporal variation and spatial distribution of soil moisture profiles would have numerous benefits in the fields of meteorology, hydrology, agriculture, and the monitoring of global climate change. This research is a first attempt to apply remotely sensed data and data assimilation techniques with the machine learning approach to estimate surface soil moisture. Thereafter, spatial and temporal forecasting of soil moisture in the root zone is attempted. This could be used to help inform the farmer /irrigator about the soil moisture status of the field, which could enhance on-farm irrigation efficiency and aid the process of decision-making related to water orders and delivery. With this in mind, the research was taken further to develop a crop classification scheme. This reflects various elements related to crops, including the growing cycle (temporary/permanent), crop species, crop variety, season, land type, crop use, type of product and cultivation methods (for example, crops grown under protective cover).

The technical contributions of this research include:

1. A data driven model employing intelligent data analysis to build a soil moisture estimation algorithm which uses remotely sensed and other readily available data for soil moisture estimation on a large scale.
2. A robust multi-output prediction algorithm for spatio-temporal forecasting of root zone soil moisture which produces predictions simultaneously at different depths, several days in the future.
3. A well-defined computational procedure for supervised classification of multispectral reflectance data which is fast and accurate and can be applied for discrimination of different surface types.

A literature review suggests that some of the general potential benefits of using the techniques examined in this research are:

1. Improved estimates of evapotranspiration through the influence on partitioning of available energy at the ground surface into sensible and latent heat exchange (Entekhabi, Nakamura, and Njoku, 1993, 1994; Giacomelli et al., 1995).
2. Increased crop yield through optimal soil moisture conditions at pre-planting and during the growing season (Topp, Davis, and Annan, 1980; Jackson, Hawley, and O'Neill, 1987; Saha, 1995).
3. Improved weather predictions through improved modeling of the interaction of land surface processes (Fast and McCorcle, 1991; Engman, 1992; Betts et al., 1994).

4. Economic and water conservation benefits through rational irrigation scheduling (Jackson et al., 1981; Jackson, 1982; Jackson, Hawley, and O'Neill, 1987; Saha, 1995).
5. Management of cultural practices, including trafficability in the fields (Wigneron et al., 1998).
6. Early drought prediction (Engman, 1990), drought monitoring (Jackson et al., 1981; Jackson, Hawley, and O'Neill, 1987) and evaluation of drought impact on agricultural production (Newton, Heilman, and van Bavel, 1983) for management of rural subsidy schemes.
7. Improved erosion prediction through improved hydrological modeling and the relationship between erosion and runoff producing zones (Beecham, 1995; Western, Green, and Grayson, 1997).

Dissertation Organization

The general structure of the dissertation is as follows. Chapter I is a precursor to this dissertation that includes a very general introduction, problem statement, purpose and objectives, research motivation, and contribution that drove this dissertation to its completion. The dissertation has three main components. Chapter II provides a detailed literature review concerning the estimation of soil moisture content using different techniques. The chapter covers the necessary details related to the machine learning approach using RVMs and SVMs. Further on, the details of soil moisture estimation using remotely sensed data assimilation with the RVM and SVM models are covered. Chapter III discusses the literature related to the spatial and temporal forecasting of soil

moisture profile. The chapter covers the basic details of the MVRVM and the methodology used to develop this model. Chapter IV discusses the multi-class supervised classification algorithm for crop identification. The chapter discusses the necessary details of the MCRVM and ancillary data used to build the model. Chapter V summarizes the findings of this research, described the important inferences derived from this research, and presents the conclusions and recommendations.

CHAPTER II

FUSION OF REMOTELY SENSED DATA FOR SOIL MOISTURE ESTIMATION USING RELEVANCE VECTOR AND SUPPORT VECTOR MACHINES

Abstract

A data assimilation (DA) methodology that used two state-of-the-art techniques, relevance vector machines (RVMs) and support vector machines (SVMs), was applied to retrieve surface (0-6 cm) soil moisture content (SMC) and SMC at 30 cm depth. The RVMs and SVMs are known for their robustness, efficiency, and sparseness and provide a statistically sound approach to solving the inverse problems and thus to building statistical models. Here, we built a statistical model which produced acceptable estimations of SMC using inexpensive and readily available data. The study area for this research was the Walnut Creek watershed in Ames, south-central Iowa, USA. The data were obtained from Soil Moisture Experiments (SMEX) 2002 conducted at Ames, Iowa. The DA methodology combined remotely sensed inputs with field measurements, crop physiological characteristics, soil temperature, soil water holding capacity and meteorological data to build a two-step model for estimation of SMC using both techniques, i.e., RVMs and SVMs. First the RVM was used to build a model which retrieved surface (0-6 cm) SMC. This information served as a boundary condition for the second step of this model, which estimated SMC at 30 cm depth. An exactly similar routine was followed with a SVM for estimation of surface (0-6 cm) SMC and SMC at 30 cm depth. Results from the RVM and SVM models were compared and statistics showed that RVMs perform better (RMSE=0.014 m^3/m^3) as compared to SVM (RMSE= 0.017

m^3/m^3) with a reduced computational complexity and more suitable real-time implementation. Cross-validation techniques were used to optimize the model. Bootstrapping was used to check over/under-fitting and uncertainty in model estimates. Computations showed good agreement with the actual SMC measurements (RVM-R^2=0.92; SVM-R^2=0.88) and statistics indicated good model generalization capability (RVM-IoA=0.97; SVM-IoA=0.96).

Introduction

Precise estimation of soil moisture is necessary for soil water balance calculations, various hydrometeorological, ecological, or biogeochemical modeling applications, and initialization of various land-atmosphere models. Soil moisture constitutes about 0.0001% of the earth's water (Islam and Engman, 1996) but seasonal changes in this small quantity of water contribute to a 1.4 cm change in sea level (Mather, 1974). Soil moisture content (SMC) information helps in explaining processes related to crop growth, forest dynamics and other vadose zone processes which play a vital role in water resources planning and management. An accurate estimation of SMC is important for many applications. The objective of this research was to generate SMC estimates by applying a new remotely sensed data assimilation technique using learning machines which automatically learn to recognize complex patterns and make decisions based on data. Both relevance vector machines (RVMs) and support vector machines (SVMs) were used to build SMC estimation functions. The results obtained from both the machines were then compared.

SMC retrieval using different techniques has been the subject of intense research for almost four decades. Gravimetric measurements of SMC are very reliable but are also time and resource consuming. Measuring SMC with imbedded sensors, such as time- and frequency-domain reflectometers (TDRs and FDRs), do not require huge investment of time and facilities. However, most of these methods suffer from some of the same disadvantages. For example, *in situ* measurements can be exhaustive and expensive if large areas are involved as these methods are mainly 'local', where each measurement has a particular foot print representing moisture conditions in only a fraction of a cubic meter of soil. Also, because of spatial heterogeneity of soil moisture due to different soil conditions, vegetation, topography, or impacts of human activities, the local measurements must be carried out at a larger scale such as fields or watersheds (Liu et al., 2003) which might result in inaccuracies. Remote sensing techniques might provide a useful tool in addressing these data acquisition difficulties.

Some of the early work in estimating soil moisture using remote sensing was performed by Idso, Jackson, and Reginato (1975, 1976), Reginato et al. (1976), Reginato, Jackson, and Pinter (1985), and Jackson (1986). These authors established that thermal remote sensing, in concert with *in situ* measurements, can be used to measure, or at least quantitatively infer SMC (Quattrochi and Luvall, 1999). The use of solar irradiance at specific wavelengths to estimate surface SMC was explored by many researchers (Bowers and Hanks, 1965; Skidmore, Dickerson, and Schimmelpfennig, 1975; Muller and Decamps, 2001; Scott, Bastiaanssen, and Ahmad, 2003; Mouazen, Baerdemaeker, and Ramon, 2005). Likewise, the possibility of estimating the surface SMC (0-7.6 cm) from visible and near-infrared reflectance has also been demonstrated (Kaleita, Tian, and

Hirschi, 2005). Methods for SMC retrieval using remotely sensed data vary from purely empirical to more complex approaches (Moran et al., 2004; Wang et al., 2004). Optical and thermal remote sensing techniques (Price, 1980; Humes et al., 1993) or passive and active microwave sensors offer large-scale monitoring of SMC (Jackson et al., 1999; Njoku and Entekhabi, 1996; Njoku et al., 2003; Artiola, Pepper, and Brusseau, 2004). Some of the meteorological satellites, such as the Advanced Microwave Scanning Radiometer (AMSR-E), the European Remote Sensing satellite (ERS) scatterometer, or the meteorological satellite (METEOSAT), offer the possibility for operational SMC monitoring. Researchers have demonstrated the possibility of retrieval of SMC up to one meter depth in the soil profile using the METEOSAT and ERS scatterometer imagery (Wagner and Scipal, 2000; Wagner et al., 2003; Ceballos et al., 2005; Verstraeten et al., 2006). However the coarse spatial resolution (ERS-Scat: 50km; AMSR-E: 56km, METEOSAT: Visible and Infrared- 5km) of the instruments are often not consistent with the scale of hydrologic processes of interest (Das and Mohanty, 2006). The measurement represents the average over the resolution cell of the sensor, and consequently the land-surface related variability of the soil moisture field is averaged out (Wagner, Lemoine, and Rott, 1999). Also, the shallow moisture sensing depth of passive microwave sensors (Jackson and Schmugge, 1989; Engman and Chauhan, 1995; Jackson, O'neill, and Swift, 1997) and the perturbation of the signal by surface roughness and vegetation biomass, limits the use of remotely sensed soil moisture for many land-atmosphere interaction studies (Li and Islam, 2002). It has been stated in the literature that “a space-borne sensor designed to interpret SMC on the basis of soil microwave emission, and therefore the

relationship between soil dielectric constant and water content, will show considerable systematic uncertainty in SMC retrieval" (Fernandez-Galvez, 2008).

Remote sensing measurements in the thermal infrared band gave rise to the thermal inertia (TI) approach for SMC retrieval. The TI approach relates SMC to the magnitudes of the differences between daily maximum and minimum soil or crop canopy temperatures (Idso, Jackson, and Reginato, 1976). This procedure retrieves SMC from models that describe thermal inertia as a function of water content (Watson, 1975; Sabins, 1978; Pratt and Ellyett, 1979; Price, 1980, 1985; Majumdar and Bhattacharya 1990; Xue and Cracknell, 1995; Wang et al., 2004; Mitra and Majumdar, 2004; Cai et al., 2007; Lu et al., 2009). The TI approach is simple to implement because the knowledge of soil physical properties and climate can produce representative SMC profiles up to a depth of 1 meter; however, the limitation of the approach is its sensitivity to the uncertainty of soil physical properties, complex to determine in the field and typically obtained with point measurements (Verstraeten et al., 2006). The TI method provides large-scale spatial coverage but the functions are empirical and have the drawback of being site- and time-specific and as such, none of them are general enough to be applied extensively (Lu et al., 2009). Soil moisture monitoring by remote sensing includes another set of approaches which permit surface SMC retrieval from the information contained in a satellite-derived surface temperature (Ts)/vegetation index (VI) scatter plot (Carlson and Buffum, 1989; Carlson, Gillies, and Perry, 1994; Carlson, Gillies, and Schmugge, 1995). However, one of the major drawbacks to the Ts/VI method is that in order to have enough points in a remote sensing image to use in determination of the boundaries of extreme conditions, a sufficiently large number of pixels must be sampled.

This limitation is a handicap when dealing with smaller scale imagery on the order of the size of a typical farm field (Kaleita, Tian, and Hirschi, 2005).

Difficulties associated with the above approaches have furthered the interest of researchers to look for data-driven modeling tools such as artificial neural networks (ANNs), support vector machines (SVMs), and relevance vector machines (RVMs) for soil moisture estimation. In the recent past, researchers have made successful attempts to apply SVM modeling (Gill et al., 2006; Gill, Kemblowski and McKee, 2007; Yang and Huang, 2009), ANNs (Atluri, Chih-Cheng and Coleman, 1999; Chang and Islam, 2000; Jiang and Cotton, 2004; Song et al., 2008) and higher-order neural networks (Elshorbagy and Parasuraman, 2008) for soil moisture retrieval. Likewise, Khalil, Gill, and Mckee (2005), applied SVMs and RVMs for soil moisture estimation. One of the major advantages of the machine learning approach to SMC estimation is that it can provide estimates having resolution commensurate with remotely sensed data. The SVM modeling provides a very promising technique and has a remarkable estimation capacity (Mukherjee, Osuna, and Girosi, 1997). However, one can identify a number of its significant and practical disadvantages (Tipping, 2001). Ideally we desire to estimate the conditional distribution of the output in order to capture uncertainty in our estimation, but the SVM estimations are not probabilistic. SVMs make unnecessarily liberal use of kernel functions, and the requisite number grows linearly with the size of the training set. Also it is necessary to estimate the error/margin trade-off parameter ‘C’ which generally entails a cross-validation procedure, which is wasteful both of data and computation. Finally, the kernel function K(.,.) must satisfy Mercer’s condition (Tipping, 2001). The

RVMs are the Bayesian treatment of the SVM function and they do not suffer from any of these limitations.

In this study, both the SVM and RVM models were built. The remotely sensed optical and thermal data were combined with ancillary ground information and the assimilated dataset was used to train the learning machines to estimate SMC in 0-6cm depth of topsoil. The surface SMC estimates were then used to retrieve SMC at 30 cm depth.

Model Background

Relevance vector machine

The RVM was originally introduced by Tipping (2001), who discussed the detailed underlying mathematical basis for the technique. This section briefly reviews the salient features of the RVM.

The data set is in the form of input-output pairs and looks like $\{\mathbf{x}_n, t_n\}_{n=1}^{N}$. The major goal in machine learning is to establish the dependency of the model target functions on the inputs with the objective of making accurate estimates of unknown values of $\boldsymbol{y}$ when given $\mathbf{x}$ (Tipping, 2001). The standard probabilistic formulation is followed and it is assumed that target t_n represents the true model $y(\mathbf{x}_n;\boldsymbol{w})$ with some additional noise ε_n:

$$t_n = y(\mathbf{x}_n;\boldsymbol{w}) + \varepsilon_n \quad (1)$$

where ε_n is mean-zero Gaussian with variance σ^2 i.e. $\varepsilon_n \sim N(0, \sigma^2)$. Thus, $p(t_n|\mathbf{x}) = N(t_n|y(\mathbf{x}_n),\sigma^2)$, which is a Gaussian distribution over t_n with mean $y(\mathbf{x}_n)$ and variance σ^2. The likelihood of the complete data set can be written as,

$$p(\boldsymbol{t}|\boldsymbol{w},\sigma^2) = (2\pi\sigma^2)^{-N/2}\exp\{-\frac{1}{2\sigma^2}\|\boldsymbol{t}-\boldsymbol{\Phi w}\|^2\} \quad (2)$$

where $\boldsymbol{t} = (t_1 \ldots t_N)$, $\boldsymbol{\Phi}$ is an Nx(N+1) design matrix with $\boldsymbol{\Phi} = [\boldsymbol{\varphi}(\mathbf{x}_1), \boldsymbol{\varphi}(\mathbf{x}_2),\ldots, \boldsymbol{\varphi}(\mathbf{x}_N)]^T$, where, $\boldsymbol{\varphi}(\mathbf{x_n}) = (1, k(\mathbf{x}_n,\mathbf{x}_1), k(\mathbf{x}_n,\mathbf{x}_2, \ldots, k(\mathbf{x}_n,\mathbf{x}_N)]^T$. The Gaussian kernel was used for RVM formulation, which has the form: $k(\mathbf{x}_i,\mathbf{x}_j)=\exp(-r^{-2}\|\mathbf{x}_i-\mathbf{x}_j\|^2)$, where r is the kernel width and $\boldsymbol{w} = (w_1,w_2,\ldots,w_N)^T$, called weights, are adjustable parameters. A prior constraint over $\boldsymbol{w}$ is imposed by adding a complexity penalty to the likelihood to avoid overfitting. The 'hyperparameters' are used to constrain an explicit zero-mean Gaussian prior probability distribution over the weights, $\boldsymbol{w}$:

$$p(w|\alpha) = \prod_{i=1}^{N} N(w_i|0,\alpha_i^{-1}) \quad (3)$$

where $\boldsymbol{\alpha}$ is a vector of $N+1$ hyperparameters. The hyperpriors over $\boldsymbol{\alpha}$ and σ^2 are defined to complete the specification of the hierarchical prior. Consequently, using Bayes' rule, the posterior probability over all the unknown parameters can be computed:

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) = p(\boldsymbol{t}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2).\ p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)/\int p(\boldsymbol{t}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)d\boldsymbol{w}d\boldsymbol{\alpha}d\sigma^2 \quad (4)$$

The analytical solution of (4) is intractable. Hence, the posterior is decomposed:

$$p(\boldsymbol{w}, \boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) = p(\mathbf{w}|\boldsymbol{t}, \boldsymbol{\alpha}, \sigma^2).\ p(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) \quad (5)$$

The first part of (5) can be expressed as: $p(\mathbf{w}|\boldsymbol{t}, \boldsymbol{\alpha}, \sigma^2)\sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with posterior covariance, $\boldsymbol{\Sigma}=(\boldsymbol{A}+\beta\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}$ and mean, $\boldsymbol{\mu} =\beta\Sigma\boldsymbol{\Phi}^T\boldsymbol{t}$, where $\boldsymbol{A} = \mathrm{diag}(\alpha_0,\alpha_1,\ldots,\alpha_N)$ and $\sigma^2 = 1/\beta$. The second part of (5) i.e. $p(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{t})$, is represented by a delta function at its mode, i.e. at its most probable values $\boldsymbol{\alpha}_{\mathrm{MP}}$ and σ^2_{MP}. In order to calculate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, hyperparameters $\boldsymbol{\alpha}$ and β are required, which maximize the second part of (5), and which is decomposed as,

$$p(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) = \mathrm{p}(\boldsymbol{t}|\boldsymbol{\alpha}, \sigma^2)\ p(\boldsymbol{\alpha})p(\sigma^2) \quad (6)$$

"Learning" becomes a search for the most probable hyperparameter posterior mode, i.e., the maximization of $p(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) \propto p(\boldsymbol{t}|\boldsymbol{\alpha}, \sigma^2)\, p(\boldsymbol{\alpha})\, p(\sigma^2)$ with respect to $\boldsymbol{\alpha}$ and σ^2. For uniform hyperpriors, one needs only to maximize the term $p(\boldsymbol{t}|\boldsymbol{\alpha},\sigma^2)$:

$$p(\boldsymbol{t}|\boldsymbol{\alpha},\sigma^2) = \int p(\boldsymbol{t}|\mathbf{w}, \sigma^2)p(\mathbf{w}|,\boldsymbol{\alpha})d\mathbf{w}$$

$$= p(\boldsymbol{t}\big|\boldsymbol{\alpha},\sigma^2) = (2\pi)^{-N/2}\left|\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\right|^{-1/2} \exp\{-\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}(\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}})^{-1}\boldsymbol{t}\} \quad (7)$$

Maximization of this quantity is known as the type II maximum likelihood method (Berger, 1985). $\boldsymbol{\alpha}$ is obtained by differentiating (7). The learning procedure calls for a repeated updating of the previous values of $\boldsymbol{\alpha}$ and σ^2. It is observed that α_i approaches infinity such that $\boldsymbol{w}$ will have few non-zero weights. Those will be the "relevance vectors" (RVs). This results in sparsity of the model. The distribution for a new query $\mathbf{x}_{N+1}$ becomes,

$$p(t_{N+1}\big|\boldsymbol{t}, \boldsymbol{\alpha}_{MP}, \sigma^2_{MP}) = \int p(t_{N+1}\big|\boldsymbol{w}, \sigma^2_{MP})\, p(\boldsymbol{w}\big|\boldsymbol{t}, \boldsymbol{\alpha}_{MP}, \sigma^2_{MP})d\boldsymbol{w} \quad (8)$$

This is readily computed, resulting in,

$$p(t_{N+1}\big|\boldsymbol{t}, \boldsymbol{\alpha}_{MP}, \sigma^2_{MP}) \sim N(t_{N+1}\big|y_{N+1}, \sigma^2_{N+1}) \quad (9)$$

where $y_{N+1} = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}_{N+1})$ is the mean and $\sigma^2_{N+1} = \sigma^2_{MP} + \boldsymbol{\varphi}(\boldsymbol{x}_{N+1})^{\mathrm{T}}\ \Sigma\boldsymbol{\varphi}(\boldsymbol{x}_{N+1})$ is the variance of the distribution.

Support vector machines

Vapnik and his co-workers developed SVMs for regression (Vapnik, 1998). Only a brief description of the principles of SVMs is provided here. More details can be found in Vapnik (1995, 2000).

The functional dependency, $f(\mathbf{x})$, between independent variables $\mathbf{x}=\{x_1,x_2,\ldots x_l\}$, with $\mathbf{x}\in \boldsymbol{R}^K$, and (dependent) variable, $\boldsymbol{y}=\{y_1,y_2,\ldots y_l\}$, with $y\in \boldsymbol{R}$ is learned through the regularized functional:

Minimize $$\frac{1}{2}\| \boldsymbol{w} \|^2 + C \sum_{i=1}^{L}(\xi_i + \xi_i^*)$$

Subject to $$y_i - \sum_{j=1}^{K}\sum_{i=1}^{L} w_j x_{ji} - b \le \varepsilon + \xi_i ; \quad \sum_{j=1}^{K}\sum_{i=1}^{L} w_j x_{ji} + b - y_i \le \varepsilon + \xi_i^* ; \quad \xi_i, \xi_i^* \ge 0 \qquad (10)$$

where $f(\mathbf{x}) = \langle \boldsymbol{w}, \mathbf{x} \rangle + b$ with $\boldsymbol{w}\in \mathbf{x}$, $b\in \boldsymbol{R}$. $f(\mathbf{x})$ can be written as

$$f(\mathbf{x}) = \sum_{j=1}^{K} w_j \boldsymbol{x}_j + b \qquad (11)$$

where $\langle \boldsymbol{w}, \mathbf{x} \rangle$ denotes the dot product and K, the dimension of $\boldsymbol{w}$ and $\mathbf{x}$. b is the bias. In the above formulation, the errors with absolute value less than ε are ignored, making the solution sparse and hence "ε-insensitive". The first term in the objective function of (10) is a regularization term; the second term is the ε-insensitive loss function. The ξ_i and ξ_i^* are slack variables. For errors smaller than ε, ξ_i and ξ_i^* are not required to be non-zero and the point does not enter the objective function. The quantity 'C' controls the trade-off between minimizing the loss function and minimizing model complexity. Equation (10) is solved in dual form:

Maximize

$$\boldsymbol{w}(\boldsymbol{\alpha}^*, \boldsymbol{\alpha}) = \varepsilon \sum_{i=1}^{L}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{L} y_i(\alpha_i - \alpha_i^*) - \frac{1}{2}\sum_{i,j=1}^{L}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

Subject to constraints: $$\sum_{i=1}^{L}(\alpha_i^* - \alpha_i) = 0 \quad ; \qquad 0 \le \alpha_i, \alpha_i^* \le C \qquad (12)$$

where $i = 1,\ldots, L$ is the sample size and the approximating function is

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)k(\mathbf{x}, \boldsymbol{x}_i) + b \qquad (13)$$

Here $\boldsymbol{\alpha}^*$, $\boldsymbol{\alpha}$ are Lagrange multipliers, and $k(\mathbf{x}, \boldsymbol{x}_i)$ is the radial basis kernel function, which has the form: $k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(-\gamma|\boldsymbol{x}_i-\boldsymbol{x}_j|^2)$, wherein γ is the kernel parameter which is selected on the basis of a trial-and-error procedure. The $\boldsymbol{x}_i$'s are "support vectors" (SVs), and N is the number of support vectors corresponding to values of the independent variable that are at least ε away from actual observations. For $| f(\boldsymbol{x}_i)-y_i | \geq \varepsilon$, the Lagrange multipliers are non-zero and for the points inside the ε-tube, the parameters α^*_i, α_i vanish. In this paper, parameter 'C' is selected through a 10-fold cross validation technique. There are different criteria for selecting ε values in the literature.

A list of symbols is shown in Appendix B.

Study Area, Data Sets, Pre- and Post-Processing

Study area

The study area is the Walnut Creek watershed located at Ames, in south-central Iowa, USA. It is a small watershed in the heart of the Corn Belt with an area of about 5,130 hectares and is characterized by fairly level topography and rich soils that developed under prairie and prairie pothole wetlands. More than 80% of this watershed is planted to corn and soybean row crops. Figure 1(*a*) shows the location of the study area and Figure 1(*b*) shows the soil texture map of the Walnut Creek watershed.

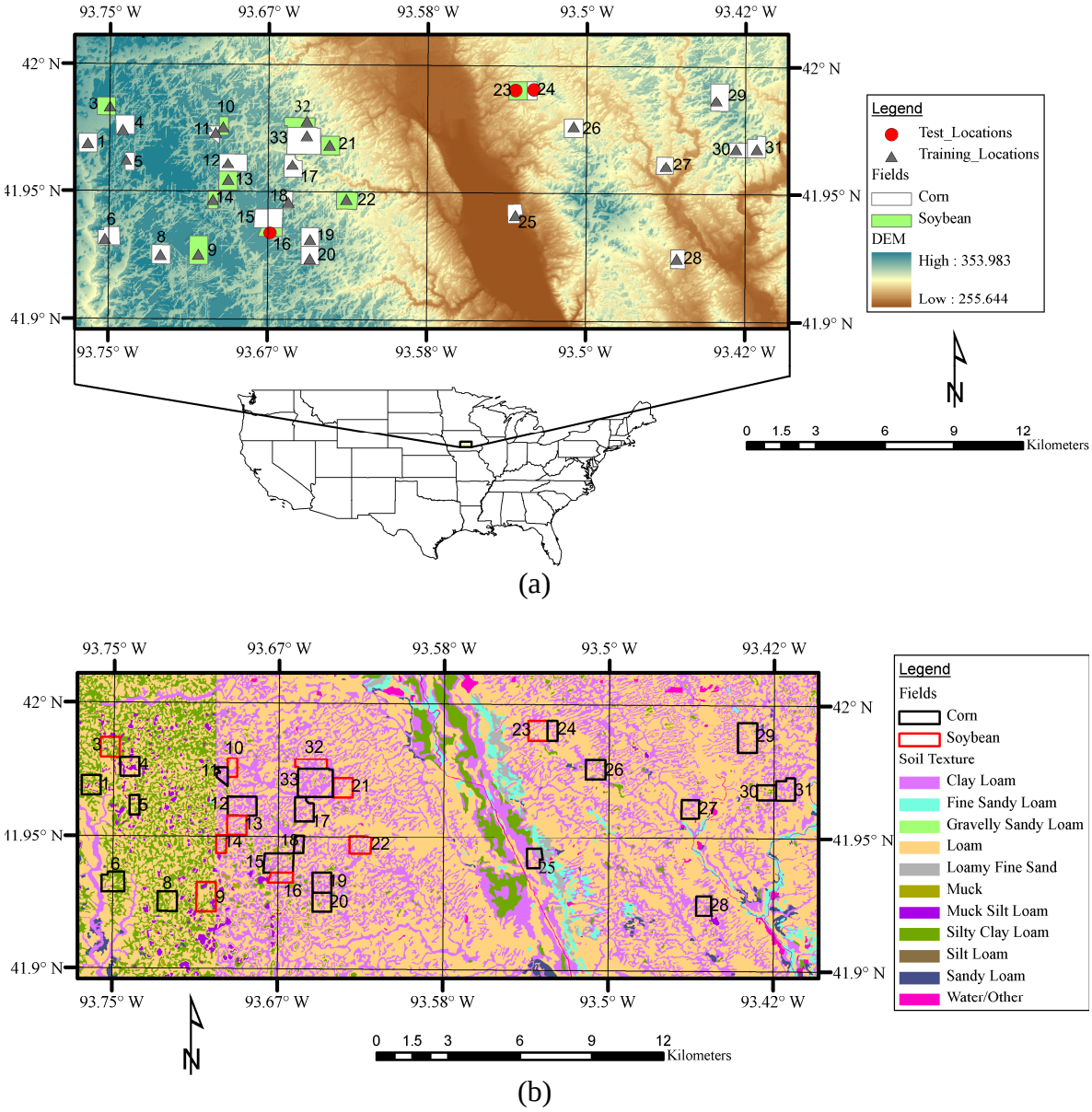


Figure 1. Maps of study area: (*a*) experimental fields, sampling locations and topography (Scale: 1:285,150 or 1 cm= 2.8515 km); (*b*) soil texture.

Datasets

In this study, reflectance data in the visible (Landsat channel 3: 0.63-0.69 µm), near-infrared (Landsat channel 4: 0.78-0.90 µm) and short-wave infrared (Landsat channel 5: 1.55-1.75 µm) were used to calculate vegetation indices, which were used as

model inputs. Land surface temperature inputs were derived from Landsat 5 and 7 thermal infrared bands, ~10.40-12.50 µm. Other inputs consisted of meteorological conditions (air temperature, precipitation), leaf area index (LAI), soil temperature, soil water holding capacity (mm/m), surface SMC (0-6 cm) and SMC at 30 cm depth.

The data used for this study were a part of the Soil Moisture Experiments (SMEX02) conducted at Ames, Iowa in 2002. The temporal coverage of the data was a one-month period between mid-June and mid-July and the spatial coverage was 41.52°N to 42.2°N, 93.23°W to 93.50°W. Brief descriptions of the five datasets from the experiments which were used for this research are provided here. More details regarding these can be found on the SMEX02 website.

Watershed soil moisture data, Walnut Creek, Iowa. These data result from daily measurements of volumetric SMC (0-6 cm) using a manually inserted probe and hand-held reader (Jackson and Cosh, 2003b) conducted at 31 moisture sampling sites in the Walnut Creek watershed (see Figure 1). The unit for volumetric SMC was cubic meter of water per cubic meter of soil (m^3/ m^3).

Landsat Thematic Mapper imagery, Iowa. The processed Landsat 5 and 7 Thematic Mapper imagery were the same as those used by Anderson et al. (2004). The dataset consisted of reflectance values for channels 3 (0.63-0.69 µm), 4 (0.78-0.90 µm) and 5 (1.55-1.75 µm) with a spatial resolution of 30 meters. Further description can be found in Jackson and Cosh (2003a). Landsat images with minimal cloud cover over the watershed acquired on 23 June, 1 July and 8 July 2002 were used. Table 1 shows the details of the cloudless scenes used for the analysis.

Table 1. Landsat scene details

Date (2002)	Day of year	Landsat	Path	Row	Watershed cloud cover (%)
June 23	174	5	26	31	0
July 1	182	7	26	31	0
July 8	189	7	27	31	0

The processed land surface temperature (LST) data set consisted of LST derived from the thermal band (~10.4-12.5 μm) of the same instrument. The unit of LST measurement was degrees Kelvin. Landsat 7 ETM+ data and Landsat 5 TM data had a spatial resolution of 60 m and 120 m respectively (for more details about this dataset, refer to Li et al., 2004).

Rain gauge network, Walnut Creek, Iowa. This data set, acquired during SMEX02 experiments, included hourly precipitation data in millimetres (mm) at 20 rain gauge stations distributed throughout the study area. Data were recorded from 1 June through 19 August 2002 (for more details about this dataset, refer to Prueger, 2004).

Soil moisture and temperature profiles, Walnut Creek, Iowa. Soil profile stations were deployed at four sites (15, 16, 23 and 24) in the Walnut Creek watershed. Sites 15 and 24 were corn fields and 16 and 23 were Soybean fields (see Figure 1). Volumetric water content (VWC) of soil was measured using a water content reflectometer (WCR) device. Soil temperature was measured in degrees Celsius (°C) using soil temperature probes (STP) at six depths: 2, 5, 10, 15, 20, and 30 cm.

SMEX02 Land surface information: Soils database. The Soils Database on the SMEX02 website included Environmental Systems Research Institute (ESRI) shapefiles containing soil classifications. The soils shapefile for the Walnut Creek watershed contained soil texture information of the study area and high, low and average available

water holding capacities (WHC) of the soils in inches per 5 ft. The average WHC was used as an input after converting its unit to millimetres/meter.

A preliminary analysis was done to visually inspect the interaction of different attributes. Figures 2-5 show the time series of daily rainfall, surface SMC (0-6 cm) and SMC at 30 cm depth for fields 15, 16, 23 and 24 from 24 June to 12 July 2002. Figures 2, 3, 4 and 5 show that the SMC at 0-6 cm and 30 cm depth had a decreasing trend. In all the cases, this downward trend decreased after the precipitation events. This showed that the precipitation events had an impact on SMC at 0-6 cm and 30 cm depth.

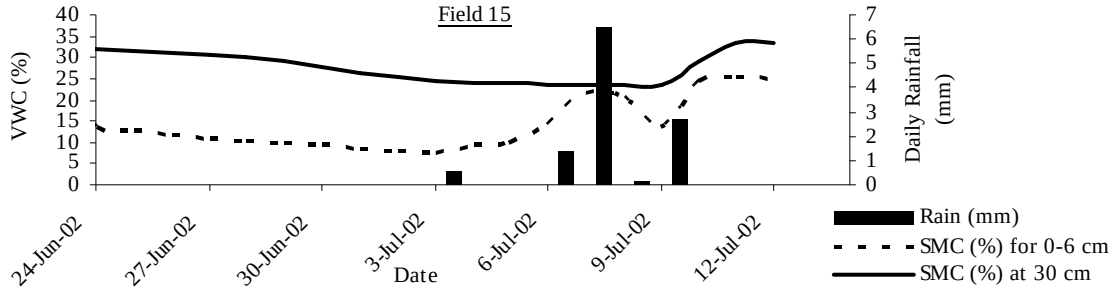


Figure 2.Variation of rainfall and SMC with time for 0-6 cm and 30 cm depth on field 15.

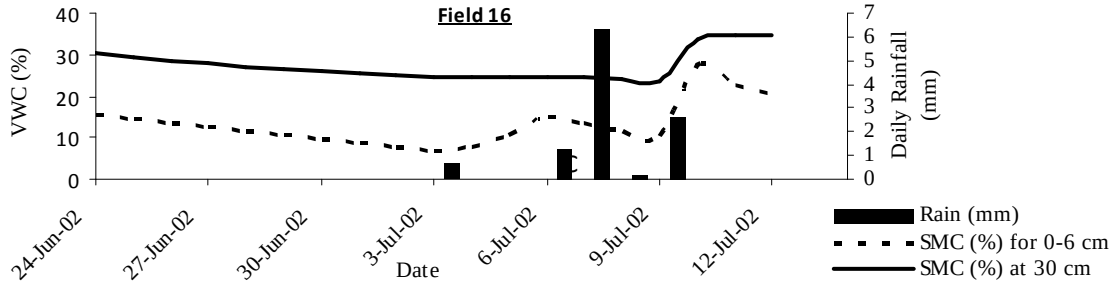


Figure 3.Variation of rainfall and SMC with time for 0-6 cm and 30 cm depth on field 16.

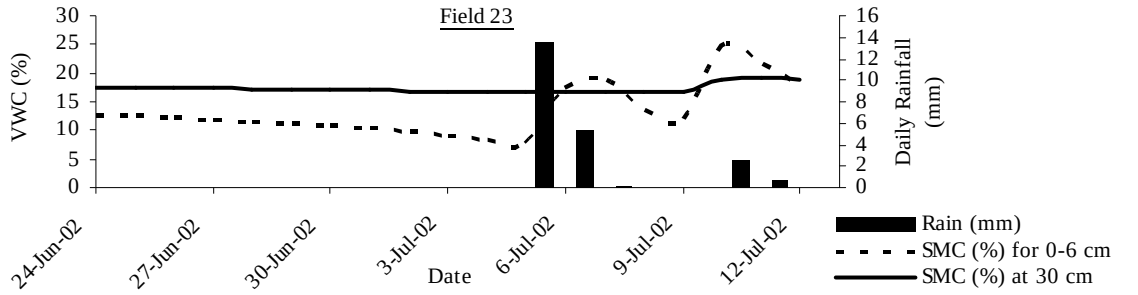


Figure 4.Variation of rainfall and SMC with time for 0-6 cm and 30 cm depth on field 23.

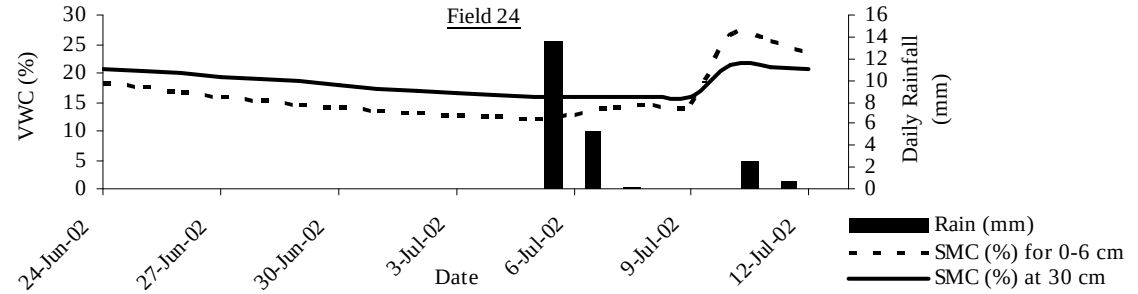


Figure 5.Variation of rainfall and SMC with time for 0-6 cm and 30 cm depth on field 24.

Pre-processing

As mentioned in section 3.2.2, processed Landsat 5 and 7 Thematic Mapper imagery were the same as those used by Anderson et al. (2004), and hence no pre-processing was required. To obtain the processed LST images the MODTRAN 4.1 radiative transfer model was used to correct the original Level-1G radiances for atmospheric effects; then the radiances were converted to LST. Details about the pre-processing method of LST imagery are described by Li et al. (2004).

Post-processing

Post-processing was done to obtain point values of SAVI, LAI, LST and soil water holding capacities which were used as inputs to the RVM and SVM models.

Spatial layer of vegetation index. Spatial layers of SAVI and NDWI were created in ERDAS Imagine using the Landsat images for 23 June, 1 July and 8 July 2002. The following equations were used.

$$SAVI = (R_{NIR} - R_{RED})\,(1 + L)/\,(R_{NIR} + R_{RED} + L) \quad (14)$$

$$NDWI = R_{NIR} - R_{SWIR} / R_{NIR} + R_{SWIR} \quad (15)$$

where, R_{NIR}, $R_{RED,}$ R_{SWIR} are the apparent reflectance values in the near-infrared (~0.8 µm), red (~0.6 µm) and short wave infrared (~1.2-25 µm) wavebands respectively. L is a calibration factor (Huete, 1988). The SAVI was one of the inputs to the learning machine models. The NDWI layer was created for use in (16) for estimation of LAI. The point values of SAVI were extracted corresponding to the latitude and longitude of the field soil moisture sampling locations (see Appendix A) using ArcGIS. SAVI was dimensionless.

Spatial layer of LAI. Figure 6 shows the three-band Landsat image (for 1 July 2002) of the study area with some land use superimposed on top. Figure 6 gives us an idea of the heterogeneity of the area. Field measurements for LAI were available but not at all points of interest. Interpolation of the field measurements was not possible due to spatial variability of the area. Hence, a spatial layer of LAI was created using the equation developed by Andersen et al. (2004) for the same study area:

$$Y = (a*VI + b) * (1 + c*exp[d*VI]) \quad (16)$$

where Y = LAI, a = 2.88, b=1.14, c=0.104, d = 4.1 are constants and VI = NDWI. For more details about (16), refer to Anderson et al. (2004). The point values of LAI were extracted from these spatial layers based on the latitude and longitude of the field soil moisture sampling locations. LAI was dimensionless.

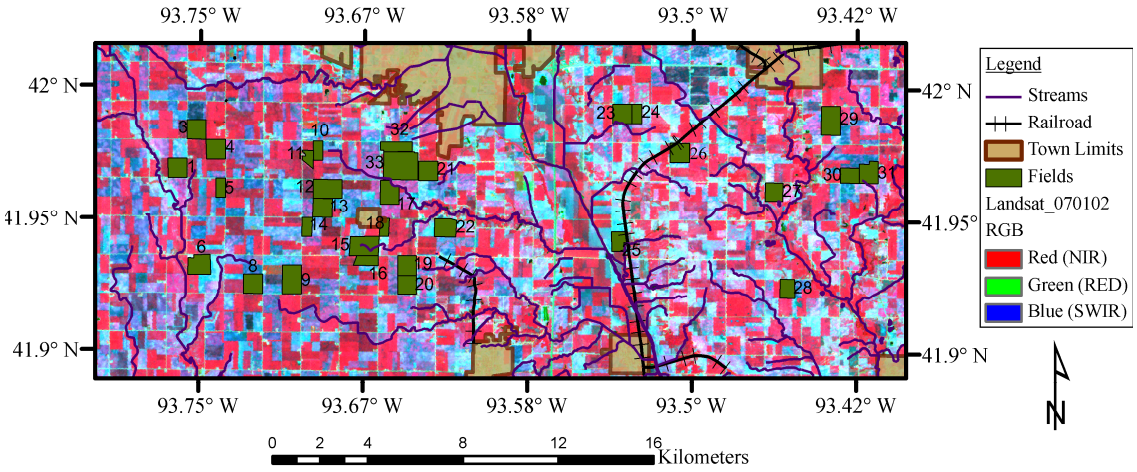


Figure 6. Landsat image of the study area superimposed with landuse (Scale: 1:318,000 or 1 cm= 3.18 km).

Land surface temperature (LST). The LSTs for 23 June, 1 July and 8 July 2002 were used as one of the inputs to the machines. The point values were extracted from the images using the latitude and longitude position of the field sampling locations.

Meteorological data. The air temperature for the study area was downloaded from the DAYMET U.S. data centre website. The website provides daily surface weather data and climatological summaries based on the latitude and longitude of the location.

The hourly precipitation data was added to get daily data. The precipitation data (see section 3.2.3) corresponding to the input points were obtained by creating a spatially interpolated precipitation layer using kriging in ArcGIS. The rainy days close to the 23 June, 1 July, and 8 July, 2002, dates were used and “days since it last rained” was included as one of the inputs. For 23 June and 1 July, the area received precipitation 11 days and 5 days prior to the acquisition of the Landsat image respectively. For 8 July, the area received rainfall for three days in a row i.e. the 5, 6, and 7 July. An average rainfall value was considered and “days since it last rained” was taken equal to 1.

Soil texture and water holding capacity. The GIS soils shapefile was used to extract data on soil texture and associated average WHC of the soils corresponding to the latitude and longitude of the field sampling locations. The maximum and minimum variation of WHC within a field was 46.83 mm/m and 20 mm/m respectively.

Attribute Selection

Achieving an optimal level of performance for any learning machine entails different design choices. The objectives of building optimal model architecture were to produce acceptable SMC estimates and to assure good model generalization abilities. Different models could be deduced given different training sets. However, for successful model construction, the training data set should carry enough symptomatic information about the processes involved. Here, one of the critical issues in preparing the training set was to select input variables that strongly influence soil moisture status. Previous studies have shown good correlation between SMC and remotely sensed surface temperature and vegetation index (Gillies and Carlson, 1995; Gillies et al., 1997; Sandholt, Rasmussen, and Andersen, 2002; Xiao, Zhang, and Tan, 2005). It was also reported that SMC is strongly influenced by climatic data, i.e. air temperature and recharge through precipitation (Young and Nobel, 1986; Coronato and Bertiller, 1996). Hence, SAVI, air temperature and precipitation were included as inputs.

The models examined in this research use visible, NIR, and SWIR reflectance values as input variables. Unlike the longer microwave wavelengths, the optical signals of these bands have limited ability to penetrate clouds and vegetation canopy, and are highly attenuated by the earth's atmosphere. Cloud contamination is therefore a problem

common to all optical techniques (Moran et al., 2004). The capacity for higher spatial resolution, broad coverage, multi-satellite sensor availability, high and regular revisit frequencies, and the possibility of real-time applications are however very promising (Verstraeten et al., 2006). With these things in mind, data in visible, NIR, and SWIR bands were used in this research. Price (1980) mentioned that remote sensing in the thermal-infrared represents a promising source of information because surface temperature is tightly coupled to surface moisture fluxes through the latent heat release of evaporation. Likewise, Humes et al. (1993), discussed that thermal IR band is sensitive to surface soil moisture. Hence Landsat thermal infrared derived LST was chosen as one of the inputs.

During the study period, the area had partial crop cover (Anderson, 2003). At canopy level, the reflectance is a combination of soil and vegetation reflectance. Daughtry et al. (2000), discussed the effect of soil brightness changes due to surface soil moisture on canopy reflectance for a range of leaf area indices. Further, plant reflectance is governed by leaf surface properties and internal structure, as well as by the concentration and distribution of biochemical components (Xu et al., 2007), such as pigments, nutrient contents (Ayala-Silva and Beyl, 2005), structural discontinuities, water content in fresh leaf and dry matter in a wilted leaf (Jacquemoud and Ustin, 2001). This assessment helps in detecting plant water stress which is related to the soil moisture stress (Waring and Cleary, 1967), plant regulation such as stomatal closure and reductions in photosynthesis rate (Jones, 2007) and atmospheric demands (Scott and Geddes, 1979). Also, variations in soil moisture cause plant stress which affects leaf area development (Meier and Leuschner, 2008). Keeping these findings in mind, LAI was included as one

of the inputs. Another important input was the soil temperature which is responsible for driving the heat fluxes and incorporates the effect of varying soil types (Gilman, 1980). Also, soil heterogeneity affects soil moisture content through variations in soil texture and soil water holding capacity (Jacobs et al., 2004). Hence, soil water holding capacity was included as an input.

The issue of estimating SMC of the soil profile from surface measurements has been investigated in a number of studies (Camillo and Schmugge, 1983; Entekhabi, Nakamura, and Njoku, 1994; Calvet, Noilhan, and Bessemoulin, 1998; Calvet and Noilhan, 2000; Walker, Willgoose, and Kalma, 2001a, b; Albergel et al., 2008). Hence for Model II which estimates SMC at 30 cm depth, topsoil moisture content (0-6 cm) was included as one of the inputs. After selecting these variables, the input matrix was prepared for training the learning machines.

Methodology

This research involves the development of three models for SMC estimation. Model I estimates surface (0-6 cm) SMC. Model II estimates SMC at 30 cm depth using field measurement of surface SMC as one of the inputs. Model III is a two-step model which combines models I and II and uses the surface SMC estimated by model I as one of the inputs to estimate SMC at 30 cm depth. The estimation at larger depths could have been attempted but the analysis was limited to 30 cm due to the non-availability of validation data.

Figure 7 illustrates the approach used for models I, II, and III in the training phase. The flow diagram shows the general inputs and the outputs.

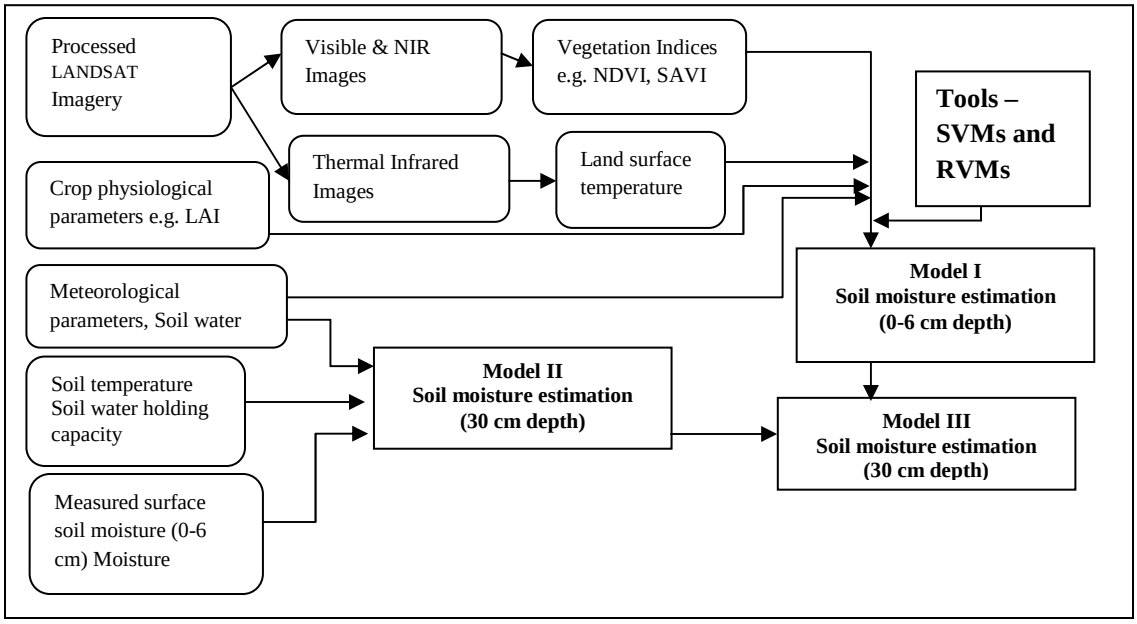


Figure 7. Flow diagram of model approaches in the training phase.

For model I, the surface SMC (0-6 cm) data acquired on 31 sites (see Figure 1) on three dates, i.e. 25 June, 1 July and 8 July 2002, were used to train the RVM and SVM models. The field measurement of surface SMC was not available on 23 June 2002 for all the sites; hence the data acquired on 25 June 2002 were used. A total of 93 points were available (31 sites and 3 dates). Site 30 had some missing data on 8 July 2002, so it was removed from the dataset. Twelve observations of surface SMC at 0-6 cm depth belonging to sites 15, 16, 23, and 24 were selected for testing model I. Site 15 had some missing data. Analysis was done using interpolated data but this deteriorated the results. Hence, the data points belonging to site 15 were removed, and only sites 16, 23 and 24 were used as test sites. Table 2 shows the coordinates of the test sample locations.

Table 2. Test point locations

Date (2002)	DOY	SiteID	Latitude	Longitude	Easting	Northing
June 25	176	WC16	41.9341	-93.6656	444821.6	4642675
June 25	176	WC23	41.9908	-93.5372	455505.1	4648888
June 25	176	WC24	41.991	-93.5276	456300.1	4648906
July 1	182	WC16	41.9341	-93.6656	444821.6	4642675
July 1	182	WC23	41.9908	-93.5372	455505.1	4648888
July 1	182	WC24	41.991	-93.5276	456300.1	4648906
July 8	189	WC16	41.9341	-93.6656	444821.6	4642675
July 8	189	WC23	41.9908	-93.5372	455505.1	4648888
July 8	189	WC24	41.991	-93.5276	456300.1	4648906

For model II, the daily data of SMC at 30 cm depth acquired between 23 June and 23 July 2002, on sites 16, 23 and 24 were used. Model II did not use any remotely sensed inputs. There were 90 available data points (30 days and 3 sites), out of which 9 were kept aside for testing. These 9 instances were the data acquired on 23 June, 1 July and 8 July 2002 (on which the Landsat images were available) on sites 16, 23 and 24.

The dataset for the first step of model III was similar to model I. For the second step, it was similar to model II. The training phase of the first and second step was similar to model I and II respectively. During the testing phase, the measured surface SMC (one of the inputs to the second step) was replaced by the modeled surface SMC estimates produced by first step. This test set was then used to estimate SMC at 30 cm depth.

Analysis

Model I. The RVM and SVM models were trained using seven inputs: crop physiological characteristic (i.e. LAI), remotely sensed inputs (vegetation index- SAVI, LST), meteorological inputs (air temperature, precipitation), number of days since it last rained and average WHC of the soil. Model output was surface SMC (0-6 cm). The

sample size was limited based on the availability of the Landsat images and correlating the image acquisition dates with the dates on which the remainder of the attributes were available.

Once the machine was trained, the surface SMC (0-6 cm) was estimated to check the model performance in the training phase. The data set that had been kept aside during the training phase was used in the testing phase.

As mentioned above, this study uses (16) for developing the spatial LAI layer and extracting LAI values at points of interest from the layer. Equation (16) uses NDWI as one of the inputs that is calculated using the SWIR (water absorption band). However, this does not limit the use of this model to the availability of SWIR band. For future work, either field measurements of LAI can be used or an empirical equation relating LAI to vegetation index (refer to Anderson et al., 2004) can be developed for the region of interest.

Model II. The second model was built using the actual field measurement of surface (0-6 cm) SMC as one of the inputs to the model. The other inputs were soil temperature for 0-6 cm depth, precipitation, number of days since it last rained and the average WHC of the soil. The output was SMC at 30 cm depth.

This model was more like an intermediate step in the development of model III. However, model II can be used independently if field measurements of surface SMC and temperature are available.

Model III. This model combines models I and II. In the first step the SVM and RVM models were trained in a similar manner as model I to estimate surface SMC. The second step was trained to estimate SMC at 30 cm depth, in a similar manner as model II.

In the test phase, the program automatically replaced the field measurements of surface SMC in the input set of the second step with the modeled values of surface SMC produced by the first step. The SVM and RVM models were tested with the estimated values of surface SMC to estimate SMC at 30 cm depth. This was done to simulate a situation where the field measurements of surface SMC are unavailable and the SMC at 30 cm depth has to be estimated using the remotely sensed inputs, crop physiological characteristics, soil WHC and meteorological data, similar to those used for model I. The general setup of the model looks like:

$$\underset{\mathbf{x}}{\textbf{Input}} \quad \rightarrow \quad \underset{\boldsymbol{y}}{\textbf{Output}}$$

For model I, **x** is the input matrix of size '*n* x *m*', where '*n*' represents the instances on each of the 31 sites for 3 different dates: '*n*' goes from 1 to 80 in the training phase, and from 1 to 9 in the test phase. For model II, **x** is the input matrix of size '*n* x *m*', where '*n*' represents the instances from 23 June to 23 July on sites 16, 23 and 24: '*n*' goes from 1 to 81 for the training phase and 1 to 9 in the testing phase for model II; '*m*' is the input dimension and equals 7 for model I and 5 for model II. The output is ***y***; this is a vector of dimension *p* x *1*, where '*p*' is surface SMC (0-6 cm) for the model I and SMC at 30 cm depth for the model II. It goes from 1 to 80 in the training phase and from 1 to 9 in the test phase for model I. For model II '*p*' goes from 1 to 81 for the training phase and 1 to 9 for the test phase.

<u>Evaluation of goodness-of-fit by comparing models to data</u>

To test the degree of association between the observed and estimated data, goodness-of-fit evaluation measures were used. The mean absolute error (MAE), a linear

measure and root mean square error (RMSE), a quadratic scoring rule, were used to measure average magnitude of error. The index of agreement (IoA) and coefficient of efficiency (CoE) were also used to check model performance. These statistics are calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}(y_i - f(\mathbf{x})_i)^2}{N}} \tag{17}$$

$$MAE = \frac{\sum_{i=0}^{n}\left|y_i - f(\mathbf{x})_i\right|}{N} \tag{18}$$

where N is the number of test samples. In this formulation, y and $f(\mathbf{x})$ are the measured and modeled values, respectively. The RMSE and MAE measure the error between the actual data and modeled values. Large values of RMSE or MAE mean that the difference between the actual measurements and the modeled values is large; hence the model is not performing well. Both the MAE and RMSE can range from zero to infinity. Lower values are better. The RMSE has the same dimensionality as the data and therefore is easy to interpret.

$$IoA = 1.0 - \frac{\sum_{i=1}^{N}(y_i - f(\mathbf{x})_i)^2}{\sum_{i=1}^{N}(\left|f(\mathbf{x})_i - \bar{y}\right| + \left|y_i - \bar{y}\right|)^2} = 1.0 - N\frac{MSE}{PE} \tag{19}$$

The IoA is calculated by comparing an observed group variance with an expected random variance. It varies from zero (inferior model) to one (excellent model). Potential Error (PE) is defined as the sum of the squared absolute values of the distances from $f(\mathbf{x})_i$ to $\bar{y}$ and from y_i to $\bar{y}$ and represents the largest value that it can attain for each actual observation/simulated value pair (Legates and McCabe, 1999).

$$CoE = 1.0 - \frac{\sum_{i=1}^{N}(y_i - f(\mathbf{x})_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (20)$$

CoE ranges from minus infinity (inferior model) to one (excellent model) (Nash and Sutcliffe, 1970). Thus, a value of zero for the CoE indicates that the observed mean, $\bar{y}$ is as good an estimator as the model, while negative values indicate that the observed mean is a better estimator than the model (Wilcox et al., 1990).

Results and Discussion

The goal of this research was to produce SMC estimates for 0-6 cm depth of topsoil and at 30 cm depth with RVMs and SVMs using a data assimilation technique. The following sections describe the selection of model parameters, RVM and SVM model performance and the bootstrap analysis.

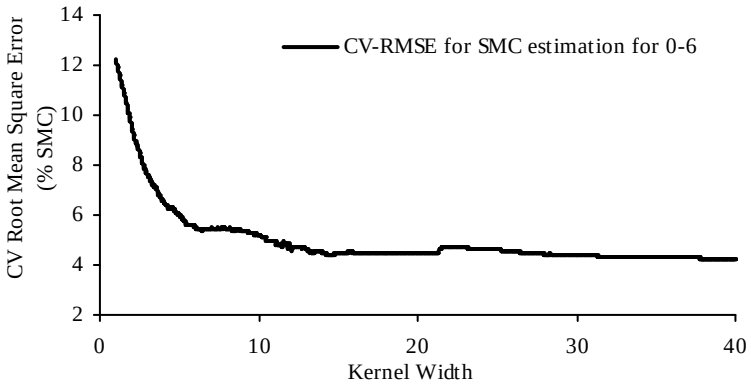
Model parameters

The RVM performance was tested based on selection of optimal kernel width and optimized iterations. The SVM performance was tested based on the selection of the cost parameter '*C*', the 'ε' insensitive parameter, and kernel parameter 'γ'. 'γ' was obtained through a trial-and-error procedure. The parameters '*C*' and 'ε' were tuned using 10-fold cross-validation. For the RVM, analogues of these parameters ($\boldsymbol{\alpha}$*'s* and σ^2) were automatically estimated by the learning procedure. The optimal kernel width '*r*' for the RVM model was obtained through 10-fold cross-validation.

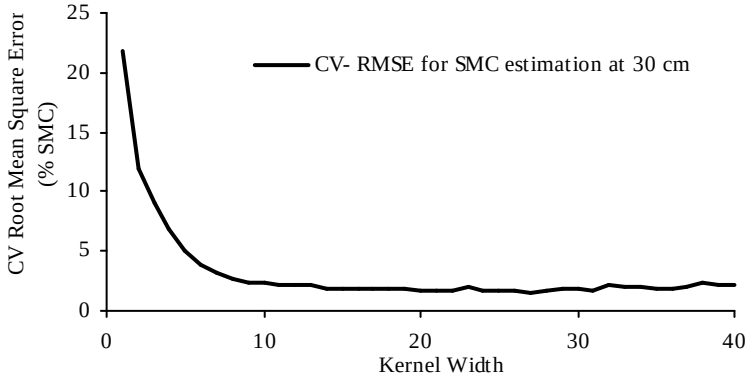
The RVM model cross validation results for surface SMC (0-6 cm) are shown in Figure 8(*a*). The least RMSE was obtained for a kernel width of *r* = 2.1. The number of

iterations was determined through a trial-and-error process and a value of 1500 was found to be optimal. The cross validation result for the RVM model for SMC estimation at 30 cm depth are shown in Figure 8(*b*). The least RMSE was obtained for $r = 3.2$. The optimal number of iterations, 1200, was determined through a trial-and-error process.

Figure 9(*a*) shows the cross-validation result of the SVM model for surface (0-6 cm) SMC estimation for determining parameter '*C*' (see Equation 10). The choice of the SVM cost parameter '*C*' can be critical for the model because it controls the trade-off between allowing training errors and forcing rigid margins. If the '*C*' value is increased, it forces the machine to create a more accurate model that may not generalize well.



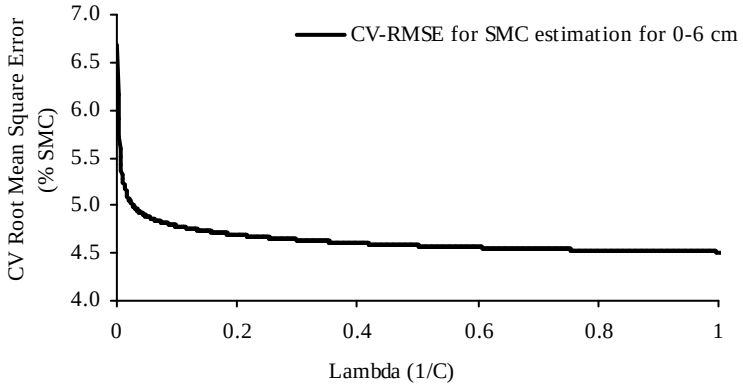
(*a*)



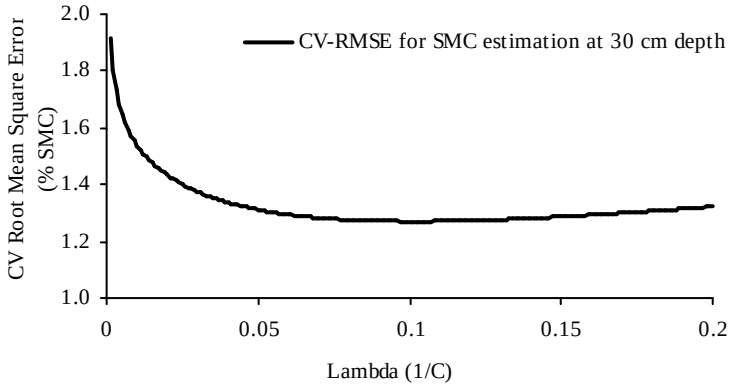
(*b*)

Figure 8. Ten- fold cross-validation results for the RVM model. RMSE vs. Kernel Width: (*a*) Model I – Estimation of surface soil moisture for 0-6cm depth; (*b*) Model II - Estimation of soil moisture at 30 cm depth.

The regularization parameter 'λ' corresponds to 1/*C*. The smallest value of RMSE for surface SMC estimation model was obtained at λ = 0.5. Figure 9(*b*) shows the cross-validation result of the SVM model for SMC estimation at 30 cm depth. The least RMSE for this model was obtained for λ = 0.02.



(*a*)



(*b*)

Figure 9. Ten- fold cross-validation results for the SVM model. RMSE vs. lambda (1/C): (*a*) Model I – estimation of surface soil moisture at 0-6 cm depth; (*b*) Model II - estimation of soil moisture at 30-cm depth.

Model performance

Model I results. The results for model I are shown in Figures 10(*a*-*h*). Both the RVM and SVM models performed well, and the modeled values of surface (0-6 cm) SMC followed the pattern of actual measurements. The goodness-of-fit test results for training and test data for both the RVM and SVM models are shown in Table 3.

The RVM model results for the training data are shown in Figures 10(*a*) and 10(*b*) and those of the SVM model are shown in Figures 10(*e*) and 10(*f*). The actual training data have maximum and minimum surface (0-6 cm) SMC of 36.2% and 4.3% respectively. The RVM demonstrated good performance with a training RMSE of 1.2%, IoA of 0.98 and CoE of 0.95 in the training phase as opposed to the SVM model, with a training RMSE of 4.3%, IoA of 0.72 and CoE of 0.87 (see Table 3), which indicated that observed data and modeled values were close. The SVM results for the training data were good (see Table 3), but the RVM results were better.

The RVM test results are shown in Figures 10(*c*) and 10(*d*) and those of the SVM are shown in Figures 10(*g*) and 10(*h*).

Table 3. Goodness-of-fit test results for Model I

Stage	**Statistic**	**SVM**	**RVM**
Training	RMSE	2.0%	1.2%
	MAE	1.6	1.01
	IoA	0.96	0.98
	CoE	0.87	0.95
Test	RMSE	4.3%	3.8%
	MAE	3.72	3.4
	IoA	0.72	0.83
	CoE	0.32	0.48

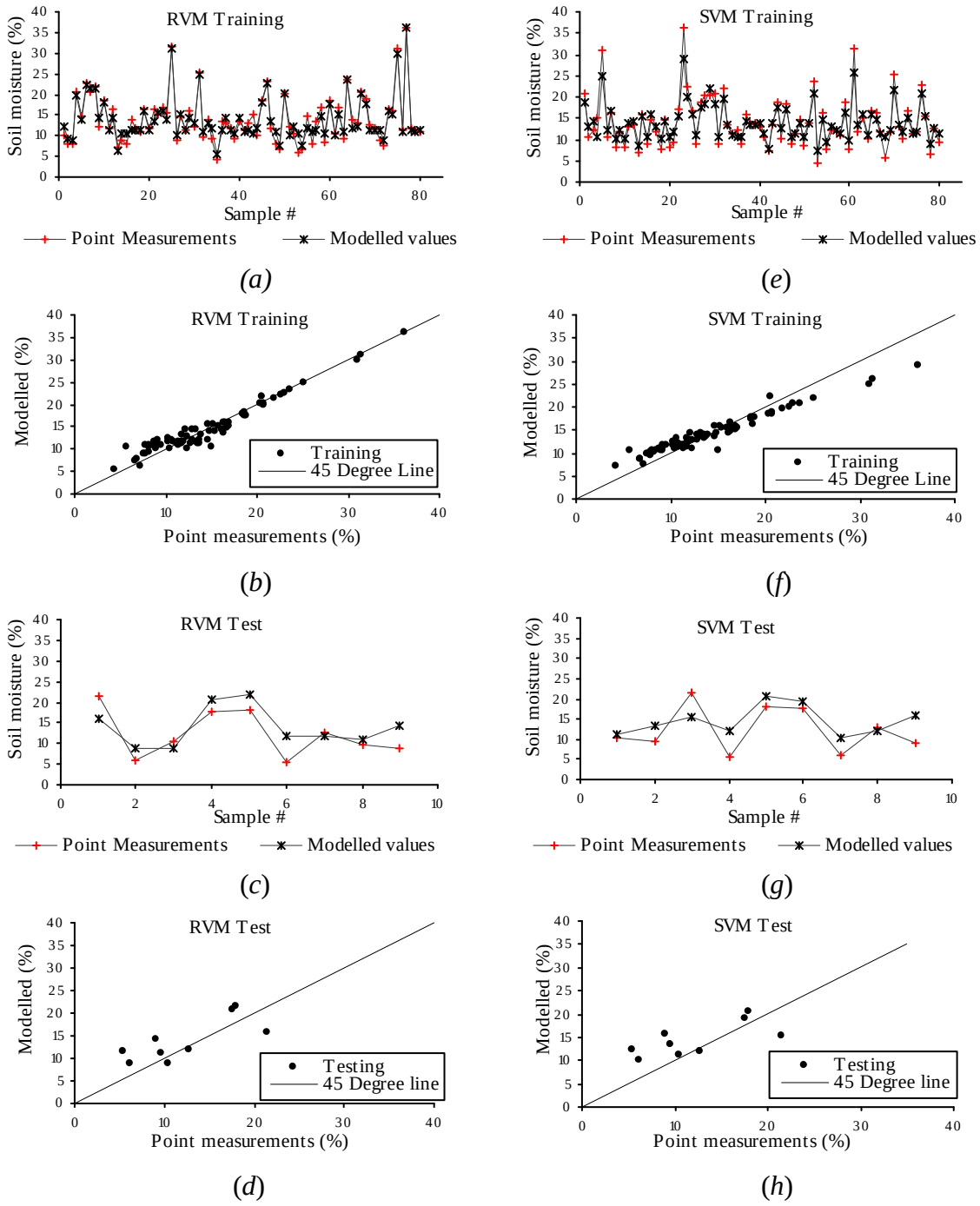


Figure 10. Model I results in terms of soil moisture (%) estimation at 0-6 cm depth: (*a*) model performance for RVM training set; (*b*) modeled versus measured values for RVM training set; (*c*) model performance for RVM test set; (*d*) modeled versus measured values for RVM test set; (*e*) model performance for SVM training set; (*f*) modeled versus measured values for SVM training set; (*g*) model performance for SVM test set; (*h*) modeled versus measured values for SVM test set.

The modeled values followed the pattern of actual measurements (see Figure 10(*c*) and 10(*g*)), which indicated good performance by the machines. The RVM model had a test RMSE of 3.8% and IoA of 0.83 as opposed to the SVM model, which had a test RMSE of 4.3% and IoA of 0.72 (see Table 3). The IoA values suggested that the total variance in the observed data was well explained by the model. The MAE was 3.4 for the RVM model and 3.72 for the SVM model, which suggested that the RVM and SVM model estimations differed from the data by 3.4% and 3.72% respectively. The number of relevance vectors (RVs) was 40 out of 80 training points and the number of support vectors (SVs) was 72 out of 80 training points. The smaller number of RVs as compared to SVs indicated a sparser RVM model. The modeled values and measured data points have been connected by lines just to illustrate the relative position of modeled and measured values (see Figures 10(*a*), (*c*), (*e*), and (*g*)).

Model II results. The results for model II are shown in Figures 11(*a*-*h*). Goodness-of-fit test results (see Table 4) showed that the RVM model performed better than the SVM version in both training and testing phases. The training results for the RVM model are shown in Figures 11(*a*) and (*b*) and those of the SVM model in Figures 11(*e*) and 11(*f*). The maximum and minimum field measurements of SMC at 30 cm depth were 35.2% and 15.9%, respectively. The RVM model had a training RMSE of 0.02%, IoA of 0.99 and CoE of 0.99 as opposed to SVM model which had a training RMSE of 0.36%, IoA of 0.99 and CoE of 0.99. Both the RVM and SVM models demonstrated excellent performance in the training phase. The number of RVs was 62 out of 81 training points and there were 74 SVs out of 81 training points. Again, the numbers of RVs used were fewer than SVs.

Figures 11(*c*) and 11(*d*) showed an excellent performance by the RVM model in the testing phase, and the pattern of the measured SMC was followed closely by the modeled values. Figure 11(*d*) shows that the points lie close to the 45 degree line which indicated that there was good agreement between modeled and measured values. The SVM test results are shown in Figures 11(*g*) and 11(*h*), indicating that the SVM model performs very well. For test data, the RVM gave a test RMSE of 0.48% with an IoA of 0.99, as opposed to the SVM, which gave a test RMSE of 0.66% and an IoA of 0.96 (see Table 4). The IoA values suggested that the total variance in the observed data was very well explained by the models. The MAE was 0.3 for the RVM and 0.32 for the SVM, which suggested that the RVM and SVM model estimations differed from the data by only 0.3% and 0.32%, respectively.

The results suggested that in situations where the field measurements of surface SMC are available, model II can be used independently for estimation of SMC at larger depths with high degree of accuracy using the learning machine models. It does not use remotely sensed inputs and infers the estimating function with surface SMC, soil WHC and meteorological data.

Table 4. Goodness-of-fit test results for Model-II

Stage	**Statistic**	**SVM**	**RVM**
Training	RMSE	0.36%	0.02%
	MAE	0.19	0.012
	IoA	0.99	0.99
	CoE	0.99	0.99
Test	RMSE	0.66%	0.48%
	MAE	0.32	0.3
	IoA	0.99	0.99
	CoE	0.98	0.99

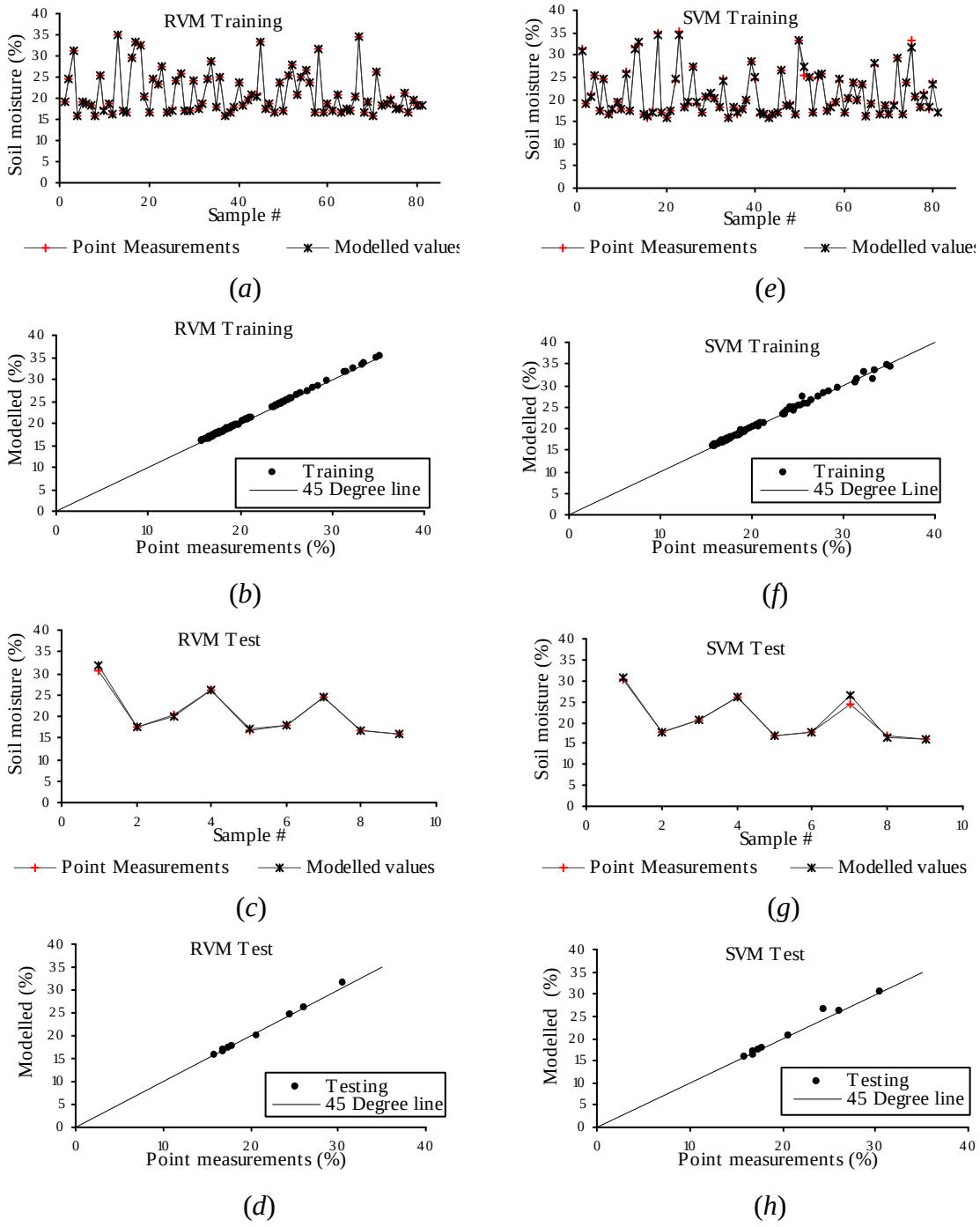


Figure 11. Model II results in terms of soil moisture (%) estimation at 30 cm depth: (*a*) model performance for RVM training set; (*b*) modeled versus measured values for RVM training set; (*c*) model performance for RVM test set; (*d*) modeled versus measured values for RVM test set; (*e*) model performance for SVM training set; (*f*) modeled versus measured values for SVM training set; (*g*) model performance for SVM test set; (*h*) modeled versus measured values for SVM test set.

Model III results. After testing the feasibility of estimating surface SMC and SMC at 30 cm depth in models I and II, a separate code was written for model III. It combined models I and II to estimate SMC at 30 cm depth. It was a two-step model where in the first step, the surface SMC was estimated in a similar manner as in model I. The second step was similar to model II, where the program automatically replaced the actual surface SMC values (one of the inputs to model II) in the test set by the surface SMC (0-6 cm) estimated by the first step and produced SMC estimates at 30 cm depth. The results showed that the modeled values for the test data were close to the actual SMC measurements.

The results for model III are shown in Figures 12*(a-h*). The goodness-of-fit test results (see Table 5) showed that the RVM performed better than the SVM in both the training and testing phases. The training results for the RVM (see Figures 12(*a*) and 12(*b*)), demonstrated excellent performance (R^2 = 0.915 and RMSE = 0.38%) with an IoA of 0.99 and CoE of 0.99, which indicated that modeled and measured values were very close in the training phase (see Table 5). The SVM model showed good training results.

Table 5. Goodness-of-fit test results for Model III

Stage	**Statistic**	**SVM**	**RVM**
Training	RMSE	0.76%	0.38%
	MAE	1.36	0.26
	IoA	0.99	0.99
	CoE	0.97	0.99
Test	RMSE	1.7%	1.4%
	MAE	1.36	0.98
	IoA	0.96	0.97
	CoE	0.87	0.92

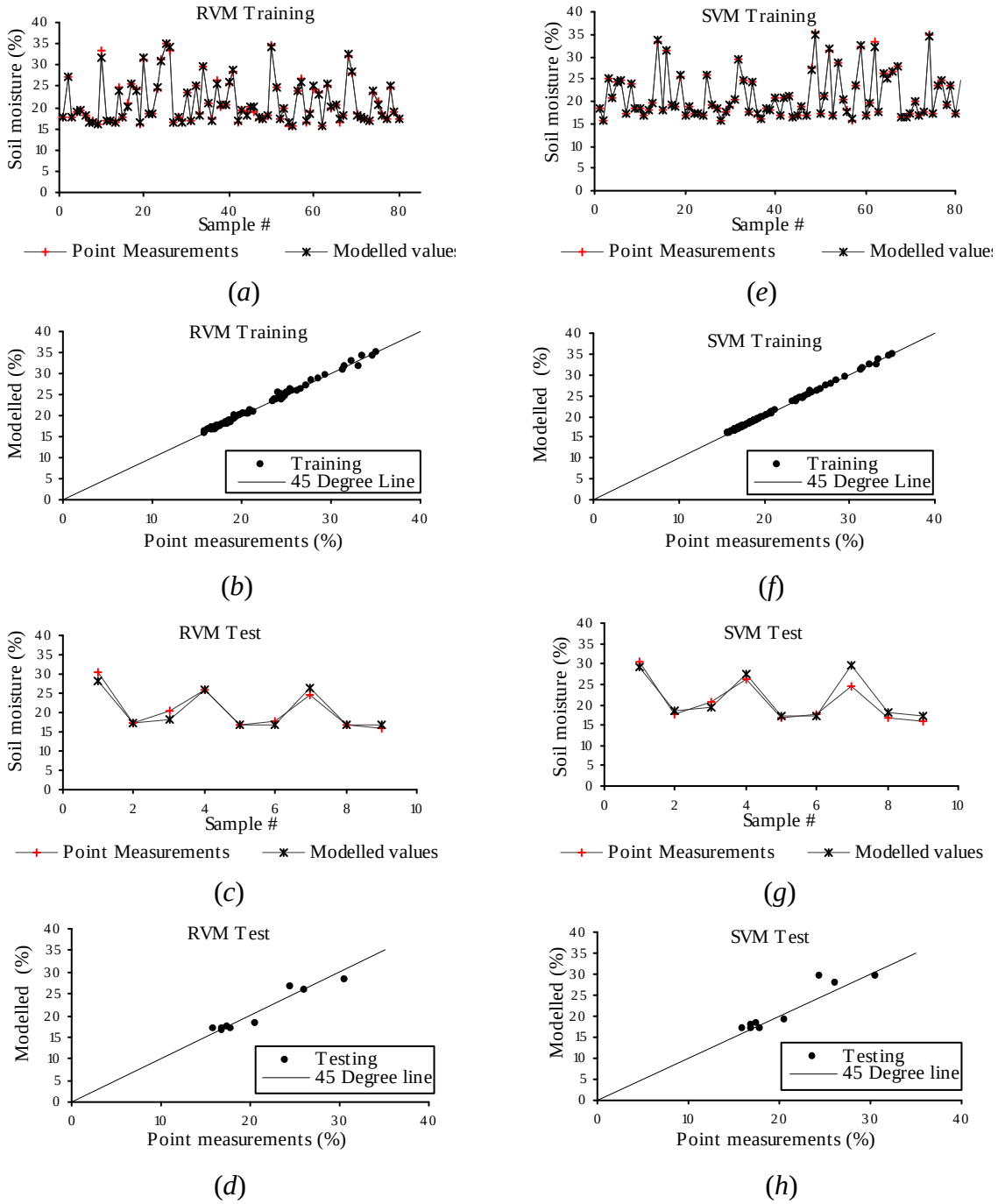


Figure 12. Model III results in terms of soil moisture (%) estimation at 30 cm depth: (*a*) model performance for RVM training set; (*b*) modeled versus measured values for RVM training set; (*c*) model performance for RVM test set; (*d*) modeled versus measured values for RVM test set; (*e*) model performance for SVM training set (*f*) modeled versus measured values for SVM training set; (*g*) model performance for SVM test set; (*h*) modeled versus measured values for SVM test set.

The RVM test results are shown in Figures 12(*c*) and 12(*d*) and that of the SVM are shown in Figures 12(*g*) and 12(*h*). The maximum and minimum field measurements of SMC at 30 cm depth were 35.2% and 15.9%, respectively. The RVM model had a test RMSE of 1.4% with IoA of 0.97, as compared to the SVM model, which has a test RMSE of 1.7% and an IoA of 0.96. The IoA values suggested that the total variance in the observed data was well explained by the model. The MAE was 0.98 for the RVM and 1.36 for the SVM model (see Table 5). This suggested that the RVM and SVM model estimations differed from the data by 0.98% and 1.36%, respectively. The number of RVs was 22 out of 81 training points and there were 68 SVs from 81 training points. Table 5 shows the training and testing results for model III.

As compared to the SVM models, the RVM models performed better in the testing phase in all the cases. It had been observed that learning machines are predisposed to overtraining. In the process of overtraining, the performance on the training examples still increases while the performance on unseen data deteriorates resulting in a loss of generalization capability. Generalization implies that the learning procedure is assumed to reach a state where it will also be able to estimate the correct output for a new set of data, thus generalizing to situations not presented during training. A better performance by the RVM models indicated that RVMs avoided overtraining and generalized better as compared to SVMs. This was also evident from the bootstrap results (see Figures 13 and 14). Also, each RVM model had fewer RVs than its corresponding SVM counterpart, and hence was sparser. The SVM attains a sparse structure by using structural risk minimization (Vapnik and Chervonenkis, 1991) while the RVM distributes the posterior probability mass over solutions with a small number of basis functions, and the given

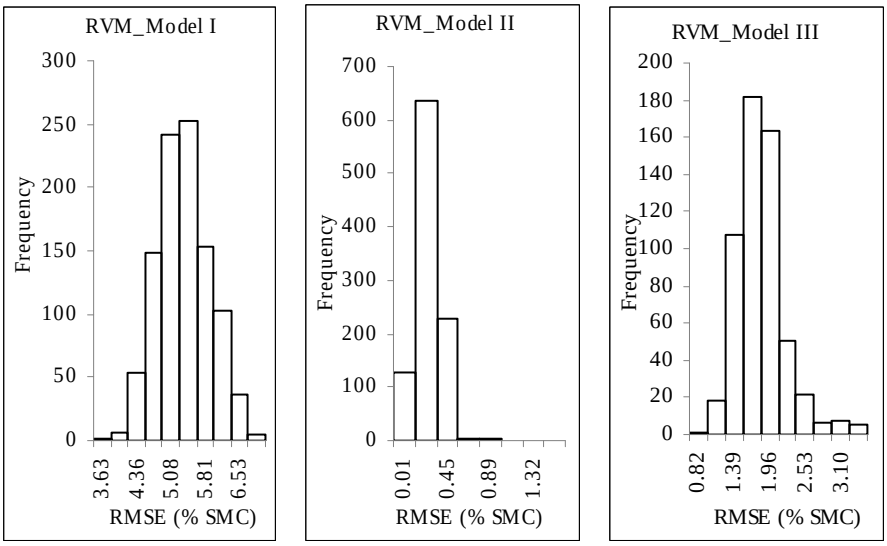
learning algorithm finds one such solution (Tipping, 2001). The RVM tends to capture the prototypical examples (i.e., those patterns that contained critical information about system dynamics and the underlying distribution) from the data.

The model results were also affected from different sources of error resulting from missing processes and parameters, errors in the measured data, approximations in the computation (e.g., numerical discrimination), temporal, spatial and scale variability, and overall model structure. Therefore, there were uncertainties attached when an attempt was made to quantify the model calculability. For example the models considered the average soil water holding capacities values for analysis. However, for some cases, the maximum variation within a single field was around 26 mm/m, which would have definitely affected the model results. Also there are some inherent errors in the estimated inputs, such as LAI which would definitely reflect in model outputs. Ultimately, the primary goal of hydrologic modeling (physically-based or data-driven) is to encapsulate the available knowledge of the underlying processes, together with the inherent uncertainties, to produce good estimates. From this viewpoint, the RVM model was best able to seize the information present in the data.

Bootstrapping

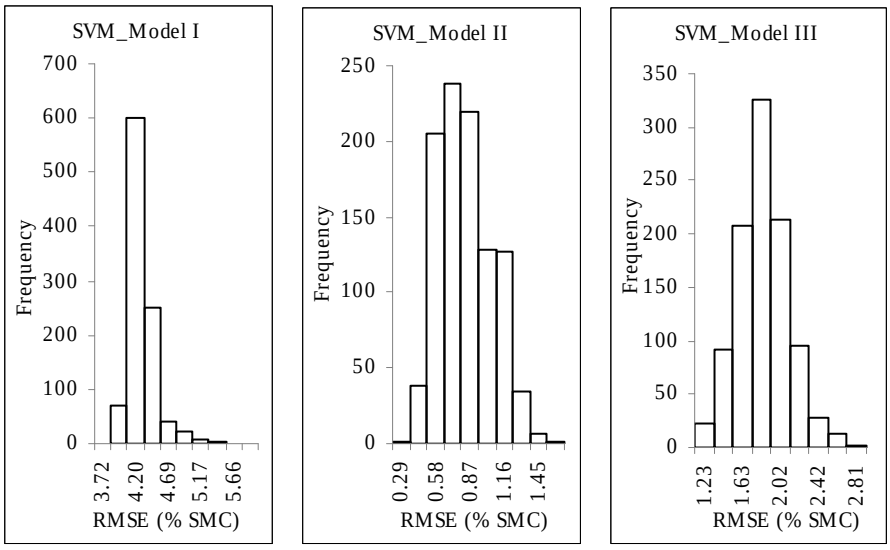
Bootstrapping was performed for both the RVM and SVM models to check over-fitting and model generalization capability. Figures 13 and 14 show bootstrap results for RMSE in percent SMC, as estimated from 1000 bootstrap samples. Conforming to the non-parametric approach, no assumption was made about the distribution of the data and repeated samples were drawn from the population with replacement. The basic idea was: if the sample is a good approximation of the population, the bootstrap method will

provide a good approximation of the sampling distribution of the statistic, in this case, the RMSE. Although beyond the scope of this paper, our goal here was to ensure good generalization of the inductive learning algorithm given scarce data and limited information. A narrow confidence interval indicated that the available training dataset was adequate to determine the machine parameters. From Figures 13 and 14, one could deduce rough confidence bounds that are more revealing of model performance than single values (Willmott, 1984). In Figures 13 and 14, we observed that the RMSE values for all the three RVM and SVM models are centered around one maximum value with highest frequency. Also RVM and SVM models show a fairly narrow confidence bound in all three modeling cases which imply that the models were robust and their parameters were well determined.



(*a*) (*b*) (*c*)

Figure 13. Bootstrap histogram of RMSE of the RVM models based on bootstrap analysis for the test phase of: (*a*) Model I: Estimation of surface soil moisture at 0-6 cm depth; (*b*) Model II: Estimation of soil moisture at 30 cm depth using ground measurement of surface soil moisture; (*c*) Model III: Estimation of soil moisture at 30 cm depth using estimated surface soil moisture (0-6 cm depth)



(*a*) (*b*) (*c*)

Figure 14. Bootstrap histogram of RMSE of the SVM models based on bootstrap analysis for the test phase of (*a*) Model I: Estimation of surface soil moisture at 0-6 cm depth; (*b*) Model II: Estimation of soil moisture at 30 cm depth using ground measurement of surface soil moisture; (*c*) Model III: Estimation of soil moisture at 30 cm depth using estimated surface soil moisture (0-6 cm depth)

Summary and Conclusions

The paper demonstrated a new technique for estimation of SMC using remotely sensed inputs as a part of a unified database that consisted of meteorological data, field measurements, and crop physiological factors. The results for model I (0-6 cm SMC estimation) showed that it was possible to get a good idea of the surface SMC by using RVMs and SVMs. Some sensitivity analysis revealed that apart from the reflectance data, soil temperature and precipitation were important contributors in inferring the estimating functions. Removing these inputs reduced the accuracy by almost 10%. Model II (SMC estimation at 30 cm depth) simulated a situation where surface SMC measurements were available and SMC at larger depths had to be estimated. The RVM results for model II were very close to the measured values of SMC which suggested that surface SMC acts

as a substantial input when estimating SMC at greater depths. This model can also be used independently for SMC estimation at larger depths using field measurements of surface SMC. Model III (SMC estimation at 30 cm depth) was built to simulate a situation where field measurements of surface SMC were unavailable but where remotely sensed data and other relevant surface information were available and the SMC at 30 cm depth had to be estimated. Results indicated that the RVMs performed better than the SVMs in all the test cases and hence demonstrated a better capability for capturing the underlying phenomena, showing good potential for SMC estimation. Computation of statistics of interest in conjunction with cross-validation and bootstrapping analyses accomplished a broad operational evaluation of the models that allows us to conclude that a remotely sensed data assimilation technique along with the learning machine model is a useful tool for soil moisture retrieval on large scales.

The procedures reported in this study are unique for four reasons: (i) They establish a link between readily available climatic variables and soil-moisture using SVMs and RVMs. (ii) They can provide estimates having resolution commensurate with remotely sensed data. The estimates can be made on finer to coarser scales depending on the data availability. (iii) The approach opens potential new horizons for modeling hydrologic systems using RVMs that might be built from limited amounts of data. (iv) The approach presented uses a concrete paradigm that is mathematically sound with manageable computational complexity. These sparse learning machines are theoretically elegant and well-regularized, in general require few parameters, and are relatively easy to calibrate.

One of the goals in developing these models was to assess the SMC during the growing season, which would give approximate information to the farmer/irrigator about the soil moisture status of the field. The results from this research are very encouraging, and we believe the SMC retrieval approach is worthy of attention because of the uniqueness previously mentioned.

Most importantly, information on soil moisture behavior helps in better management and understanding of hydrologic systems, irrigation scheduling and can result in improved forecasting, especially for agricultural basins. This research is a preliminary step in this direction. Future work would include the spatio-temporal forecasting of soil moisture in the root zone using RVMs.

CHAPTER III

SPATIO-TEMPORAL PREDICTION OF ROOT ZONE SOIL MOISTURE USING MULTIVARIATE RELEVANCE VECTOR MACHINES

Abstract

Root zone soil moisture at 1- and 2-meter depths are forecasted four days into the future. Prediction of soil moisture can be of paramount importance owing to its applicability in soil water balance calculations, various hydrometeorological, ecological, and biogeochemical modeling and initialization of various land-atmosphere models. In this article, we propose a new multivariate output prediction approach to root zone soil moisture assessment using learning machine models. These models are known for their robustness, efficiency, and sparseness; they provide a statistically sound approach to solving the inverse problem and thus to building statistical models. The multivariate relevance vector machine (MVRVM) is used to build a model that predicts future soil moisture states based upon current soil moisture and soil temperature conditions. The predicting function learns the input-output response pattern from the training dataset. Soil moisture measurements acquired by the Soil Climate Analysis Network (SCAN) site at Rees Center, Texas are used for this study. The methodology combines the data at different depths from 5 cm to 50 cm, the largest of which corresponds to the depth at which the soil moisture sensors are generally operational, to produce soil moisture predictions at larger depths. The MVRVM model demonstrates superior performance. The results for soil moisture predictions at 1 m and 2 m depth on the fourth day are excellent with RMSE = 0.0131 m^3_{water}/m^3_{soil} for 1 m; and RMSE = 0.0015 m^3/m^3 for 2 m

forecasted values. The statistics of predictions for fourth day (IoA = 0.96; CoE = 0.87 for 1 m and IoA = 0.99; CoE = 0.96 for 2 m) indicate good model generalization capability and computations show good agreement with the actual soil moisture measurements with R^2 = 0.88 and R^2 = 0.97 for 1 m and 2 m depths, respectively. The MVRVM produces good results for all four days. Bootstrapping is used to check over/under-fitting and uncertainty in model estimates.

Introduction

Root zone soil moisture is regarded as key factor governing surface water and energy balances and plays a vital role in hydroclimatic and environmental predictions. Soil moisture content (SMC) measurements are important for irrigation scheduling and crop yield forecast modeling, understanding rainfall/runoff generation processes. Information on soil moisture helps in explaining processes related to crop growth, forest dynamics and other vadose zone processes which play a vital role in water resources planning and management.

Soil moisture varies both in space and time because of spatial and temporal variations in precipitation, soil properties, topographic features, and vegetation characteristics (Das and Mohanty, 2006). The spatio-temporal prediction of SMC is difficult, though, capturing these variations and having an accurate estimation of soil moisture is necessary for soil and land survey (Webster and Butler, 1976; McKenzie and Austin, 1993), soil and land evaluation (Fu and Gulinck, 1994), hydrologic modeling and watershed management (Western et al., 1999; Qiu et al., 2003). Also, there is a need to develop methods for estimating SMC which make the best possible use of ancillary

information, particularly that which is relatively cheap to obtain (Moore et al., 1993; Lark, 1999; Qiu et al., 2001). Much work has been done in the past where soil moisture at larger depths was retrieved using surface soil moisture estimates (Camillo and Schmugge, 1983; Entekhabi, Nakamura, and Njoku, 1994; Calvet, Noilhan, and Bessemoulin, 1998; Calvet and Noilhan, 2000; Albergel et al., 2008). Sabater et al. (2007) state that surface soil moisture content is physically related to root-zone soil moisture through diffusion processes, and both surface and root-zone soil layers are commonly simulated by land surface models (LSMs). Li and Islam (2002) demonstrated the relationship between the soil moisture profile and surface soil moisture and fluxes. It was found that soil moisture can be predicted using low-level atmospheric and meteorological inputs (Mahfouf, 1991; Gill et al., 2006).

SMC retrieval using different techniques has been the subject of research for almost four decades. In general, soil moisture measurements are made as point measurements, mainly using gravimetric, nuclear, electromagnetic, tensiometric, or hygrometric techniques (Song et al., 2008), or by measuring SMC with imbedded sensors, such as time- and frequency-domain reflectometers (TDRs and FDRs). Physically based models for soil moisture estimation include the Soil-Plant-Atmosphere-Water (SPAW) model of Saxton, Johnson, and Shaw (1974) (Arora, Singh and Singh, 1997; Rao and Saxton, 1995; Hill and Neary, 2009), the U.S. Department of Agriculture Hydrograph Laboratory (USDAHL) model (Holtan et al., 1975; Comer and Henson, 1976), and the Sacramento Soil Moisture Accounting (SAC-SMA) Model (Peck, 1976; Sorooshian, Duan, and Gupta, 1993) used by the National Weather Service River Forecast System (NWSRFS) (Burnash and Singh, 1995), soil vegetation atmosphere

transfer (SVAT) models, among others. However, the difficulty associated with measurement of the physical parameters required by these models serves as an impediment.

This has furthered the interest of researchers to look for data-driven modeling tools such as artificial neural networks (ANNs) (Atluri, Chih-Cheng, and Coleman, 1999; Chang and Islam, 2000; Jiang and Cotton, 2004), higher order neural networks (Elshorbagy and Parasuraman, 2008), support vector machines (SVMs) (Gill, Kemblowski, and McKee, 2007; Yang and Huang, 2009). Gill et al. (2006) used soil moisture and meteorological data to generate SVM predictions for four and seven days ahead. The RVM and SVM models were used for forecasting soil moisture five days in the future by Khalil, Gill, and Mckee (2005). In the present study, we are applying a relatively new data-driven tool, the multivariate relevance vector machine (MVRVM) for soil moisture estimation. The purpose of this research was to develop a new model which forecasts soil moisture at different root zone depths, so both spatial and temporal predictions are done simultaneously. With this goal in mind, soil moisture at shallower depths, soil temperature and precipitation were used as inputs to a MVRVM model that forecasts soil moisture at large depths and for several days in the future. This model used available data acquired by the data collection station for previous days. The past measurements soil moisture data at shallower depths were used to train the machine. This learning machine tool automatically learns to recognize complex patterns that reside in data and that can be exploited to model input-output relationships. Therefore, this model was capable of recognizing a pattern between future soil moisture conditions and soil

moisture values in the past. This technique, which has generated promising results, has never been tried before.

Multi-variate Relevance Vector Machine (MVRVM)

"Sparse Bayesian Learning" is used to describe the application of Bayesian automatic relevance determination (ARD) concepts to models that are linear in their parameters. The motivation behind the approach is that one can infer a regression or classification model that is both accurate and sparse in that it makes its predictions using only a small number of relevant basis functions that are optimally selected from a potentially large initial set. A special case of this concept is the RVM which is applied to linear kernel models. The RVM was originally introduced by Tipping (2000). Thayananthan et al. (2006) proposed an extension of the sparse Bayesian regression model developed by Tipping and Faul (2003) and this extension enables a single relevance vector machine (RVM) to handle multiple output dimensions. The multivariate regression code developed by Thayananthan et al. (2006) is an open source code. This code was used as a base to build the MVRVM model which was particular to this application.

The data set is in the form of input-output pairs, $\{\mathbf{x}_n, \mathbf{t}_{rn}\}_{r=1,n=1}^{P,N}$, (P = number of output dimensions and N = number of observations). The major goal is to learn a model of dependency of the outputs on the inputs with the objective of making accurate predictions for previously unseen values of $\mathbf{x}$ (Tipping, 2001). Each output vector (t_r) is written as $t_r = (t_1, \ldots, t_N)^T$ and is expressed as the sum of an approximation vector

$y = (y(x_1), \ldots, y(x_N))^T$ and an "error" vector, the elements of which are considered as independent samples from some noise process $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)^T$:

$$t_r = y_r + \epsilon_r,$$

$$= \boldsymbol{\Phi} \boldsymbol{w}_r + \epsilon_r, \qquad (21)$$

where:

$\boldsymbol{w}_r$ = weight vector for the r^{th} component of the output vector t_r

$\boldsymbol{\Phi} = [\Phi(\mathbf{x}_1) \ldots \Phi(\mathbf{x}_M)]$, the N x M 'design' matrix whose columns comprise the complete set of M 'basis vectors'.

According to the sparse Bayesian approach (Tipping, 2001), "the errors are conventionally assumed to be zero-mean Gaussian, with variance σ_r^2. The parameter σ^2 is estimated from the data and the error model implies a multivariate Gaussian likelihood for the target vector t_r:

$$p(t_r \mid \boldsymbol{w}_r, \sigma_r^2) = (2\pi)^{-N/2} \sigma_r^{-N} \exp\{-\frac{\|t_r - y_r\|}{2\sigma_r^2}\}^2 \qquad (22)$$

"There are as many parameters in the model as training examples, therefore we would expect maximum likelihood estimation of $\boldsymbol{w}_r$ and σ^2 from (22) to lead to severe over-fitting" (Tipping, 2001). A prior constraint over $\boldsymbol{w}_r$ is imposed by adding a complexity penalty to the likelihood to avoid overfitting. The 'hyperparameters' are used to constrain an explicit zero-mean Gaussian prior probability distribution over the weights, $\boldsymbol{w}_r$:

$$p(\boldsymbol{w}_r \mid \alpha) = (2\pi)^{-M/2} \prod_{m=1}^{M} \alpha_m^{1/2} \exp -\left(\frac{\alpha_m \boldsymbol{w}_{rm}^2}{2}\right) \qquad (23)$$

Tipping and Faul (2003) introduced "M" independent hyperparameters, $\alpha = (\alpha_1, ..\alpha_M)^T$, each one individually controlling the strength of the prior over its associated

weight. It is this form of prior that ultimately makes the model sparse. Given α, the posterior parameter distribution conditioned on the data is given by combining the likelihood and prior within Bayes' rule:

$$p(\boldsymbol{w}_r \mid t_r, \alpha, \sigma_r^{\,2}) = p(\boldsymbol{t}_r \mid \boldsymbol{w}_r, \sigma_r^{\,2})\, p(\boldsymbol{w}_r \mid \alpha) / p(t_r \mid \alpha_r, \sigma_r^{\,2}) \tag{24}$$

and is Gaussian $N(\mu_r, \Sigma_r)$ with

$$\Sigma_r = (A + \sigma_r^{-2}\, \boldsymbol{\Phi}^T\, \boldsymbol{\Phi})^{-1} \qquad \mu = \sigma_r^{-2}\, \Sigma_r\, \boldsymbol{\Phi}^T\, t_r,$$

where A is defined as diag $(\alpha_1, \ldots, \alpha_M)$.

The algorithm proposed by Thayananthan et al. (2006) for training an RVM with multivariate outputs by finding the optimal hyperparameters is as follows:

1. Initialization of the noise variance σ_r and the hyperparameter α :

σ_r = variance of t_r x 0.1, r ϵ1….P

α = infinity (∞)

P = number of output dimensions

2. Iterate

2.1. Compute $\{\mu_r, \Sigma_r\}_{r=1}^{P}$

$$\Sigma_r = (\sigma_r^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mathbf{A})^{-1} \tag{25}$$

$$\mu_r = \sigma_r^{-2}\Sigma_r\boldsymbol{\Phi}^T t_r \tag{26}$$

2.2. Compute $\{s_{ri}, q_{ri}\}_{r=1,i=1}^{P,M}$ using,

$$s_{ri} = \frac{\alpha_i S_{ri}}{\alpha_i - S_{ri}} \text{ And } q_{ri} = \frac{\alpha_i Q_{ri}}{\alpha_i - S_{ri}}$$

$$S_{ri} = \sigma_r^{-2}\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i - \sigma_r^{-4}\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}\Sigma_r\boldsymbol{\Phi}^T\boldsymbol{\Phi}_i \tag{27}$$

$$Q_{ri} = \sigma_r^{-2}\boldsymbol{\Phi}_i^T t_i - \sigma_r^{-4}\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}\Sigma_r\boldsymbol{\Phi}^T t_i \tag{28}$$

M = number of basis functions

2.3. Find the basis function, Φ_m, and the corresponding optimal hyperparameter α_m^{opt} that minimize $L(\alpha)$ using the above Equations:

$$\alpha_i^{opt} = \arg \min_{\alpha i} l(\alpha_i)$$

$$= \arg \min l(\alpha_i) \left[\sum_{r=1}^{M} \left\{ \log \alpha_i - \log(\alpha_i + s_{ri}) + \frac{q_{ri}^2}{\alpha_i + s_{ri}} \right\} \right] \quad (29)$$

$$m = \arg \min_{i} l(\alpha_i^{opt})$$

If $\alpha_m^{old} = \infty$ and $\alpha_m^{opt} < \infty$, then add Φ_m to the model with $\alpha_m = \alpha_m^{opt}$

If $\alpha_m^{old} < \infty$ and $\alpha_m^{opt} = \infty$, then remove the Φ_m from the model with $\alpha_m = \infty$

If $\alpha_m^{old} < \infty$ and $\alpha_m^{opt} < \infty$, then update α_m with α_m^{opt}

2.4. Re-estimate the noise parameters using,

$$\sigma_r^2 = \frac{\left\| t_r - \boldsymbol{\varphi} \boldsymbol{\mu}_r \right\|^2}{\boldsymbol{M} - \sum_i^M \gamma_i} \qquad r \in 1 \ldots P \quad (30)$$

The optimal hyperparameters and the noise parameters are then used to obtain the optimal weight matrix:

$$\mathbf{A}^{opt} = \text{diag}(\alpha_1^{opt}, \ldots, \alpha_M^{opt}) \quad (31)$$

$$\Sigma_r^{opt} = ((\sigma_r^{opt})^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{A})^{-1} \quad (32)$$

$$\mu_r^{opt} = (\sigma_r^{opt})^{-2} \Sigma_r^{opt} \boldsymbol{\Phi}^T t_r \quad (33)$$

Readers interested in detailed descriptions of the model are referred to Thayananthan et al. (2006).

Data Description

The data used for this study were taken from the Soil Climate Analysis Network (SCAN) site at Rees Center, Texas, USA. There are about 86 SCAN stations across the United States where daily and hourly measurements for meteorological and soil moisture data are made using various sensors and instruments. Of these 86 stations, most collect soil moisture data up to a depth of 40 inch (around 100 cm), and there a few which collect soil moisture data up to a depth of 80 inch (around 200 cm). Rees Center, Texas is one such SCAN station which collects soil moisture data up to a depth of 200 cm.

In this particular application, meteorological inputs (precipitation and soil temperature) and soil moisture data were used. The location of the data collection station at Rees Center, Texas is 33° 37' N and 102° 02' W, at and elevation of 3333 feet (1015.9 meter (m)) (Figure 15). The period of record is from March 10, 2005 to present.

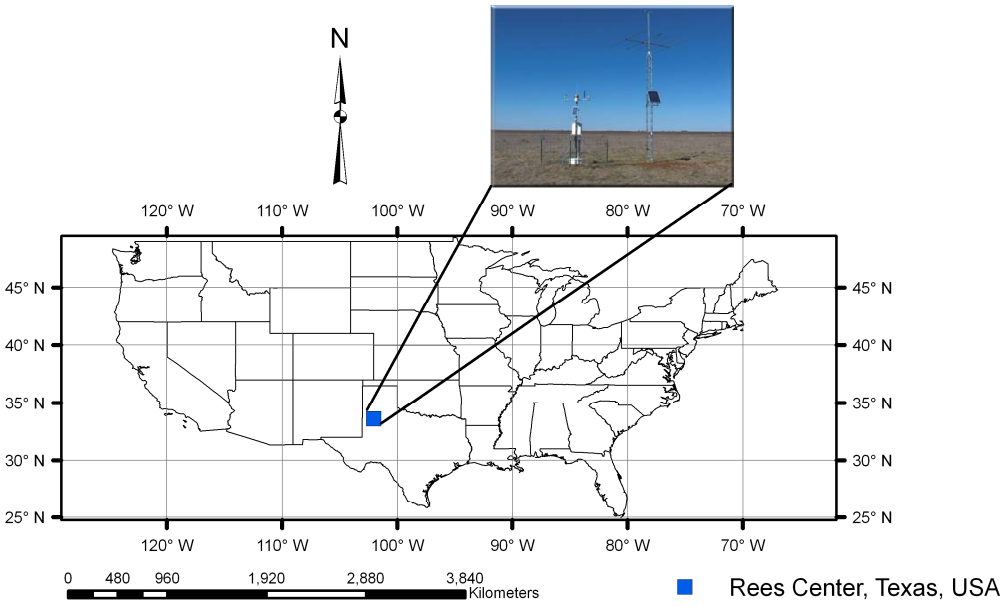


Figure 15. Location of data collection station.

Methodology

Our goal was to forecast root zone soil moisture at 1 and 2 m depths. This was done by assimilating soil moisture (m^3/m^3) at shallower depths (5, 10, 20, 30 and 50 cm), soil temperature (Celsius) and precipitation (mm), and predicting soil moisture at depths of 1 and 2 m.

The time series soil moisture data for 12 months were downloaded from the Natural Resource Conservation service (NRCS) website. A stratified sampling of the 12 months of data was carried out and the training and testing data were extracted from this stratified sample. The test data were kept aside for validating the performance of the machine. The stratified sampling was done to train the MVRVM model for different values of soil moisture in different seasons. It was observed that normalization of the data between -1 and 1 produced better results as compared to the case where raw data were used. Hence the data were normalized.

The MVRVM model was trained with 227 days of soil moisture and corresponding soil temperature and meteorological data. The inputs to the model were precipitation, soil temperature, and soil moisture data on days “d-4”, “d-3”, “d-2”, “d-1”. The output of the model was forecasted soil moisture values at “d”, “d+1”, “d+2” and “d+3”. Time steps were in days. The performance of the model was tested with 100 days of input data. Figure 16 shows the flow diagram of the model approach in the training phase.

Three analyses were done with different inputs. For the first analysis, the MVRVM model was trained using soil moisture at depths of 5, 10, 20, 30, and 50 cm below ground surface, soil temperature (Celsius), and precipitation (mm) as inputs (see

Figure 16). The input data for four days in the past: d-4, d-3, d-2, and d-1, were used. The second analysis was similar to the first one but it only used soil moisture at 5 and 10 cm, soil temperature (Celsius), and precipitation (mm) as inputs to train the MVRVM model. The third analysis used soil moisture at 30 and 50 cm, soil temperature (Celsius), and precipitation (mm) as inputs to train the MVRVM model. The output in all three cases was soil moisture at 1 and 2 m depths on d, d+1, d+2 and d+3, i.e. four days into the future

The latter two analyses were carried out to observe the variation in the model output when moisture in the topsoil (5, 10 cm) and then at larger depths (30, 50 cm) were used to train the learning machine. Figure 16 shows the MVRVM model approach. The model inputs were:

$\mathbf{x} = [XP_{d-4}, XS_{d-4}, XT_{d-4}, XP_{d-3}, XS_{d-3}, XT_{d-3}, XP_{d-2}, XS_{d-2}, XT_{d-2}, XP_{d-1}, XT_{d-1}, XS_{d-1}]$

where, d= time (day)

and the model outputs were : $\boldsymbol{y} = [Y1_d, Y2_d, Y1_{d+1}, Y2_{d+1}, Y1_{d+2}, Y2_{d+2}, Y1_{d+3}\, Y2_{d+3}]$

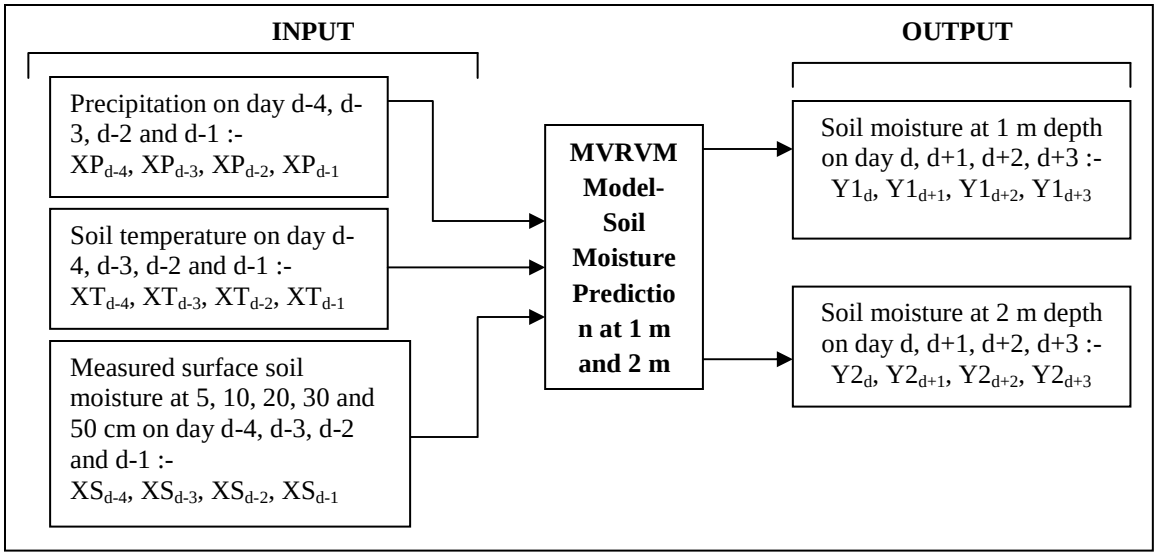


Figure 16. Flow Diagram for MVRVM model approach.

To test the degree of association between the observed and estimated data, goodness-of-fit evaluation measures were used. The mean absolute error (MAE), a linear measure and root mean square error (RMSE), a quadratic scoring rule, were used to measure average magnitude of error. The index of agreement (IoA) and coefficient of efficiency (CoE) were also used to check model performance. These statistics are calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}(y_i - f(x)_i)^2}{N}} \tag{34}$$

$$MAE = \frac{\sum_{i=0}^{n}\left|y_i - f(x)_i\right|}{N} \tag{35}$$

where N is the number of testing samples. In this formulation, y and $f(x)$ are the measured and modeled values, respectively. The RMSE and MAE measure the error between the actual data and modeled values. Large values of RMSE or MAE mean that the difference between the actual measurements and the modeled values is large; hence the model is not performing well. Both the MAE and RMSE can range from zero to infinity. Lower values are better. The RMSE has the same dimensionality as the data and therefore it is easy to interpret.

$$IoA = 1.0 - \frac{\sum_{i=1}^{N}(y_i - f(x)_i)^2}{\sum_{i=1}^{N}(\left|f(x)_i - \bar{y}\right| + \left|y_i - \bar{y}\right|)^2} = 1.0 - N\frac{MSE}{PE} \tag{36}$$

The IoA is calculated by comparing an observed group variance with an expected random variance. It varies from zero (inferior model) to one (excellent model). Potential Error (PE) is defined as the sum of the squared absolute values of the distances from $f(x)_i$

to $\bar{y}$ and from y_i to $\bar{y}$ and represents the largest value that it can attain for each actual observation/simulated value pair (Legates and McCabe, 1999).

$$CoE = 1.0 - \frac{\sum_{i=1}^{N}(y_i - f(x)_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (37)$$

CoE ranges from minus infinity (inferior model) to one (excellent model) (Nash and Sutcliffe, 1970). Thus, a value of zero for the CoE indicates that the observed mean, $\bar{y}$, is as good an estimator as the model, while negative values indicate that the observed mean is a better estimator than the model (Wilcox et al., 1990).

Results and Discussion

The goal of this research was to obtain spatio-temporal estimates of root zone soil moisture four days into the future at depths of 1 and 2 m below ground surface. A MVRVM was used to build the model. This section discusses the selection of model parameters, MVRVM model performance, and the bootstrap analyses.

Evaluation of RVM performance was based on selection of optimal kernel width and optimized iterations. Several trials were performed for obtaining the optimal values of these parameters. For the MVRVM, the parameters $\boldsymbol{\alpha}$ and σ^2 were automatically estimated by the learning procedure. The optimal kernel width for the MVRVM model was obtained through a trial and error, and the optimal number of iterations was obtained by plotting the parameter beta against the number of iterations (see Figure 17).

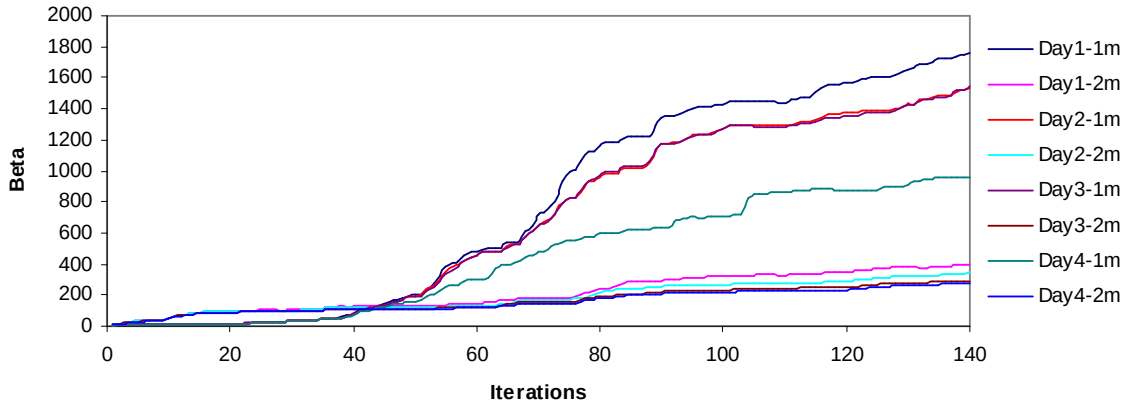


Figure 17. Variation of parameter beta with number of iterations for different outputs.

The number at which the value of parameter beta became almost constant was considered to be optimal.

The MVRVM model exhibits good performance. Figures 18 and 19 show predicted outputs versus original data and the confidence bounds for the test phase. The data from four days in the past were used to predict soil moisture estimates four days into the future.

The results produced by the MVRVM model in the testing phase for root zone SMC estimate at 1 m depth for four consecutive days are shown in Figures 18a - 18d. The MVRVM model showed good results with the forecasted values of soil moisture closely following the pattern of the field measurements. Table 6 shows the goodness-of-fit test results for the test set. The average maximum value of soil moisture at 1 m depth is about 30%, and the minimum is about 15%. The correlation result for the MVRVM model on the fourth day at 1 m depth (see Table 6) demonstrated good performance (R^2 = 0.877 and RMSE = 1.31%), with an IoA value of 0.96 and the CoE of 0.87. This indicated that the observed data and modeled values were close. The bias is very small indicating that

the estimator is robust. The average MAE for 1 m depth was 0.5 which suggested that model estimates differed from the data on an average by 0.50%.

The results produced by the MVRVM model in the testing phase for root zone SMC estimates at a depth of 2 m for four consecutive days in the future are shown in Figures 19a - 19d. Again, the MVRVM model showed excellent results with forecasted values closely following the pattern of the time series. Table 6 shows the goodness-of-fit test results for test data. The average maximum value of soil moisture at 1 m depth is about 19%, and the minimum is about 15%. The correlation result for the MVRVM model on the fourth day at a depth of 2 m (see Table 6) again demonstrated good performance (R^2 = 0.968 and RMSE = 0.15%), with an IoA value of 0.99 and a CoE value of 0.97. This indicated that observed data and modeled values were very close. The average MAE for 1 m depth was 0.08, which suggested that model estimates differed from the data by an average of only 0.08%.

The number of relevance vectors (RVs) used in the MVRVM model was 81 out of 227 training points.

Table 6. MVRVM results (Kernel Width, r = 3, Iterations = 140)

Statistics	Multivariate Relevance Vector Machine Model							
	Day d		Day d+1		Day d+2		Day d+3	
	1 m depth	2 m depth	1 m depth	2 m depth	1 m depth	2 m depth	1 m depth	2 m depth
RMSE, %	1.14	0.13	0.99	0.12	1.27	0.13	1.31	0.15
R^2	0.904	0.974	0.926	0.977	0.873	0.972	0.877	0.968
CoE	0.898	0.972	0.92	0.974	0.870	0.970	0.869	0.965
IoA	0.971	0.993	0.977	0.993	0.962	0.992	0.96	0.991
Bias	-0.1548	0.0168	-0.1168	0.0178	0.0251	0.0107	0.0151	0.0193
MAE	0.51	0.08	0.44	0.08	0.53	0.084	0.55	0.087

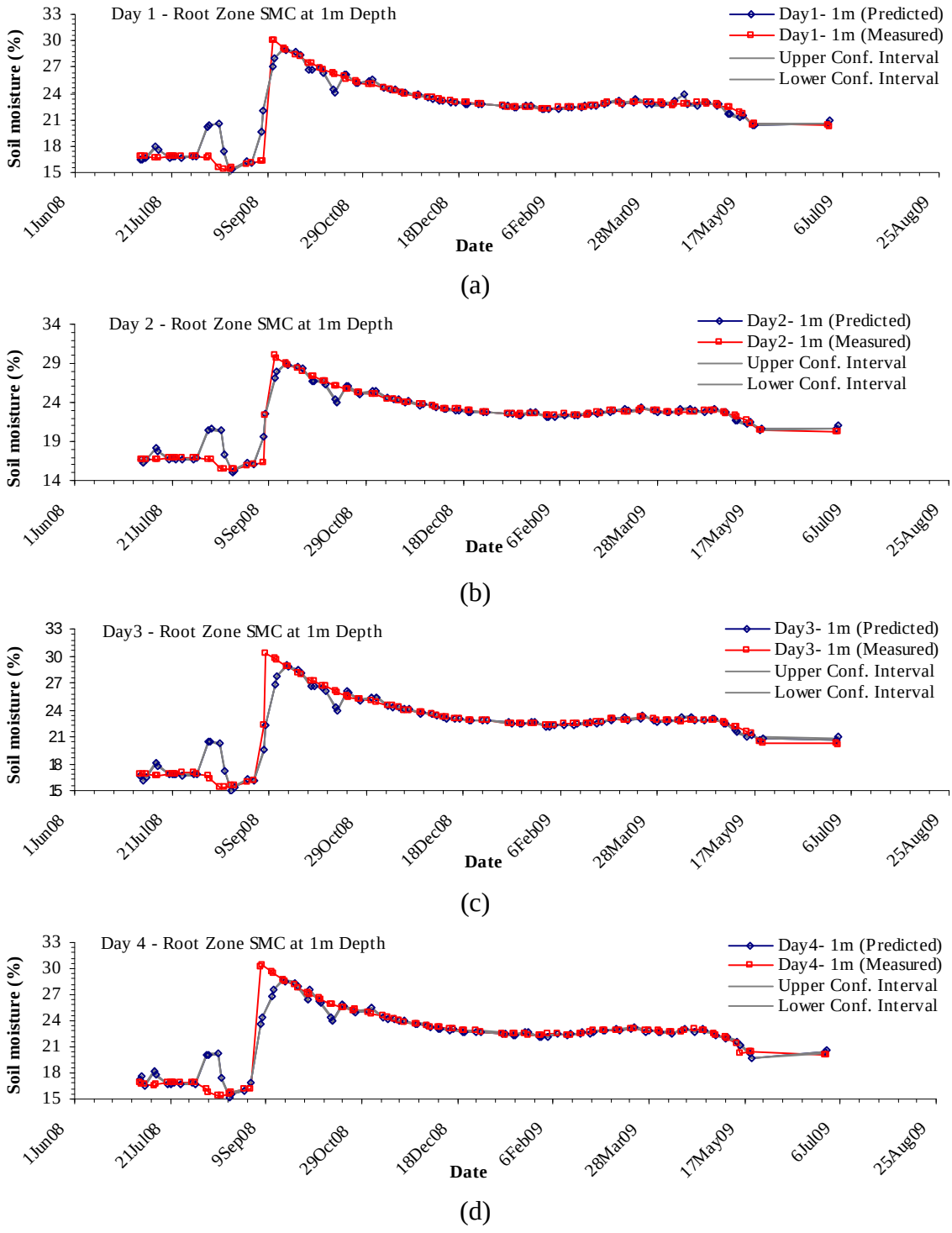
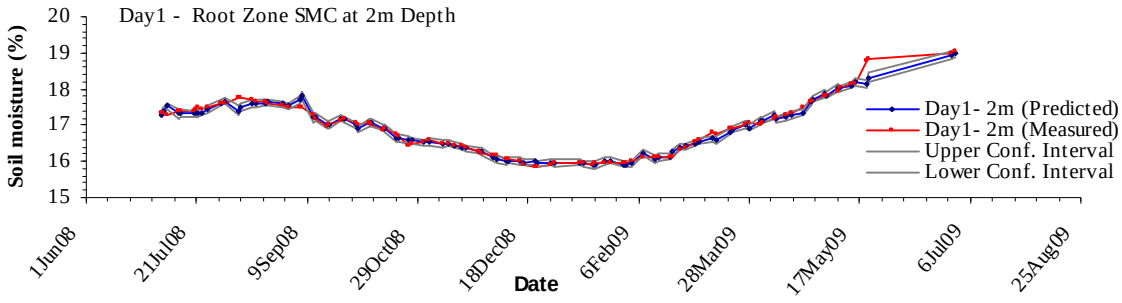
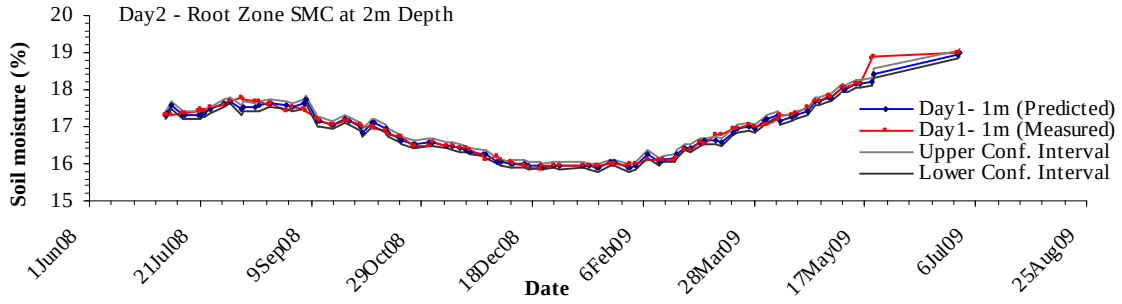


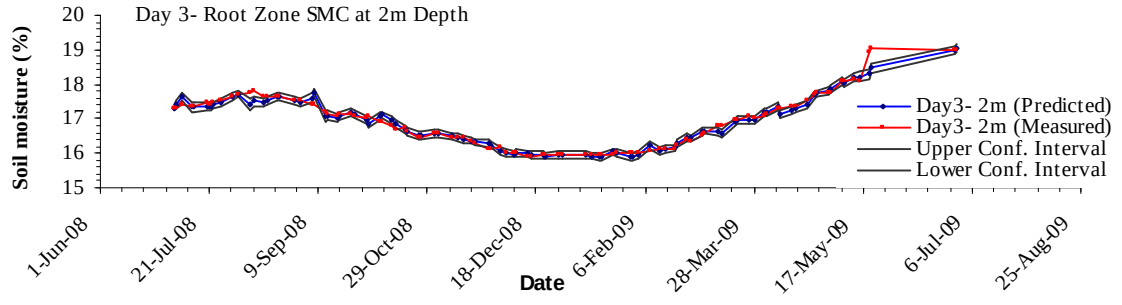
Figure 18. Root zone soil moisture prediction at 1 meter depth on day: (a) d; (b) d+1; (c) d+2; (d) d+3.



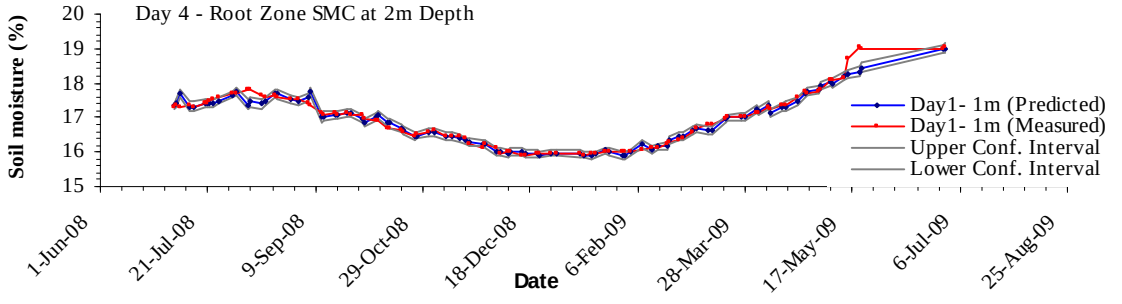
(a)



(b)



(c)



(d)

Figure 19. Root zone soil moisture prediction at 2 meter depth on day: (a) d; (b) d+1; (c) d+2; (d) d+3.

Better results were obtained for soil moisture forecast at 2 m depth in comparison to those obtained at 1 m depth. The variation in moisture at larger depths is smaller than at shallower depths. If there is no deep-rooted vegetation, there would be no extraction of moisture at 2 m depth. In this case, the machine has a simpler problem of forecasting soil moisture at 2 m depth. Figures 19a - 19d show that the soil moisture pattern is followed very accurately by the MVRVM model. The machine was able to capture the spatio-temporal variation of soil moisture at the root zone depths during peak agricultural seasons. Figures 18 and 19 show that the forecast of future soil moisture values has fairly narrow confidence bounds (at 95% confidence interval), which indicates that there is low variance in predictions. The plots (see Figures 18 and 19) show that most of the measured data points lie inside the confidence bounds, indicating that the model is robust.

The full MVRVM model used soil moisture values at four different depths as inputs: 5, 10, 30, and 50 cm for soil moisture prediction at deeper depths. Two additional analyses were done to reveal the effect of using the data at 5 cm and 10 cm (see Table 7) for prediction of root zone soil moisture at 1 m and 2 m, and then using data at 30 cm and 50 cm for the same prediction. The SMC predictions obtained by using input data at 30 cm and 50 cm were closer to the actual soil moisture measurements (see Table 8) and this model produced better results compared to the results generated by the MVRVM model which used data at 5 cm and 10 cm (Table 7). The results for both the analyses were good but not as good as were obtained from the full MVRVM model. However, depending upon the availability of data, the MVRVM model can be applied for soil moisture prediction at larger depths. This article brings into light the capability of the MVRVM

Table 7. MVRVM model results when only surface SMC at a depth of 5 cm and 10 cm are used as inputs (kernel width, r = 4, iterations = 140)

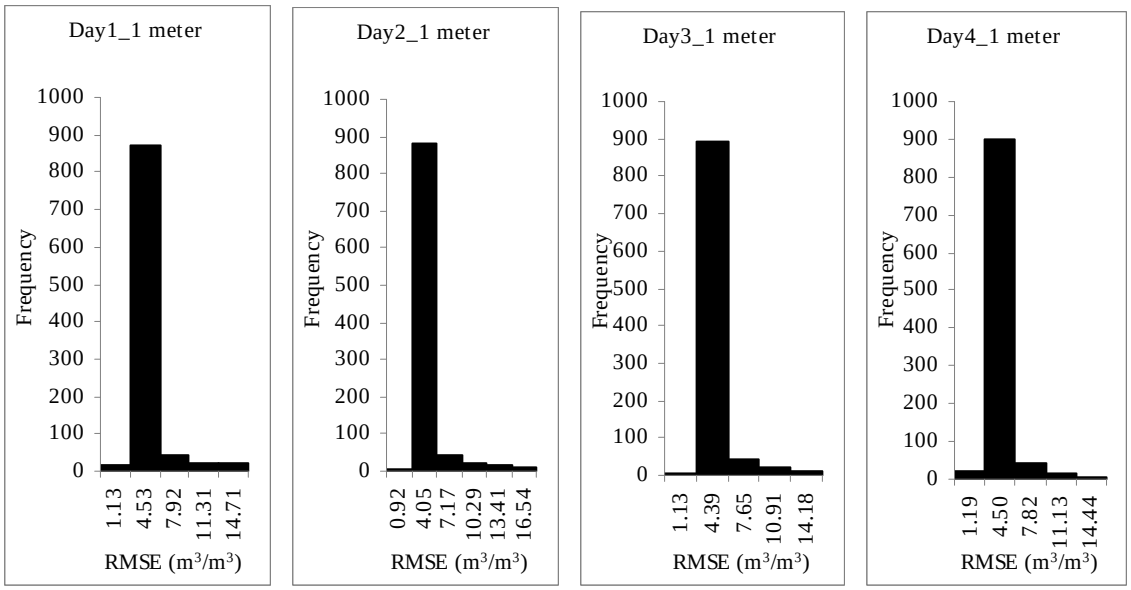
Statistics	Multivariate Relevance Vector Machine Model							
	Day d		Day d+1		Day d+2		Day d+3	
	1 m depth	2 m depth	1 m depth	2 m depth	1 m depth	2 m depth	1 m depth	2 m depth
RMSE %	1.96	0.31	1.83	0.30	1.81	0.32	1.90	0.32
CoE	0.70	0.83	0.72	0.85	0.73	0.83	0.72	0.83
IoA	0.91	0.95	0.92	0.96	0.92	0.95	0.91	0.95
Bias	-0.099	0.119	-0.06	0.081	0.074	0.077	0.07	0.119

Table 8. MVRVM model results when SMC at a depth of 30 cm and 50 cm are used as inputs (kernel width, r =3, iterations = 140)

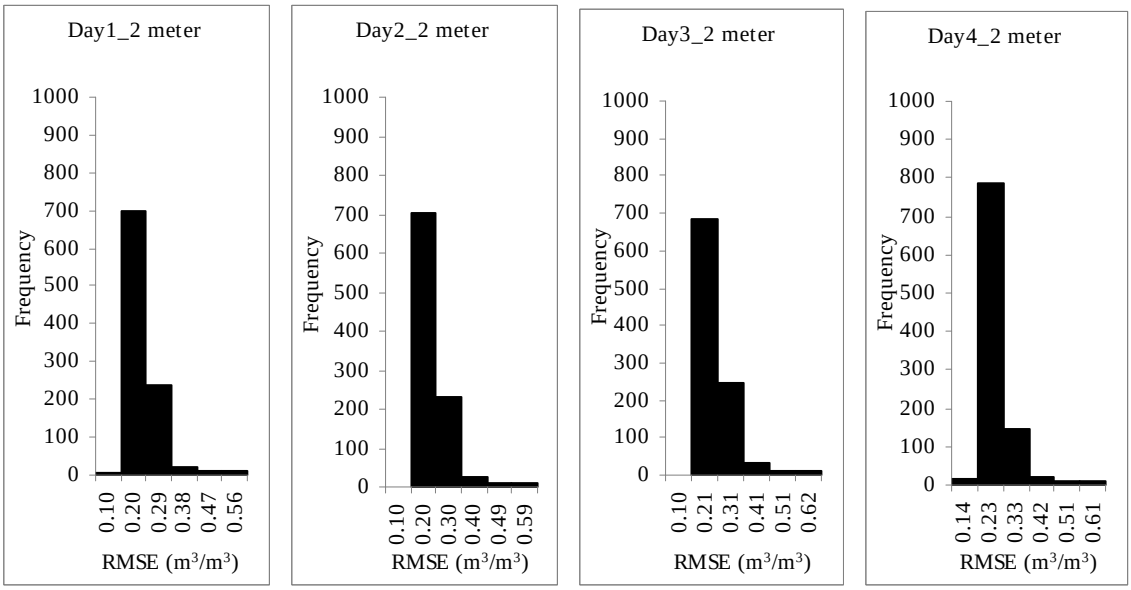
Statistics	Multivariate Relevance Vector Machine Model							
	Day d		Day d+1		Day d+2		Day d+3	
	1 m depth	2 m depth	1 m depth	2 m depth	1 m depth	2 m depth	1 m depth	2 m depth
RMSE %	1.84	0.16	1.78	0.15	1.95	0.17	1.98	0.17
CoE	0.74	0.95	0.74	0.96	0.69	0.95	0.70	0.95
IoA	0.94	0.98	0.94	0.99	0.93	0.98	0.92	0.98
Bias	-0.093	0.044	-0.017	0.033	0.138	0.028	0.120	0.027

model to learn the pattern of soil moisture variation and predict acceptable estimates of soil moisture.

Bootstrapping was performed for the MVRVM model to check for over-fitting and evaluate model generalization capability. Figure 20 shows bootstrap results for RMSE, as estimated from 1000 bootstrap samples. Conforming to the nonparametric approach, no assumption was made about the form of the data, and repeated samples were drawn from the population with replacement. The basic idea is that if the sample is a good approximation of the population, the bootstrap method will provide a good approximation of the sampling distribution of the statistic, in this case, the RMSE. Although beyond the scope of this article, our goal here was to ensure good generalization of the inductive learning algorithm.



(a)



(b)

Figure 20. Bootstrap Analysis Results for Uncertainty in the RMSE of the MVRVM Model with 1000 Bootstrap Samples for the Test Phase: (a) Prediction of Root Zone Soil Moisture at 1 meter Depth for Days d, d+1, d+2, and d+3; (b) Prediction of Root Zone Soil Moisture at 2 meter Depth for Days d, d+1, d+2, and d+3.

A narrow confidence interval indicated that the available training dataset was adequate to determine the machine parameters. From Figures 20a and 20b, one could deduce rough confidence bounds that are more revealing of model performance than single values (Willmott, 1984). In Figure 20, we see that the RMSE values for all the three MVRVM models are centered around one maximum value with highest frequency. Also the MVRVM model shows a fairly narrow confidence bound in all the cases, which implies that the model was robust and its parameter values were well determined.

Summary and Conclusions

This article presents a first attempt to forecast spatial and temporal variation of soil moisture simultaneously using machine learning techniques. This model is based on a sparse Bayesian learning machine approach wherein the machine learns the input output pattern with high accuracy. A MVRVM model is built for developing the prediction functions that forecast soil moisture at 1 m and 2 m depth four days into the future. Three different analyses were done using input data at different depths. The best results were obtained for the full MVRVM model where the input data at 5, 10, 30, and 50 cm depths were used. The results showed excellent performance by the machine for all four days. The forecasted root zone soil moistures were very close to the measured values. It was observed that the SMC predictions at 2 m depth were more accurate than those at 1 m depth due to less variation of moisture at larger depths, which made it easy for the model to learn the input output pattern. The second analysis where the SMC were predicted at 1 m and 2 m depths using data at 5 cm and 10 cm suggested that it is possible to estimate soil moisture in the root zone using surface data by applying the MVRVM model. The

inputs for the third analysis were chosen keeping in mind that the soil moisture sensors are generally operational at these depths, i.e. 30 cm and 50 cm. The MVRVM model performance for this third analysis was also very good, leading to the conclusion that soil moisture conditions at larger depths can be predicted using the MVRVM model if soil moisture data from the sensors are available at 30 cm and 50 cm.

Computation of statistics of interest in conjunction with bootstrapping analyses accomplished a broad operational evaluation of the full MVRVM soil moisture model. These analyses allow us to conclude that the model can predict spatial and temporal variation of soil moisture at large depths with a high degree of accuracy. The model also had good generalization capabilities providing robustness, as demonstrated by the bootstrap analyses results. The MVRVM scheme discussed in this article can be employed to obtain soil moisture estimates from the model in real time and is a potentially useful approach for obtaining short term forecasts in situations where new data can be rapidly exploited as they become available. The results are encouraging and confirm the relevance of the proposed methodology which can benefit soil moisture monitoring and can be extended to other fields of hydrologic science.

CHAPTER IV

ASSIMILATION TECHNIQUE FOR CLASSIFICATION USING SPECTRAL REFLECTANCE DATA AND MULTICLASS RELEVANCE VECTOR MACHINE

Abstract

Classification techniques using ancillary data in addition to spectral data have demonstrated that, in many cases, the proper addition of ancillary data to spectral data can lead to greater class distinctions. We propose a data assimilation technique which fuses spectral reflectance data with other ancillary data to train the state-of-the-art Multiclass Relevance Vector Machine (MCRVM) for building a classification model. The work presents the development and testing of this data assimilation procedure to carry out a supervised classification which is based on statistical learning theory. The model would classify the assimilated data into a predefined number of categories based on a given set of predictors. The study area for this research was the Little Washita watershed, Oklahoma, USA. The data was a part of the Soil Moisture Experiments (SMEX) 2003 conducted at Oklahoma. The paper uses the multispectral radiometer reflectance data, acquired during SMEX03, that has spectral bandpass characteristics similar to selected channels of the Landsat Thematic Mapper and MODIS instruments. Data assimilation technique which fused remotely sensed data (reflectance, vegetation indices) with field measurements of crop physiological characteristics was used to perform supervised classification using the MCRVM model. Once trained, the machine was capable of identifying different classes. The MCRVM routine was trained and tested on two datasets. The first was the vegetation data with six classes (corn, alfalfa, soybeans,

quarry, lake, and bare soil) and seven attributes; the second was the classic Iris flower data with three classes (Setosa, Versicolour, and Virginica) and four attributes. The latter does not use any ancillary data and was just used to test the accuracy of the MCRVM classification routine. The multiclass classification accuracy achieved on the test sets of vegetation and Iris data was 95.2% and 98.7%, respectively. The results showed good performance by the machine with six misclassifications out of 125 instances for the first dataset and one misclassification for the Iris flower dataset. The misclassifications generally occurred where the value of posterior probabilities of class membership for two classes were very close. Predictions showed good agreement with actual data as demonstrated by confusion matrices, receiver operating characteristic (ROC) graphs, and Kappa coefficients. The statistics indicated good model generalization capability.

Introduction

Ancillary data, either in addition to or derived from remotely sensed data, has the potential for increasing classification accuracy and precision (Strahler, Logan, and Bryant, 1978; Hutchinson, 1982; Trotter, 1991; Lawrence and Wright, 2001; Bahadur, 2009). Colstoun, Eric, and Walthall (2006) state that ancillary information may be useful for reducing errors encountered with the use of spectral/temporal data alone. Oftentimes, the classes under investigation are spectrally overlapping as the reflectance recorded by remote sensing satellites for many of these thematic classes is dependent on several extraneous factors like terrain, soil type, moisture content, acquisition time, atmospheric conditions, etc. (Vatsavai et al., 2008) and situation can be improved if one has ancillary data that are correlated with the attributes of interest (Magnussen, McRoberts, and Tomppo, 2009). Hence, the objective of this study was to propose an alternative method

to combine simple predictors with spectral reflectance data to obtain high classification accuracy using MCRVM. This can be beneficial for two main reasons. First, adding ancillary data can produce higher classification accuracy (Jensen, 1996) without dramatically increasing the complexity of the generated model. This was shown by performing sensitivity analysis with subsets of data. Second, the model is not dependent on the scale of data and can provide class estimates having resolution commensurate with remotely sensed data. One of the major goals behind development of this model was to review the suitability of the algorithm for vegetation and crop discrimination.

Accurate and timely information on landcover and the distribution of vegetation on the earth's surface helps us understand the effect of changes in land cover on phenomena as diverse as the atmospheric CO_2 concentrations, loss of prime agricultural lands, terrestrial primary productivity, the hydrologic cycle, and the energy balance at the surface-atmosphere interface (Tucker, Townshend, and Goff, 1985; Anderson et al., 1976). The vast acreages associated with the global agricultural resource base make mapping and monitoring the state of this resource very important (Huang, Davis, and Townshend, 2002). Also, activities to support agriculture, such as crop mapping provide significant information for marketing and trading decisions (Ozdarici and Turker, 2006).

Remote sensing scientists and land cover mapping practitioners have developed new and better techniques for remotely sensed-based landcover mapping (Xu et al., 2004; Le Bris and Boldo, 2008; Tymkow and Borkowski, 2008) and crop discrimination (McNairn, 2002; Doraiswamy, Akhmedov, and Stern, 2007; Mathur and Foody, 2008). Various methods have been applied for classifying remotely sensed data, e.g. nearest neighbor (Barandela and Juarez, 2002), maximum likelihood classifier (MLC) (Ozdarici

and Turker, 2007) artificial neural networks (Hepner et al., 1990; Heermann and Khazenie, 1992; Foody, 1995; Gopal and Woodcock, 1996; Mas, 2004), support vector machines (Huang, Davis, and Townshend, 2002; Melganni and Bruzzone, 2004; Foody and Mathur, 2004; Hermes et al., 1999; Roli and Fumera, 2001; Mercier and Lennon, 2003; Camps-Valls et al., 2003; Munoz-Mari et al., 2008) and, more recently, the relevance vector machine (RVM) (Foody, 2008; Demir and Erturk, 2007). RVMs have a natural extension to the multiclass case and determine hyperparameters in a single run. They ensure a fast and efficient classification process (Wong and Cipolla, 2005) RVMs have been successfully applied in variety of different fields and have been shown to be more suitable for real-time implementation with reduced computational complexity and comparable accuracies (Foody, 2008). Wei et al. (2005) proposed the RVM technique for detection of micro-calcification clusters in digital mammograms. The authors show that though the RVM training time was greater than that of a support vector machine (SVM), the testing time was much less for the RVM while maintaining its best detection accuracy. In Zhang and Malik (2005), an extension of the RVM technique to multiclass problems was derived and applied to digit classification. A two-level hierarchical hybrid SVM-RVM model was used in Silva and Ribeiro (2006) to perform text classification. Recently the RVM multiclassifier has been introduced for classification of remotely sensed data (Foody, 2008) where the data were classified based on reflectance in three spectral wavebands. This paper discussed how the probabilistic nature of the RVM-based classification indicates the class allocation uncertainty on a per-case basis. In Demir and Erturk (2007), RVMs were used for hyperspectral data classification. The authors showed that RVMs produced comparable classification accuracy with a significantly smaller

number of RVs and, therefore, much faster testing time. While RVMs have been successful in producing comparable classification accuracies and producing probabilistic estimates which help understand the class uncertainty on a per case basis (Foody, 2008), failure to incorporate ancillary data into the classification algorithm might fail to fully exploit the range of information available (Lawrence and Wright, 2001). When ancillary data have been incorporated into traditional classification algorithms as logical channels (combining the ancillary data as an additional data layer with the spectral bands), the full range of information available in the ancillary data was used (Strahler, Logan, and Bryant, 1978; Elurnnoh and Shrestha, 2000; Ricchetti, 2000).

In this research, a data assimilation technique was explored using the multi-class relevance vector machine (MCRVM) as a modeling tool for classification of data where ancillary information, relevant to the type of study being carried out, is merged with the reflectance data. The data were assimilated in a non-redundant fashion with LAI, vegetation indices (VIs), and reflectance as inputs. The model produced high classification accuracies and results showed good model performance. This technique as a whole has never been tried before. The model was prepared mainly for crop classification purposes, and inputs that are more sensitive to vegetation differences were used in the training set. Some rigorous accuracy assessment was done to assure that the allocation of classes is not accidental and has been learned by the model. The receiver operating characteristic (ROC) curves were used to check the MCRVM model performance. The model works well with small datasets.

Study area and Data Description

Study area

The study area was the Little Washita watershed, south-west Oklahoma, USA. The data used for the study was a part of the Soil Moisture Experiment (SMEX03) conducted in Oklahoma in 2003. The Vegetation data acquired during the experiments in the Little Washita watershed were used in this paper. The temporal coverage of the data was from 1-17 July 2003 (Jackson and McKee, 2004). Figure 21 shows the Landsat image of the Little Washita watershed area and the different surface types.

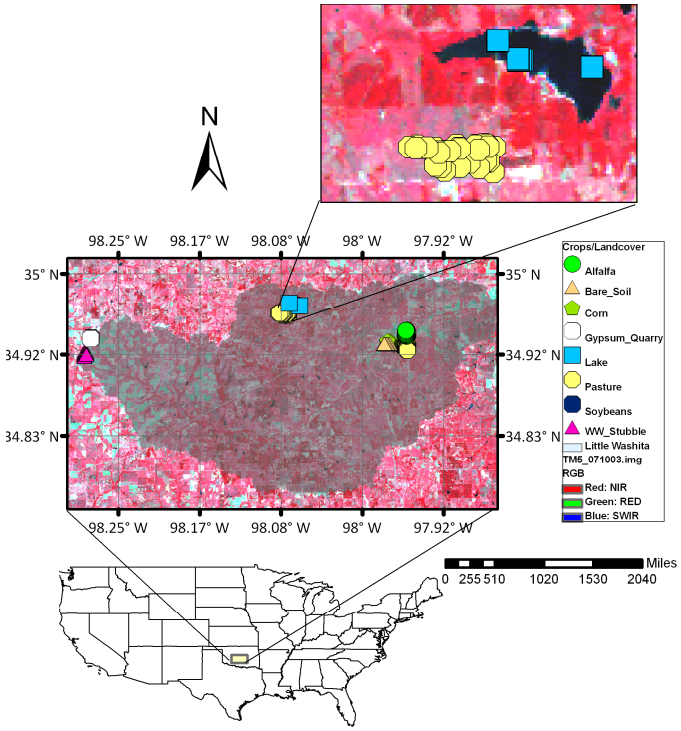


Figure 21. Study area showing the sampling locations of different crop types.

The Vegetation data was downloaded from the NSIDC website. Several Little Washita watershed sites, which represented the dominant types of vegetation, were sampled. Sampling was performed on sites approximately a quarter section (0.8 km by 0.8 km) in size and was concentrated in the Little Washita watershed. Reflectance and Leaf Area Index (LAI) measurements were collected at 9 different sites which included measurements over a lake and a quarry for calibration purposes. The vegetation types were Corn, Alfalfa, Soybeans, Winter-wheat stubble, Pasture and Bare soil. Out of these data acquired over Corn, Alfalfa, Soybeans, Bare soil, Quarry and Lake were used for this paper. The attributes used for training the MCRVM model were LAI (m^2/m^2), multispectral radiometer reflectance (%) and VIs. The following sections provide a brief description of the datasets.

Vegetation data

Multi-spectral radiometer reflectance measurements. During the SMEX 2003 experiments, the investigators used MSR-16R multispectral radiometers manufactured by CropScan to measure reflectance (Jackson and McKee, 2004). Table 9 shows the sampling scheme for the data.

Table 9. Sampling scheme of vegetation data

No. of Samples	LW03 Pasture	LW12 Pasture	LW20 stubble	LW31 Corn	LW32 Bare-Soil	LW33 Alfalfa	SOY1 Soybean	Lake	Quarry
7/1/2003	14								
7/2/2003		14		14	14			4	4
7/3/2003			14			14	14		
7/9/2003	14	14				14	14		
7/10/2003				14	14			5	5
7/16/2003		14		14	14	14			
7/17/2003	14						14		
7/18/2003								5	5

The wavelengths measured were: 485, 560, 650, 660, 830, 850, 1240, 1640, and 1650 nm bands. These bands provide data for selected channels of the Landsat Thematic Mapper and MODIS instruments. Channels were chosen to provide a variety of vegetation water content indices (Jackson and McKee, 2004). The average % reflectance measurements in wavebands 485, 560, 660 and 1650 nm were used directly as inputs.

Leaf area index (LAI) measurements. LAI is defined as the ratio of total upper leaf surface of vegetation divided by the surface area of the land on which the vegetation grows. During the SMEX 2003 experiments, LAI was measured using LI-COR LAI-2000 plant canopy analyzers using an indirect contact method based on light transmittance through the canopy (Jackson and McKee, 2004). LAI is a dimensionless value (m^2/m^2).

Calculation of VIs. The soil adjusted vegetation index (SAVI) and normalized difference water index (NDWI) were used as inputs. The MSR-16R multi-spectral radiometer reflectance data recorded in the bands 650, 830, 850, and 1240 nm were used to calculate the VIs The following equations were used.

$$SAVI = (R_{NIR} - R_{RED})\,(1 + L)/\,(R_{NIR} + R_{RED} + L) \tag{38}$$

$$NDWI = R_{NIR} - R_{SWIR} / R_{NIR} + R_{SWIR} \tag{39}$$

where, R_{NIR}, $R_{RED,}$ R_{SWIR} are the apparent reflectance values in the near-infrared (~0.8 µm), red (~0.6 µm) and short wave infrared (~1.2–25 µm) wavebands, respectively. L is a calibration factor (Huete, 1988). SAVI and NDWI were dimensionless values.

Iris data

The second dataset was the Iris flower data. This is perhaps the best known dataset found in the pattern recognition literature (Güngör and Unler, 2007). Figure 22 shows images of the three types of Iris flower.

The dataset consists of 3 classes with 50 instances each, where each class refers to a type of Iris plant, Setosa, Versicolour, and Virginica. The dataset has four attributes: sepal length, sepal width, petal length, and petal width in cm.

Methodology

In this study, given a set of assimilated data with labeled instances which are selected from a finite dataset, an inductive procedure was built to deduce an inferring function i.e the MCRVM model, which was able to map unseen instances to their appropriate classes. The section describes how the multi-classifier was built, trained and tested with the vegetation and the Iris data. Further the section describes the accuracy assessment methods used for checking the robustness, convergence, speed, and accuracy of the model from the performance viewpoint.

(a) (b) (c)

Figure 22. Iris flower of type (a) Setosa; (b) Versicolour; (c) Virginica

Multi-class Relevance Vector Machine (MCRVM)

The RVM was originally introduced by Tipping (2001). Thayananthan et al. (2006) proposed an extension of the sparse Bayesian model developed by Tipping to handle multiple outputs. Thayananthan's MCRVM code is an open source code which extends Tipping's binary Relevance Vector Machine classification scheme (Tipping, 2001) to a MCRVM classification algorithm. This code was used as a base to build the multi-classifier which was particular to this application.

General background of RVM. "Sparse Bayesian Learning" is used to describe the application of Bayesian automatic relevance determination (ARD) concepts to models that are linear in their parameters. The motivation behind the approach is that one can infer a regression or classification model that is both accurate and sparse in that it makes its predictions using only a small number of relevant basis functions that are automatically selected from a potentially large initial set. A special case of this concept is the RVM which is applied to the linear kernel models.

The data set is in the form of input-output pairs, $\{\mathbf{x}_n, y_n\}_{n=1}^{N}$. The major goal is to learn a model of dependency of the targets on the inputs with the objective of making accurate predictions for previously unseen values of **x** (Tipping, 2001). This model is defined as some function $y(\mathbf{x})$ whose parameters are found as:

$$y(\mathbf{x};\mathbf{w}) = \sum_{i=1}^{M} \boldsymbol{w}_i \, \varphi_i(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) \tag{40}$$

where the output $y(\mathbf{x};\mathbf{w})$ is a linearly weighted sum of M, generally nonlinear and fixed basis functions $\boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}),\ldots \varphi_M(\mathbf{x}))^T$ and $\mathbf{w} = (w_1, w_2, \ldots, w_M)^T$, called weights,

are adjustable parameters. Equation (40) can result in a number of different models, of which RVMs are a special case.

This procedure is highly perceptive with a Bayesian probabilistic framework that helps in extracting predictors that are very sparse, with few non-zero **w** parameters. Only those basis functions that are necessary for making accurate predictions are retained.

Baye's rule states that the posterior probability of **w** is obtained by combining the likelihood and prior as:

$$p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha},\sigma^2) = p(\mathbf{t}|\mathbf{w},\sigma^2)\,p(\mathbf{w}|\boldsymbol{\alpha}) / p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) \quad (41)$$

where σ^2 is the error variance, $p(t|w,\sigma^2)$ is the likelihood of target **t**, $p(\mathbf{w}|\boldsymbol{\alpha})$ is the prior, and $p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2)$ is the evidence.

RVM classification follows an identical framework as regression (see Chapter II – Relevance vector machines). To account for the change in target quantities (classes), the logistic sigmoid link function $\sigma(y) = 1/(1+e^{-y})$ is applied to $y(\mathbf{x})$ and, the Bernoulli distribution is adopted for $p(\mathbf{t}|\mathbf{w})$. The likelihood can be written as:

$$p(\mathbf{t}|\boldsymbol{w}) = \prod_{n=1}^{N} \sigma\{\boldsymbol{y}(\mathbf{x}_n;\mathbf{w})\}^{t_n} [1 - \sigma\{\boldsymbol{y}(\mathbf{x}_n;\mathbf{w})\}]^{1-t_n} \quad (42)$$

where t_n is the target class $\in \{1, 2, 3, 4, 5, 6\}$ in this paper. In (Zhang and Malik, 2005) a true multiclass likelihood was specified. It was obtained by generalizing (42) to multinomial form given by,

$$p(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N}\prod_{k=1}^{K} \sigma\{y_k; y_1, y_2, \ldots y_k\}^{t_{nk}} \quad (43)$$

where the predictor y_k of each class was coupled with the multinomial logit function given by,

$$\sigma(y_k; y_1, y_2, ...y_k) = \frac{e^{y_k}}{e^{y_1} + ... + e^{y_k}} \tag{44}$$

For obtaining probabilistic outputs, a sigmoid link function is applied to the output $y(\mathbf{x})$, $f(y)=1/(1+e^{-y})$. A zero mean Gaussian prior distribution is applied over $\mathbf{w}$ and is given by,

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{n=1}^{N} \sqrt{\frac{\alpha_n}{2\pi}} \exp\left(\frac{\alpha_n \boldsymbol{w}_n^2}{2}\right) \tag{45}$$

Here each of the N independent hyperparameters, $\boldsymbol{\alpha} = (\alpha_0,\alpha_1,\ldots,\alpha_N)^T$, individually control the strength of the prior over it's associated weight and is eventually responsible for the sparsity of the model (Tipping, 2001).

The closed-form expression for the weight posterior $p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha})$ and evidence of hyperparameters $p(\mathbf{t}|\boldsymbol{\alpha})$ cannot be obtained since the weights in (43) cannot be integrated out. Hence a Laplacian approximation is used. Since $p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha}) \propto p(\boldsymbol{t}|\boldsymbol{w})\, p(\mathbf{w}|\boldsymbol{\alpha})$, with a fixed given $\boldsymbol{\alpha}$, the maximum a posteriori estimate (MAP) of weights can be obtained by maximizing $\log(p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha}))$ or by minimizing the following cost function (Camps-Valls, Marsheva, and Zhou, 2007),

$$\log(p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha},\sigma^2)) = \sum_{n=1}^{N} \left(\frac{\alpha_n \boldsymbol{w}_n^2}{2} - \boldsymbol{t}_n \log \boldsymbol{y}_n + (1-\boldsymbol{t}_n)\log(1-\boldsymbol{y}_n)\right) \tag{46}$$

The Hessian of $\log(p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha}))$ is given by,

$$\mathbf{H} = \nabla^2(\log(p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha}))) = \boldsymbol{\Phi}^T \boldsymbol{B}\boldsymbol{\Phi} + \boldsymbol{A} \tag{47}$$

where matrix $\boldsymbol{\Phi}$ is the Nx$(N+1)$ 'design' matrix with $\boldsymbol{\varphi}_{nm}=k(\mathbf{x}_n,\mathbf{x}_{m-1})$. $k(\mathbf{x}_n,\mathbf{x}_{m-1})$ is the Gaussian kernel and has the form: $k(\mathbf{x}_n,\mathbf{x}_{m-1}) = \exp(-r^{-2}\|x_n - x_{m-1}\|^2)$, where r is the kernel width. $\boldsymbol{A} = \text{diag}\{\alpha_1,\ldots,\alpha_n\}$, and $\mathbf{B} = \text{diag}(\beta_1,\beta_2,\ldots,\beta_N)$ is the diagonal matrix with $\beta_n =$

$\sigma\{y(\mathbf{x}_n)\}[1-\sigma\{y(\mathbf{x}_n)\}]$. The hyperparameters $\boldsymbol{\alpha}$ are iteratively updated using the covariance Σ and mean $\boldsymbol{\mu}_{MP}$ of the Gaussian approximation.

The covariance Σ is given by the inverse of the Hessian (47),

$$\boldsymbol{\Sigma} = (\mathbf{H})^{-1} = (\boldsymbol{\Phi}^T\mathbf{B}\boldsymbol{\Phi} + \boldsymbol{A})^{-1} \quad (48)$$

and the mean is given by,

$$\boldsymbol{\mu}_{MP} = \boldsymbol{\Sigma}\boldsymbol{\Phi}^T\mathbf{B}\hat{\mathrm{t}} \quad (49)$$

$$\hat{\mathbf{t}} = \boldsymbol{\Phi}\boldsymbol{\mu}_{MP} + \mathbf{B}^{-1}(\mathbf{t} - \mathbf{y}) \quad (50)$$

The following equation is used for updating the hyperparameters (Tipping, 2001):

$$\alpha_i^{new} = \frac{1 - \alpha_i \Sigma_{ii}}{\mu_i^2} \quad (51)$$

where μ_i denotes the i^{th} posterior mean weight from (10), Σ_{ii} is the i^{th} diagonal element of the posterior weight covariance (48) , and the quantity $1-\alpha_i\Sigma_{ii}$ is a measure of the degree to which the associated parameter w_i is determined by the data (Khalil, Mckee, and Kaluarachchi, 2005). During the re-estimation process the α_i tend to infinity making $p(\boldsymbol{w}_i|\mathbf{t},\boldsymbol{\alpha})$ highly peaked at zero. This makes the associated weights zero and hence the associated basis functions are discarded making the machine sparse.

Data assimilation and training and testing of the MCRVM model

Two different datasets were used for training and testing the model. The first dataset was the vegetation data from SMEX 2003 which had seven inputs which were LAI, SAVI, NDWI and reflectance at 485, 560, 660 and 1650 nm and six output classes which were Corn, Alfalfa, Soybeans, Quarry, Lake, and Bare soil. The second was the

Iris flower dataset with 4 attributes (Sepal length, sepal width, petal length and petal width) and 3 classes (Setosa, Versicolour, and Virginica).

The fist step in developing the classification scheme was data cleaning where the missing and the inconsistent data was removed. We know that use of ancillary data in classification must rely on in-depth knowledge of the target to select the attribute that best characterizes it (Ricchetti, 2000). The aim was to extract the structural features from the data which would be used by the classifier to assemble a robust predictor and a generalized multiclass learning machine. The purpose was to build a model for vegetation/crop discrimination. Hence, several runs were performed with different combinations of reflectance values with VIs and LAI. It was observed that reflectance at 485, 560, 660, and 1650 nm along with SAVI, NDWI and LAI which produced the best results and enhanced class separability. The VIs were calculated using reflectance in bands 650, 830, 850, and 1240 nm and the bands that were already used for the calculation of VIs were not used in the input training matrix.

After the data were assimilated, a small representative set of points were selected from the vegetation dataset through stratified random sampling for training the MCRVM model. The vegetation data training set comprised of 70 instances and an independent testing set consisting of 125 instances. The trained machine was then used to classify the test data.

After the test results were obtained, which were the posterior probabilities of the class memberships, the ultimate class was selected based on the maximum Bayesian posterior probability rule applied to these posterior probabilities. Figure 23 summarizes the methodology for the multi-class classification RVM model.

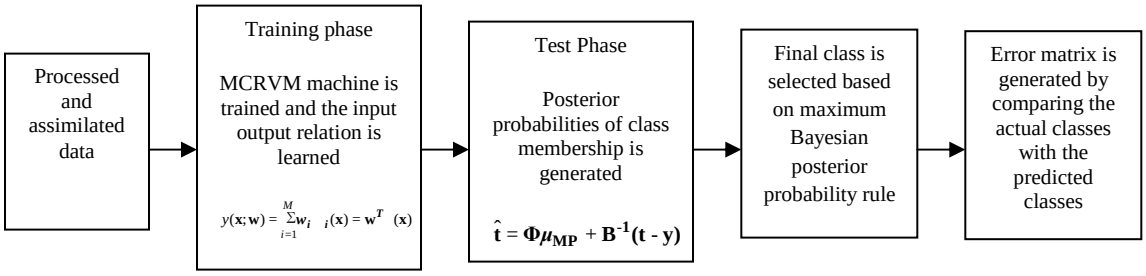


Figure 23. Diagram of MCRVM data classification process.

Sensitivity analysis was done wherein LAI was removed and the model was run for the remaining six inputs. Another analysis was done with just the reflectance data to observe the effect of data assimilation. A rigorous accuracy assessment was done where the ROC curves, confusion matrix, and Cohen's Kappa coefficient (see Equation 53) were calculated for each dataset. The classification accuracy was expressed as the percentage of the testing cases correctly classified.

The Iris dataset was used for testing the classifier generalization capability and accuracy. The data consists of 150 instances. It was divided equally into training and testing sets of 75 instances each by stratified sampling. The MCRVM was trained and tested with each of these sets.

Accuracy assessment

A meticulous assessment of classification accuracy accomplishes a broad operational evaluation of the model. There are many classification accuracy measures reported in the literature, the most extensively used ones are derived from the error or confusion matrix (Congalton, 1991; Foody, 2002). Recent years have seen an increase in the use of ROC curves in machine learning and data mining. In addition to being a useful performance graphing method, they have properties that make them especially useful for

domains with skewed class distributions and unequal classification error costs (Fawcett, 2004). Cohen's Kappa coefficient is considered to be a robust measurement of classification accuracy yet is widely discredited. Though it has also been stated in the literature that it takes into account agreement by chance, and in some circumstances it should be considered as a standard measure of classification accuracy (Smits, Dellepiane, and Schowengerdt, 1999). The following section described these measures of accuracy.

ROC curves. The receiver operator characteristic (ROC) curves analyze the hit rates/false alarm rates (Hayat, 2007) of diagnostic decision-making. In a two-class problem, the area under the ROC curve (AUC) is a single scalar value, but a multiclass problem introduces the problem of combining the multiple pair wise discriminability (Fawcett, 2006). The approach used in this paper is taken from the discussion given in (Fawcett, 2006), following the approach used in (Provost and Domingos, 2001). The multiclass AUCs are calculated by producing an ROC curve for each class, measuring the area under the curve, and then adding up the AUCs weighted by the reference class's prevalence in the data. It is defined by,

$$\mathrm{AUC}_{\mathrm{total}} = \sum_{c_i \in C} \mathrm{AUC}(c_i).p(c_i) \tag{52}$$

where AUC(c_i) is the area under the class reference ROC curve for c_i.

Confusion matrix. A confusion matrix is a tool used in supervised learning to judge the accuracy of the classifier. This method has an advantage of producing single accuracy indexes which can be used for further evaluation and comparison (Samaniego, Bardossy, and Schulz, 2008). Tables 10 and 11 show the error matrices for the vegetation and iris data and the user's and producer's accuracy show the model performance for each class.

Kappa coefficient. The confusion matrix obtained through the MCRVM model was analyzed using the Kappa coefficient. The kappa coefficient (K) measures pairwise agreement between the classified data and real data, correcting for expected chance agreement:

$$K = \frac{N\sum_{i=1}^{n} x_{ii} - \sum_{i=1}^{n}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{n}(x_{i+} \times x_{+i})} \tag{53}$$

where n is the number of classes, x_{ii} is the number of observations on the diagonal of the confusion matrix corresponding to row *i* and column *i*, x_{i+} and x_{+i} are the marginal totals of row *i* and column *i*, respectively, and N is the total number of instances. Kappa is always less than or equal to 1. A value of 1 implies perfect agreement and values less than 1 imply less than perfect agreement.

Results and Discussion

The final classes predicted by the MCRVM model was compared with the original classes and of the 125 cases in the testing set of vegetation data, only 6 were misclassified. For the Iris data, out of 70 cases in the testing set, only one 1 was misclassified. The overall classification accuracy obtained for the vegetation data was 95.2% (Table 10) and Cohen's kappa coefficient was found to be 0.94 (Table 11). The kappa confidence interval was 0.867 to 0.974 which reflected the strength of the inter-rater agreement and showed that the observed agreement was not accidental.

Table 10. Confusion matrix along with the users accuracy (UA%) and producers accuracy (PA%) yielded by the MCRVM classifier in the test set (vegetation data)

		Classification data							
		Bare-soil	Corn	Quarry	Lake	Alfalfa	Soybean	Row Total	PA (%)
Reference Data	Bare-soil	27	0	0	0	1	0	28	100
	Corn	0	24	0	0	0	0	24	100
	Quarry	0	0	9	0	0	0	9	100
	Lake	0	0	0	9	0	0	9	100
	Alfalfa	0	2	0	0	23	0	25	92.0
	Soybean	0	1	0	0	2	27	30	90.0
	Column Total	27	27	9	9	26	27	125	97.0
	UA (%)	100	88.9	100	100	88.5	100	96.2	95.2

The average user's and producer's accuracy for the vegetation data was 96.23% and 97%, respectively. Out of six misclassification for the vegetation data, four were confident misallocations and for the rest 2, the posterior probabilities of class membership were very close. Use of LAI helped the algorithm to classify other data type such as water and Quarry as these had a 0 LAI value.

The MCRVM model was applied to the Iris data set (Fisher, 1936), which is considered as a standard benchmark in the pattern recognition literature. The accuracy achieved was 98.67% (Table 11), which is at par with the maximum accuracy achieved with Iris data (Fung and Managsarian, 2005). The average User's and Producer's accuracy was 98.67% and 98.72%, respectively. The Kappa coefficient was 0.98 (Table 12).

Once the MCRVM model was trained, the model took very less time to generate the posterior probabilities of class membership. Table 12 shows the training and testing times for both the datasets.

Table 11. Confusion matrix along with the users accuracy (UA%) and producers accuracy (PA%) yielded by the MCRVM classifier in the test set (IRIS data)

		Classification data				
		Setosa	Versicolor	Virginica	Row Total	PA (%)
Referenc.	Setosa	25	0	0	25	100
	Versicolor	0	24	0	24	100
	Virginica	0	1	25	26	96.2
	Column Total	25	25	25	75	98.72
	UA (%)	100	96.0	100	98.67	98.67

Table 12. MCRVM classifier robustness, speed, and accuracy

Dataset	No. of training/ test samples	Accuracy (%)	Kappa coefficient	Training time (sec)	Testing time (sec)
SMEX Veg data	70/125	95.2	0.94	31	0.02
Iris data	75/75	98.7	0.98	9	0.014

The inferred classifiers were sparse and used only an average of 11 RVs out of 70 training points for the SMEX vegetation dataset, and 17 RVs out of 75 training points for the Iris data. The probable reason for the larger number of RVs for the Iris data might be that one class (Setosa) is linearly separable from the other two, but the latter are not linearly separable from each other.

The multiclass AUCs were calculated using method used by Provost and Domingos (2001). The advantage of this AUC formulation is that AUC_{total} is calculated directly from class reference ROC curves which can be generated and visualized easily. The disadvantage is that class reference ROC is sensitive to class distributions and error costs, so this AUC_{total} is as well (Fawcett, 2006). The multiclass AUC_{total} for the SMEX vegetation data was 0.995, and for the Iris data it was 0.994. Figure 24 shows the true positive (TP) rate versus false positive (FP) rate for six classes of the SMEX vegetation data. The ROC curves for classes 3 and 4 are perfect. The ROC curves for classes 1, 2, 5

and 6 show that the model performance is good as the curves lie towards the northwest corner of the ROC space. A similar conclusion can be drawn for the Iris data (Figure 25), which shows that all three ROC curves lie towards northwest corner of the ROC space.

Sensitivity analysis (see Table 13) was done to test the performance of the machine without the LAI input and then without including LAI and VI. Results show that addition of LAI to the dataset increased the accuracy by almost 1%. LAI measurement is often a part of large experimental project like SMEX.

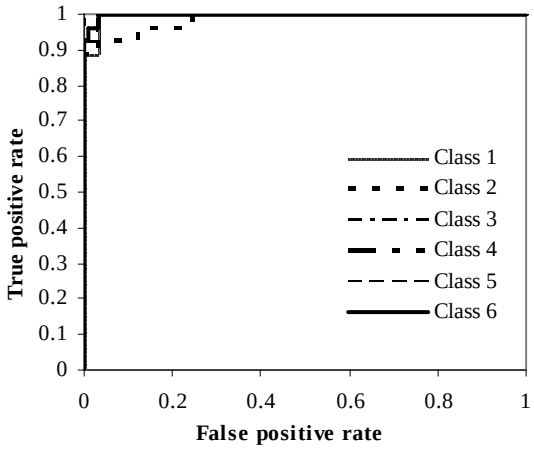


Figure 24. ROC curves for 6 classes of vegetation data.

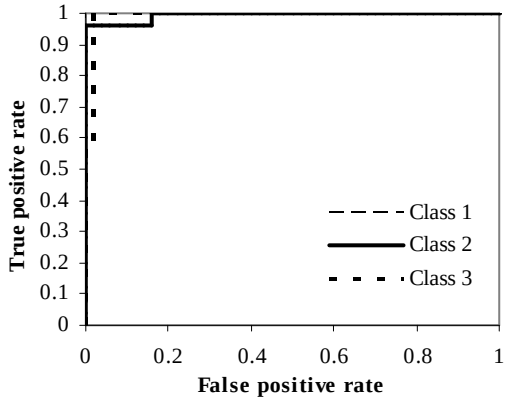


Figure 25. ROC curves for three classes of Iris data.

Table 13. MCRVM classifier accuracy obtained with different subsets of data

Inputs	No. of training/ test samples	Overall classification accuracy (%)	Kernel	Kernel Width	Training time (sec)
LAI, VI, Reflectance	70/125	95.2	Gaussian	45	31
VI, Reflectance	70/125	94.4	Gaussian	3.2	35
Reflectance	70/125	92	Gaussian	8	0.13

If the data are readily available, it can be used in conjunction with other inputs which might help in improving the accuracy of the learning machine. The MCRVM classifier produced an accuracy of 92% when only the reflectance data were used which was 3.2% less than the case where the data assimilation technique was used.

The use of Gaussian kernel resulted in the maximum accuracy of the MCRVM classifier with a kernel width of 45. Table 14 shows the results obtained for different kernel functions. It was observed that the Laplacian and Cauchy kernels produced the second best result with 91.2% accuracy.

The two vegetation indices used in the input data set were SAVI and NDWI, both of which are derived using the reflectance in the near-infrared band as one of the variables. This might result in some cross-correlation between the input variables. However, the RVM produces a maximum likelihood covariance matrix that implicitly involves perfectly uncorrelated sources; correlation among the actual sources has absolutely no effect on the RVM global minimum. This model covariance is then used in place of the measured one to improve performance when data is limited and/or when sources are correlated (Wipf and Nagarajan, 2007).

Table 14. MCRVM classifier accuracy obtained with different Kernel functions in the test set of Vegetation data

Kernel type	Kernel width	Accuracy (%)
Gaussian	45	95.2
Laplace	5	91.2
Spline	31	88.8
Cubic (cube of distance)	40	45.6
Cauchy	9	91.2
r (distance)	19	87.0
tps (thin plate spline)	1	76.0

Conclusions

We have shown that data assimilation technique using the MCRVM model can be used in the crop classification context to yield very accurate or meaningful results. The purpose of this study was to evaluate the effectiveness of using ancillary data along with spectral reflectance data to improve the interpretability of class prediction, and the automatic classification of spectral data using MCRVM. This paper presented a new and efficient technique for land cover classification. It introduced the use of data assimilation for classification and demonstrated that the classification accuracy was significantly improved from 92% to 95.2% by training the MCRVM model with assimilated inputs that affect the data being classified. Exhaustive accuracy assessment of the technique suggested that the MCRVM model is robust as demonstrated by its high classification accuracy and small number of RVs. This compact model form required much less testing time than training time and avoided the need to set additional regularization parameters. This allowed us to conclude that the MCRVM offer a suitable paradigm for the inclusion of ancillary information in the classification process as was also evident from the high classification accuracies generated by the model. The probabilistic nature of the MCRVM

results helped to evaluate the performance of the model on a per case basis and the six misclassifications in the case of vegetation data could be explained. Supervised classification requires analyst-specified classification data and it was observed that the performance of the model heavily depended on the accuracy of the data and also on the size of training and test sets. Kernel width, type of kernel, and iterations were the parameters that controlled model performance. The existing training algorithm worked well with the vegetation dataset used in this research. This should draw attention toward the use of data assimilation techniques with this sophisticated learning machine tool to improve classification accuracies in the future. This technique may uncover important patterns hidden in the data which can contribute greatly to knowledge bases.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Chapters II to IV present the body of the work and the main scientific results of the dissertation. Here, I summarize and emphasize the important conclusions and recommend avenues for future research.

Summary

This study has investigated the usefulness of remote sensing, data assimilation and statistical learning theory for solving agricultural water resources management problems. The outcome of this dissertation provided a theoretically sound approach for soil moisture estimation in the topsoil and deeper layers in the soil profile. Furthermore it laid the foundation for a new breed of techniques which use data assimilation along with learning machines for landcover classification. This dissertation is comprised of three main components:

Task I: Fusion of remotely sensed data for soil moisture estimation using relevance vector and support vector machines

In this application, a new technique for estimation of soil moisture content (SMC) is introduced. It uses remotely sensed inputs as a part of a unified database that consists of meteorological data, field measurements, and crop physiological factors. The methodology is divided into three models. The first model uses remotely sensed data and other ancillary data to retrieve soil moisture in the 0-6 cm layer of topsoil. The second model estimates soil moisture at 30 cm depth by using field measurements of SMC in the

top 0-6 cm layer. The third is a two-step model which combines the previous two models. This model estimates soil moisture at 30 cm depth by using the surface soil moisture estimates produced in the first step. Hence, the third model simulates a case where soil moisture at 30 cm depth is estimated at a large scale using the remotely sensed data, meteorological inputs and crop physiological properties. The results for the first model show that it is possible to get good estimates of the surface SMC in the top 0-6cm layer by using RVMs and SVMs. Further, using these estimates of the topsoil soil moisture, it is possible to estimate soil moisture up to a depth of 30 cm. The RVM demonstrates excellent performance. Results indicate that the RVMs perform better than the SVMs in all the test cases for the three models and hence demonstrate a better capability for capturing the underlying phenomena, showing good potential for SMC estimation.

This methodology is simple as it uses data that are easily assimilable. The output of this study provides an essential input for soil water balance calculations and updating of the soil moisture model in an operational setting.

Task II: Spatio-temporal prediction of root zone soil moisture using multivariate relevance vector machines (MVRVM)

A root zone soil moisture profile estimation algorithm has been developed for estimating the soil moisture dynamics at a point scale. This has involved the development of a computationally efficient soil moisture profile forecasting model and an application of the MVRVM by using surface soil moisture and meteorological information as inputs. This study presents a first attempt to forecast spatial and temporal variation of soil moisture simultaneously using machine learning techniques. The sparse Bayesian MVRVM model learns the input-output pattern with high accuracy and infers the

prediction functions that forecast soil moisture for up to two meters depth for several days in the future. The forecasted root zone soil moisture values are very close to the measured values, which allows us to conclude that the MVRVM model can predict spatio-temporal variation of soil moisture at large depths with a high degree of accuracy.

The MVRVM scheme discussed in this paper can be employed to obtain soil moisture estimates from the model in real time and is a potentially useful approach for obtaining short-term forecasts in situations where new data can be rapidly exploited as they become available. The results are encouraging and confirm the relevance of the proposed methodology which can benefit soil moisture monitoring and can be extended to other fields of hydrologic science.

Task III: Assimilation technique for classification using spectral reflectance data and multiclass relevance vector machine (MCRVM)

This application shows that data assimilation technique using the MCRVM model can be used in the crop classification context to yield very accurate and meaningful results. A new and efficient technique for land cover classification is introduced. The purpose was to evaluate the effectiveness of using ancillary data along with spectral reflectance data in improving the interpretability of class prediction, and the automatic classification of spectral data using MCRVM. This compact model form required much less testing time than training time and avoided the need to set additional regularization parameters. This allowed us to conclude that the MCRVM offer a suitable paradigm for the inclusion of ancillary information in the classification process as was also evident from the high classification accuracies generated by the sparse model. The probabilistic

nature of the MCRVM results helped to evaluate the performance of the model on a per case basis. This crop mapping scheme can provide significant information for marketing and trading decisions. Also, the vast acreages associated with the global agricultural resource base make crop mapping and monitoring very important.

Conclusions

An integrated, effective agricultural water management approach requires the users to get involved in the decision-making and management process. The goal is to build an appropriate knowledge base and strengthen analytical capacity in the region to better plan and manage water resources and service delivery. This is often implemented with the help of canal and reservoir operators. There is often a lag between the order and delivery of water to the farm/field. Knowledge about the moisture status of the field helps the decision maker to make the right choices leading to more efficient handling of the available water and less wastage.

With these goals in mind, this dissertation attempts to develop procedures which give a rough idea to farmers/irrigators about the moisture status of their fields and also about the productivity. This information could help in the overall improvement of the water management practices. The framework devised in this dissertation attempts to provide tools to support irrigation system operational decisions. The three components of this research dealt with timely information about the soil moisture status at the farm level with the help of remotely sensed data, root zone soil moisture assessment, and crop identification. The first provides surface soil moisture information on a large scale. It could be implemented on a real-time basis depending on the availability of data, and

could be used as inputs to the second and the third procedures. The second is to get accurate information about soil moisture at large depths using surface information and ancillary weather data. The method introduced a potential new tool for accurate estimation of soil moisture at larger depths by using surface information. This procedure provides the essential input for updating of the soil water balance calculations in an operational setting. These findings can be used in the third application which identifies crop type and vegetation cover. The classification scheme may uncover important data patterns contributing greatly to knowledge bases, and to scientific and medical research. This application, apart from providing information about crop yields and acreages can also be used for identifying weeds and other invasive species in agricultural fields.

These three components were tested and corroborated separately but they are interconnected and can be the building blocks of soil water balance calculation models. Though the dissertation is one of the most current efforts in advancing the use of learning machine tools (RVM, SVM, MVRVM, MCRVM) in water resources planning and its application in water resources management, these concepts can go well beyond the areas presented in this research for development of other water resource management models.

Recommendations

Keeping in mind the concepts developed in this research and the results demonstrated, recommendations for future research fall into these categories:

1 The data assimilation technique for soil moisture estimation using remotely sensed data can be extended to provide more accurate estimations. Data availability was an impediment for this study and users should look for more

readily available remotely sensed data. If applied on a real-time basis, this application might solve the problem of soil moisture estimation on large agricultural areas. Also depending on data availability, it can be tested whether this algorithm can provide good soil moisture estimates to depths larger than 30 cm using remotely sensed data.

2 The MVRVM root zone soil moisture forecasting algorithm can be extended from a point scale to a field scale where a number of similar MVRVM models are run at the point scale. These predictions could then be used to create spatially interpolated layers of root zone soil moisture in a GIS setting. It should be tested for one irrigation season before it becomes a part of a decision support system.

3 The MCRVM crop classification procedure should draw attention toward the use of data assimilation techniques with the sophisticated MCRVM tool and it can be taken to the next level by using it for pixel-based classification instead of data point classification. Then the results could be used to estimate crop acreage and crop yield in a GIS setting.

REFERENCES

Albergel, C., C. Rüdiger, T. Pellarin, J. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, B. Piguet, and E. Martin. 2008. From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. Hydrology and Earth System Sciences 12: 1323-1337.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, U.S. Geological Survey, Reston, VA. 28 p.

Anderson, M. 2003. SMEX02 watershed vegetation sampling data, Walnut Creek, Iowa. Boulder, CO: National Snow and Ice Data Center. Digital media.

Anderson, M., C. Neale, F. Li, J. Norman, W. Kustas, H. Jayanthi, and J. Chavez. 2004. Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and landsat imagery. Remote Sensing of Environment 92: 447-464.

Arora, V. K., C. Singh, and K. Singh. 1997. Comparative assessment of soil water balance under wheat in a subtropical environment with simplified models. The Journal of Agricultural Science 128: 461-468.

Artiola, J. F., I. L. Pepper, and M. L. Brusseau. 2004. Environmental monitoring and characterization. Elsevier Academic Press, San Diego. 418 p.

Atluri, V., H. Chih-Cheng, and T. L. Coleman. 1999. An artificial neural network for classifying and predicting soil moisture and temperature using Levenberg-Marquardt algorithm. In: Southeastcon '99. Proceedings. IEEE, Lexington. p. 10-13.

Ayala-Silva, T., and C. Beyl. 2005, Changes in spectral reflectance of wheat leaves in response to specific macronutrient deficiency. Advances in Space Research 35(2): 305-317.

Bahadur, K. 2009. Improving Landsat and IRS Image classification: Evaluation of unsupervised and supervised classification through band ratios and DEM in a mountainous landscape in Nepal. Remote Sensing 1: 1257-1272.

Barandela, R., and M. Juarez. 2002. Supervised classification of remotely sensed data with ongoing learning capability. International Journal of Remote Sensing 23: 4965-4970.

Beecham, R. 1995. Patterns of spatial and temporal variability of factors affecting nutrient export from Chaffey Dam Catchment. In: P., Binning, H., Bridgeman, and

B., Williams (Eds.), Proceedings of MODSIM 95 International Congress in Modelling and Simulation, Newcastle. p. 183-187.

Berger, J. 1985. Statistical decision theory and Bayesian analysis. Second edition. Springer, New York.

Betts, A. K., J. H. Ball, A. C. M. Baljaars, M. J. Miller, and P. Viterbo. 1994. Coupling between land-surface, boundary-layer parameterizations and rainfall on local and regional scales: Lessons from the wet summer of 1993. In: Fifth Conference on Global Change Studies: American Meteorological Society, Nashville. p. 174-181.

Bishop, C., and M. Tipping. 1998. A hierarchical latent variable model for data visualization. IEEE Transactions on Pattern Analysis and Machine Intelligence 20: 281-293.

Bowers, S. and R. Hanks. 1965. Reflection of radiant energy form soil. Soil Science 100: 130-138.

Burnash, R. J. C., and V.P. Singh. 1995. The NWS River Forecast System—catchment modelling. Computer Models of Watershed Hydrology 311-366.

Cai, G., Y. Xue, Y. Hu, Y. Wang, J. Guo, Y. Luo, C. Wu, S. Zhong, and S. Qi. 2007. Soil moisture retrieval from MODIS data in northern China plain using thermal inertia model. International Journal of Remote Sensing 28(16): 3567–3581.

Calvet, J., and J. Noilhan. 2000. From near-surface to root-zone soil moisture using year-round data. Journal of Hydrometeorology 1: 393-411.

Calvet, J., J. Noilhan, and P. Bessemoulin. 1998. Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: A feasibility study based on field measurements. Journal of Applied Meteorology 37: 371-386.

Camillo, P., and T. Schmugge. 1983. Estimating soil moisture storage in the root zone from surface measurements. Soil Science 135 (4): 245-264.

Camps-Valls, G., L. Gomez-Chova, J. Calpe-Maravilla, E. Soria-Olivas, J. Martin-Guerrero, and J. Moreno. 2003. Support vector machines for crop classification using hyperspectral data. Lecture notes in computer science 134-141.

Camps-Valls, G., T. B. Marsheva, and D. Zhou. 2007. Semi-supervised graph-based hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 45: 3044-3054.

Carlson, T., and M. Buffum. 1989. On estimating total daily evapotranspiration from remote surface temperature measurements. Remote Sensing of Environment 29(2): 197-207.

Carlson, T., R. Gillies, and E. Perry. 1994. A method to make use of thermal infrared temperature and ndvi measurements to infer surface soil water content and fractional vegetation cover. Remote Sensing Reviews 9: 161-173.

Carlson, T., R. Gillies, and T. Schmugge. 1995. An interpretation of methodologies for indirect measurement of soil water content. Agricultural and Forest Meteorology 77(3-4): 191-205.

Catarina, S., and R. Bernardete. 2006. Two-level hierarchical hybrid SVM-RVM Classification model. In: 5th International Conference on machine learning and applications, (ICMLA '06), Orlando. p. 89-94.

Ceballos, A., K. Scipal, W. Wagner, and J. Martinez-Fernandez. 2005. Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero basin, Spain. Hydrological Processes 19(8): 1549-1566.

Chang, D., and S. Islam. 2000. Estimation of soil physical properties using remote sensing and artificial neural network. Remote Sensing of Environment 74: 534-544.

Colstoun, de, B., C. Eric, and C. Walthall. 2006. Improving global scale land cover classifications with multi-directional POLDER data and a decision tree classifier. Remote Sensing of Environment 100: 474-485.

Comer, G. H., and W. H. Henson. 1976. An optimization technique adapted to usdahl-74 revised model of watershed hydrology. American Water Resources Association 12: 139-146.

Congalton, R. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37: 35-46.

Coronato, F., and M. Bertiller. 1996. Precipitation and landscape related effects on soil moisture in semi-arid rangelands of Patagonia. Journal of Arid Environments 34(1): 1-9.

Das, N. N., and B. P. Mohanty. 2006. Root zone soil moisture assessment using remote sensing and vadose zone modeling. Vadose Zone Journal 5(1): 296-307.

Daughtry, C., C. Walthall, M. Kim, E. De Colstoun, and J. McMurtrey. 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sensing of Environment 74: 229-239.

Demir, B., and S. Erturk. 2007. Hyperspectral Image Classification Using Relevance Vector Machines. Geoscience and Remote Sensing Letters, IEEE 4: 586-590.

Doraiswamy, P., B. Akhmedov, and A. Stern. 2007. Crop classification in the US Corn Belt using MODIS imagery. In: International Geoscience and Remote Sensing Symposium Barcelona, Spain 4.

Elshorbagy, A., and K. Parasuraman. 2008. On the relevance of using artificial neural networks for estimating soil moisture content. Journal of Hydrology 362:1-18.

Elurnnoh, A., and R. P. Shrestha. 2000. Application of DEM data to Landsat image classification: Evaluation in a tropical wet-dry landscape of Thailand. Photogrammetric Engineering and Remote Sensing 66(3): 297-304.

Engman, E., and N. Chauhan. 1995. Status of microwave soil moisture measurements with remote sensing. Remote Sensing of Environment 51(1): 189-198.

Engman, E. T. 1990. Progress in Microwave Remote Sensing of Soil Moisture. Canadian Journal of Remote Sensing 16(3): 6-14.

Engman, E. T. 1992. Soil Moisture Needs in Earth Sciences. In: Proc. International Geoscience and Remote Sensing Symposium (IGARSS) pp. 477-479.

Entekhabi, D., H. Nakamura, and E. G. Njoku. 1993. Retrieval of Soil Moisture by Combined Remote Sensing and Modeling. In: B. J. Choudhury, Y. H. Kerr, E. G. Njoku, and P. Pampaloni, (Eds.), ESA/NASA International Workshop on Passive Microwave Remote Sensing Research Related to Land-Atmosphere Interactions, St. Lary, France. p. 485-498.

Entekhabi, D., H. Nakamura, and E. Njoku. 1994. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. IEEE Transactions on Geoscience and Remote Sensing 32: 438-448.

Fast, J. D., and M. D. McCorcle. 1991. The effect of heterogenous soil moisture on a summer baroclinic circulation in the central United States. Monthly Weather Review 119: 2140-2167.

Fawcett, T. 2004. ROC graphs: Notes and practical considerations for researchers. Machine Learning 31: 1-38.

Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27: 861-874.

Fernandez-Galvez, J. 2008. Errors in soil moisture content estimates induced by uncertainties in the effective soil dielectric constant. International Journal of Remote Sensing 29(11): 3317-3323.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of

Eugenics 7: 179–188.

Foody, G. M. 1995. Using prior knowledge in artificial neural network classification with a minimal training set. International Journal of Remote Sensing 16: 301-312.

Foody, G. M. 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80(1): 185–201.

Foody, G. M. 2008. RVM-based multi-class classification of remotely sensed data. International Journal of Remote Sensing 29: 1817-1823.

Foody, G. M., and A. Mathur. 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on Geoscience and Remote Sensing 42: 1335-1343.

Fu, B., and H. Gulinck. 1994. Land evaluation in an area of severe erosion: the Loess Plateau of China. Land Degradation and Development 5: 33-40.

Fung, G., and O. Mangasarian. 2005. Multicategory proximal support vector machine classifiers. Machine Learning 59: 77-97.

Giacomelli, A., U. Bacchiega, P. A. Troch, and M. Mancini. 1995. Evaluation of surface soil moisture distribution by means of SAR remote sensing techniques and conceptual hydrological modelling. Journal of Hydrology 166: 445-459.

Gill, M., T. Asefa, M. Kemblowski, and M. Mckee. 2006. Soil moisture prediction using support vector machines. Journal of the American Water Resources Association 42: 1033-1046.

Gill, M., M. Kemblowski, and M. McKee. 2007. Soil Moisture Data Assimilation Using Support Vector Machines and Ensemble Kalman Filter. Journal of the American Water Resources Association 43: 1004-1015.

Gillies, R., and T. Carlson. 1995. Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. Journal of Applied Meteorology 34 (4): 745-756.

Gillies, R., T. Carlson, J. Cui, W. Kustas, and K. Humes. 1997. A verification of the'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the normalized difference vegetation index (NDVI) and surface radiant temperature. International Journal of Remote Sensing 18: 3145-3166.

Gilman, K. 1980. Estimating the soil heat flux in an upland drainage basin. Hydrological Sciences-Bulletin-des Sciences Hydrologiques 25(4): 435.

Gopal, S., and C. Woodcock. 1996. Remote sensing of forest change using artificial neural networks. IEEE Transactions on Geoscience and Remote Sensing 34: 398-404.

Güngör, Z., and A. Ünler. 2007. K-harmonic means data clustering with simulated annealing heuristic. Applied Mathematics and Computation 184: 199-209.

Han, J., and M. Kamber. 2006. Data mining: Concepts and techniques. Morgan Kaufmann, San Francisco. 772 p.

Hayat, M. A. 2007. Cancer Imaging: Instrumentation and applications. Elsevier, Academic Press, London. 733 p.

Healey, G., and A. Jain. 1996. Retrieving multispectral satellite images using physics-basedinvariant representations. IEEE Transactions on Pattern Analysis and Machine Intelligence 18: 842-848.

Heermann, P., and N. Khazenie. 1992. Classification of multispectral remote sensing data using aback-propagation neural network. IEEE Transactions on Geoscience and Remote Sensing 30:81-88.

Heinzel, V., B. Waske, M. Braun, and G. Menz. 2007. Remote sensing data assimilation for regional crop growth modelling in the region of Bonn (Germany). In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'07), Barcelona, Spain. p. 3647-3650.

Hepner, G., T. Logan, N. Ritter, and N. Bryant. 1990. Artificial neural network classification using a minimal training set- Comparison to conventional supervised classification. Photogrammetric Engineering and Remote Sensing 56: 469-473.

Hermes, L., D. Frieauff, J. Puzicha, and J. Buhmann. 1999. Support vector machines for land usage classification in Landsat TM imagery. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'99), Hamburg, Germany. p. 348–350.

Hill, A., and V. Neary. 2009. Hydrologic response of a forested sinkhole wetland to different land management scenarios. Journal of Environmental Hydrology 17.

Holtan, H., G. Stiltner, W. Henson, and N. Lopez. 1975. USDAHL-74 Revised model of watershed hydrology: A United States contribution to the international hydrological decade. Agricultural Research Service, US Dept. of Agriculture. Technical Bulletin 1518, Washington, D.C. 99 p.

Huang, C., L. Davis, and J. Townshend. 2002. An assessment of support vector machines for land cover classification. International Journal of Remote Sensing 23: 725-749.

Huete, A. 1988. A soil-adjusted vegetation index(SAVI). Remote Sensing of Environment 25: 295-309.

Humes, K., W. Kustas, T. Jackson, T. Schmugge, and M. Moran. 1993. Combined use of optical and microwave remotely sensed data for the estimation of surface energy balance components over a semi-arid watershed. Proceedings of IEEE Topical Symposium on Combined Optical, Microwave, Earth and Atmosphere Sensing, 22-25 Mar. 1993, Albuquerque, New Mexico. p. 86-89.

Hutchinson, C. F. 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. Photogmmmetric Engineering and Remote Sensing 48: 123-130.

Idso, S., R. Jackson, and R. Reginato. 1975. Estimating evaporation: A technique adaptable to remote sensing. Science 189: 991-992.

Idso, S., R. Jackson, and R. Reginato. 1976. Compensating for environmental variability in the thermal inertia approach to remote sensing of soil moisture. Journal of Applied Meteorology 15: 811-817.

Islam, S. I., and E. T. Engman. 1996. Why bother for 0.0001% of earth's water? Challenges for soil moisture research. Eos, Transactions American Geophysical Union 77: 420.

Jackson, T. 1986. Soil water modeling and remote sensing. IEEE Transactions on Geoscience and Remote Sensing 24(1):37-46.

Jackson, T., D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doriaswamy, and E. Hunt. 2004. Vegetation water content mapping using landsat data derived normalized difference water index for corn and soybeans. Remote Sensing of Environment 92 (4): 475-482.

Jackson, T., P. O'Neill, and C. Swift. 1997. Passive microwave observation of diurnal surface soil moisture. IEEE Transactions on Geoscience and Remote Sensing 35 (5):1210-1222.

Jackson, T., and T. Schmugge. 1989. Passive microwave remote sensing system for soil moisture: Some supporting research. IEEE Transactions on Geoscience and Remote Sensing 27 (2): 225-235.

Jackson, T. J. 1982. Survey of applications of passive microwave remote sensing for soil moisture in the USSR. EOS Transactions of the American Geophysical Union 63(19): 497-499.

Jackson, T. J., and M. H. Cosh. 2003a. SMEX02 Landsat Thematic Mapper Imagery, Iowa. Boulder, CO: National Snow and Ice Data Center. Digital media. Available online at: http://nsidc.org/data/nsidc-0199.html

Jackson, T. J., and M. H. Cosh. 2003b. SMEX02 Watershed soil moisture data, Walnut Creek, Iowa. Boulder, CO: National Snow and Ice Data Center. Digital media. Available online at: ftp://sidads.colorado.edu/pub/DATASETS/AVDM/data/soil_moisture/ SMEX02

Jackson, T. J., M. E. Hawley, and P. E. O'Neill. 1987. Preplanting soil moisture using passive microwave sensors. Water Resources Bulletin 23(1): 11-19.

Jackson, T. J., D. M. 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 Transactions on Geoscience and Remote Sensing 37 (5):2136-2151.

Jackson, T. J., and L. McKee. 2004. SMEX03 Oklahoma Vegetation Data. Boulder, CO: National Snow and Ice Data Center. Digital media.

Jackson, T. J., T. J. Schmugge, A. D. Nicks, G. A. Coleman, and E. T. Engman. 1981. Soil moisture updating and microwave remote sensing for hydrological simulation. Hydrological Sciences Bulletin 26(3): 305-319.

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

Jacquemoud S., and S. Ustin. 2001. Leaf optical properties: a state of the art. In 8th International Symposium of Physical Measurements & Signatures in Remote Sensing, Aussois, France 8: 223–32.

Jensen, J.R. 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. 2nd ed. Prentice-Hall, Inc, Upper Saddle River, NJ. 316 pp.

Jiang, H., and W. Cotton. 2004. Soil moisture estimation using an artificial neural network: a feasibility study. Canadian Journal of Remote Sensing 30:827-839.

Jones, H. 2007. Monitoring plant and soil water status: Established and novel methods revisited and their relevance to studies of drought tolerance. Journal of Experimental Botany 58 (2):119.

Kaleita, A., L. Tian, and M. Hirschi. 2005. Relationship between soil moisture content and soil surface reflectance. Transactions of the ASAE 48: 1979-1986.

Khalil, A., M. K. Gill, and M. Mckee. 2005. New applications for information fusion and soil moisture forecasting. In: 8th International Conference on Information Fusion, Philadelphia, PA 1622-1628.

Khalil, A., M. Mckee, and J. Kaluarachchi. 2005. Applicability of statistical learning algorithms in groundwater quality modeling. Water resources research 41: 5010-5010.

Lark, R. 1999. Soil–landform relationships at within-field scales: an investigation using continuous classification. Geoderma 92: 141-165.

Lawrence, R., and A. Wright. 2001. Rule-based classification systems using classification and regression tree (CART) analysis. Photogrammetric Engineering and Remote Sensing 67: 1137-1142.

Le Bris, A., and D. Boldo. 2008. Extraction of land cover themes from aerial ortho-images in mountainous areas using external information. The Photogrammetric Record 23: 387-404.

Legates, D., and G. McCabe. 1999. Evaluating the use of" Goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resources Research 35: 233-241.

Li, F., T. Jackson, W. Kustas, T. Schmugge, A. French, M. Cosh, and R. Bindlish. 2004. Deriving land surface temperature from Landsat 5 and 7 during SMEX02/ SMACEX. Remote Sensing of Environment 92(4): 521-534.

Li, J., and S. Islam. 2002. Estimation of root zone soil moisture and surface fluxes partitioning using near surface soil moisture measurements. Journal of Hydrology 259:1-14.

Liu, W., F. Baret, X. Gu, B. Zhang, Q. Tong, and L. Zheng. 2003. Evaluation of methods for soil surface moisture estimation from reflectance data. International Journal of Remote Sensing 24: 2069-2083.

Lu, S., Z. Ju, T. Ren, and R. Horton. 2009. A general approach to estimate soil water content from Thermal Inertia. Agricultural and Forest Meteorology 149 (10): 1693-1698.

Magnussen, S., R. McRoberts, and E. Tomppo. 2009. Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. Remote Sensing of Environment 113: 476-488.

Mahfouf, J. F. 1991. Analysis of soil moisture from near-surface parameters: A feasibility study. Journal of Applied Meteorology 30(11): 1534-1547.

Majumdar, T. J., and B. B. Bhattacharya. 1990. Simulation of thermal inertia imagery with daytime HCMM data. International Journal of Remote Sensing 11(1): 139 – 147.

Mas, J. 2004. Mapping land use/cover in a tropical coastal area using satellite sensor data, GIS and artificial neural networks. Estuarine, Coastal and Shelf Science 59:219-230.

Mather, J.K. 1974. Climatology fundamentals and applications, McGraw-Hill, New York 151-153.

Mathur, A., and G. Foody. 2008. Crop classification by support vector machine with intelligently selected training data for an operational application. International Journal of Remote Sensing 29: 2227-2240.

McKenzie, N., and M. Austin. 1993. A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. Geoderma 57: 329-355.

McNairn, H., J. Ellis, J. Van Der Sanden, T. Hirose, and R. Brown. 2002. Providing crop information using RADARSAT-1 and satellite optical imagery. International Journal of Remote Sensing 23: 851-870.

Meier, I., and C. Leuschner. 2008. Leaf size and leaf area index in Fagus Sylvatica forests: Competing effects of precipitation, temperature, and nitrogen availability. Ecosystems 11 (5): 655-669.

Melgani, F., and L. Bruzzone 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing 42: 1778-1790.

Mercier, G., and M. Lennon. 2003. Support vector machines for hyperspectral image classification with spectral-based kernels. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '03), Toulouse, France. p. 288-290.

Miller, S., D. Guertin, and D. Goodrich 2007. Hydrologic modeling uncertainty resulting from land cover misclassification. Journal of the American Water Resources Association 43(4): 1065-1075.

Mishra, A., S. Kar, and V. Singh. 2007. Prioritizing Structural Management by Quantifying the Effect of Land Use and Land Cover on Watershed Runoff and Sediment Yield. Water Resources Management 21: 1899-1913.

Mitra, D., and T. Majumdar. 2004 Thermal inertia mapping over the Brahmaputra basin, India using NOAA-AVHRR data and its possible geological applications. International Journal of Remote Sensing 25(16): 3245-3260.

Mohanty, K. K., and T. J. Majumdar. 1996. An artificial neural network (ANN) based software package for classification of remotely sensed data. Computers and Geosciences 22: 81-87.

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

Moran, M., C. Peters-Lidard, J. Watts, and S. Mcelroy. 2004. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. Canadian Journal of Remote Sensing 30(5): 805-826.

Mouazen, A., J. De Baerdemaeker, and H. Ramon. 2005. Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. Soil & Tillage Rresearch 80: 171-183.

Mukherjee, S., E. Osuna, and F. Girosi. 1997. Nonlinear prediction of chaotic time series using support vector machines. In: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing 7, Amelia Island, p. 511-520.

Muller, E., and H. Decamps. 2001. Modeling soil moisture-reflectance. Remote Sensing of Environment 76: 173-180.

Muñoz-Marí, J., L. Gómez-Chova, G. Camps-Valls, and J. Calpe-Maravilla. 2008. Image classification with semi-supervised one-class support vector machine. Image and Signal Processing for Remote Sensing XIV. Edited by Bruzzone, Lorenzo; Notarnicola, Claudia; Posa, Francesco. Proceedings of the SPIE 7109:71090B-71090B.

Nash, J., and J. Sutcliffe. 1970. River flow forecasting through conceptual models part I-A discussion of principles. Journal of Hydrology 10: 282-290.

Newton, R. W., J. L. Heilman, and C. H. M. van Bavel. 1983. Integrating Passive Microwave Measurements with a Soil Moisture/Heat Flow Model. Agricultural Water Management 7: 379-389.

Njoku, E., and D. Entekhabi. 1996. Passive microwave remote sensing of soil moisture. Journal of Hydrology. 184 (1-2):101-129.

Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem. 2003. Soil moisture retrieval from AMSR-E. IEEE Transactions on Geoscience and Remote Sensing 41(2): 215-229.

Ozdarici A., and M. Turker. 2007. Field-Based Classification of Different Resolution Images and the Filtering Effects on the Accuracies. In: 3rd International Conference on Recent Advances in Space Technologies (RAST '07), Istanbul, Turkey. p. 321-325.

Pal, M., and P. Mather. 2005. Support vector machines for classification in remote sensing. International Journal of Remote Sensing 26: 1007-1011.

Peck, E. L. 1976. Catchment modeling and initial parameter estimation for the National Weather Service river forecast system. NOAA Technical Memorandum NWS HYDRO 31.

Petropoulos, G., T. N. Carlson, M. J. Wooster, and S. Islam. 2009. A review of ts/vi remote sensing based methods for the retrieval of land surface energy fluxes and soil surface moisture. Progress in Physical Geography 33(2): 224-250.

Pratt, D., and C. Ellyett. 1979. Thermal inertia approach to mapping of soil moisture and geology. Remote Sensing of Environment 8: 151-168.

Price, J. 1980. The potential of remotely sensed thermal infrared data to infer surface soil moisture and evaporation. Water Resources Research 16(4): 787–795.

Price, J. 1985. On the analysis of thermal infrared imagery: The limited utility of apparent thermal inertia. Remote Sensing of Environment 18(1): 59-73.

Provost, F., and P. Domingos. 2001. Well-trained PETs: Improving probability estimation trees. Information Systems Department, Stern School of Business, New York University.

Prueger J. 2004. SMEX02 Rain Gauge Network, Walnut Creek, Iowa. Boulder, CO: National Snow and Ice Data Center. Digital media. Available online at: ftp://sidads.colorado.edu/pub/DATASETS/AVDM/data/soil_moisture/SMEX02/meteorological/WC_raingage/

Qiu, Y., B. Fu, J. Wang, and L. Chen. 2001. Soil moisture variation in relation to topography and land use in a hillslope catchment of the Loess Plateau, China. Journal of Hydrology 240: 243-263.

Qiu, Y., B. Fu, J. Wang, and L. Chen, 2003. Spatiotemporal prediction of soil moisture content using multiple-linear regression in a small catchment of the Loess Plateau, China. CATENA 54:173-195.

Quattrochi, D., and J. Luvall. 1999. Thermal infrared remote sensing for analysis of landscape ecological processes: Methods and applications. Landscape Ecology 14: 577-598.

Rao, A., and K. Saxton 1995. Analysis of soil water and water stress for pearl millet in an Indian arid region using the SPAW Model. Journal of Arid Environments 29: 155-167.

Ratanopad S., and W. Kainz. 2006. Land cover classification and monitoring in Northeast Thailand using Landsat 5 TM data. In: ISPRS Technical Commission II Symposium, Vienna, Austria. p. 12-14.

Reginato, R., S. Idso, R. Jackson, J. Vedder, M. Blanchard, and R. Goettelman. 1976. Soil water content and evaporation determined by thermal parameters obtained from ground-based and remote measurements. Journal of Geophysical Research 81: 1617-1620.

Reginato, R., R. Jackson, and P. Pinter. 1985. Evapotranspiration calculated from remate multispectral and ground station meteorological data. Remote Sensing of Environment 18: 75-89.

Ricchetti, E. 2000. Multispectral satellite image and ancillary data integration for geological classification. Photogrammetric Engineering and Remote Sensing 66(4):429-435.

Richards J. A., and X. Jia. 2006. Remote sensing digital image analysis: An Introduction. Fourth Edition. Springer-Verlag, New York. 439 p.

Roli, F., and G. Fumera. 2001. Support vector machines for remote-sensing image classification. In: Image and Signal Processing for Remote Sensing VI, 27-29 September, 2000, Barcelona, Spain. p. 4170:160.

Sabater, J., L. Jarlan, J. Calvet, F. Bouyssel, and P. De Rosnay. 2007. From near-surface to root-zone soil moisture using different assimilation techniques. Journal of Hydrometeorology 8: 194-206.

Sabins, F. F. 1978. Remote Sensing-Principles and Interpretation. W.H. Freeman, San Francisco. 426 p.

Saha, S. K. 1995. Assesment of Regional Soil Moisture Conditions by Coupling Satellite Sensor Data with a Soil-Plant System Heat and Moisture Balance Model. International Journal of Remote sensing 16(5): 973-980.

Samaniego, L., A. Bardossy, and K. Schulz. 2008. Supervised classification of remotely sensed imagery using a modified k-NN technique. IEEE Transactions on Geoscience and Remote Sensing 46: 2112-2125.

Sandholt, I., K. Rasmussen, and J. Andersen. 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sensing of Environment 79: 213-224.

Saxton, K. E., H. P. Johnson, and R. H. Shaw. 1974. Modeling evapotranspiration and soil moisture. Transactions of the ASAE 17: 673-677.

Scott, C. A., W. G. M. Bastiaanssen, and M. U. D. Ahmad. 2003. Mapping root zone soil moisture using remotely sensed optical imagery. Journal of Irrigation and Drainage Engineering 129: 326-335.

Scott, H., and R. Geddes. 1979. Plant water stress of soybean (Glycine max) and common cocklebur (Xanthium Pensylvanicum): A comparison under field conditions. Weed Science 27(3): 285-289.

Silva, C., and B. Ribeiro. 2006. Automated Learning of RVM for Large Scale Text Sets: Divide to Conquer. In: Intelligent Data Engineering and Automated Learning – IDEAL 2006: 878-886.

Skidmore, E., J. Dickerson, and H. Schimmelpfennig. 1975. Evaluating surface-soil water content by measuring reflectance. Soil Science Society of America Journal 39: 238-242.

Smith, M. B., K. P. Georgakakos, and X. Liang. 2004. The distributed model intercomparison project (DMIP). Journal of Hydrology 298: 1-3.

Smits, P., S. Dellepiane, and R. Schowengerdt. 1999. Quality assessment of image classification algorithms for land-cover mapping: A review and a proposal for a cost-based approach. International Journal of Remote Sensing 20:1461-1486.

Song, J., D., Wang, N., Liu, L. Cheng, L. Du, and K. Zhang. 2008. Soil moisture prediction with feature selection using a neural network. Computing: Techniques and Applications, DICTA'08. Digital Image, Washington, DC,USA 130-136.

Sorooshian, S., Q. Duan, and V. Gupta. 1993. Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model. Water Resources Research 29: 1185-1194.

Strahler, A. H., T. L. Logan, and N. A. Bryant. 1978. Improving forest cover classification accuracy from Landsat by incorporating topographic information. Proceedings of 12th International Symposium on Remote Sensing of Environment, 20-26 April, Manila, Philippines 927-942.

Thayananthan, A., R. Navaratnam, B. Stenger, P. Torr, and R. Cipolla. 2006. Multivariate relevance vector machines for tracking. Lecture Notes in Computer Science 3953: 124.

Tipping, M. E., and A. C. Faul. 2003. Fast marginal likelihood maximization for sparse bayesian models. In: C. M. Bishop, and B. J. Frey, (Eds.), Proceedings of Ninth

International Workshop on Artificial Intelligence and Statistics, Society for Artificial Intelligence and Statistics.

Tipping, M. 2000. The relevance vector machine. Advances in Neural Information Processing Systems 12: 652-658.

Tipping, M. 2001. Sparse bayesian learning and the relevance vector machine. Journal of Machine Learning Research 1: 211-244.

Topp, G. C., J. L. Davis, and A. P. Annan. 1980. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. Water Resources Research 16(3): 574-582.

Trotter, C.M. 1991. Remotely-sensed data as an information source for geographical information systems in natural resource management: A review. International Journal of Geographic Information Systems 5: 225-239.

Tucker, C., J. Townshend, and T. Goff. 1985. African land-cover classification using satellite data. Science 227: 369-375.

Tymkow, P., and A. Borkowski. 2008. land cover classification using airborne laser scanning data and photographs. ISPRS08:B3b: 185 ff.

Vapnik, V., and A. Chervonenkis. 1991. The necessary and sufficient conditions for consistency in the empirical risk minimization method. Pattern Recognition and Image Analysis 1(3): 283-305.

Vapnik, V. 1995. The nature of statistical learning theory. Springer Verlag, New York. 188 p.

Vapnik, V. 1998. Statistical learning theory. Wiley-Interscience, New York. 768 p.

Vapnik, V. 2000. The nature of statistical learning theory. Second edition. Springer Verlag, New York. 314 p.

Vatsavai, R., B. Badhuri, S. Shekhar, and T. Burk. 2008. Multisource data classification using a hybrid semi-supervised learning scheme. Wetlands 6: 28.38-66.50.

Velickov, S., and D. P. Solomatine. 2000. Predictive data mining: practical examples. In: O. Schleider and A. Zijderveld (Eds.), AI methods in Civil Engineering Applications, Cottbus. p. 3-19.

Verstraeten, W. W., F. Veroustraete, C. J. Van Der Sande, I. Grootaers, and J. Feyen, 2006. Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests. Remote Sensing of Environment 101(3): 299-314.

Wagner, W., and K. Scipal. 2000. Large scale soil moisture monitoring using C-Band Scatterometer data, Remote Sensing and Hydrology Symposium. In: Proceedings of a symposium held at Santa Fe, New Mexico, 267: 405–408.

Wagner, W., G. Lemoine, and H. Rott. 1999. A method for estimating soil moisture from ers scatterometer and soil data. Remote Sensing of Environment 70(2): 191-207.

Wagner, W., K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf. 2003. Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. Journal of Geophysical Research-Atmospheres 108 (D19): 4611.

Walker, J., G. Willgoose, and J. Kalma. 2001a. One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: A simplified soil moisture model and field application. Journal of hydrometeorology 2: 356-373.

Walker, J., G. Willgoose, and J. Kalma. 2001b. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms. Advances in Water Resources 24: 631-650.

Wang, P., L. Xiao-Wen, S. Wei, L. Xing-Min, Z. Shu-Yu, and L. An-Lin. 2004. Soil thermal inertia estimation by combining afternoon and morning AVHRR data with a modified diurnal land surface temperature change modeled. Geoscience and Remote Sensing Symposium IGARSS '04 6: 4299-4301.

Waring, R., and B. Cleary. 1967. Plant moisture stress: Evaluation by pressure bomb. Science 155: 1248-1254.

Watson, K. 1975. Geologic applications of thermal infrared images. Proceedings of the IEEE 63(1): 128-137.

Webster, R., and B. Butler. 1976. Soil classification and survey studies at Ginninderra. Australian Journal of Soil Research 14: 1-24.

Wei, L., Y. Yang, R. Nishikawa, M. Wernick, and A. Edwards. 2005. Relevance vector machine for automatic detection of clustered microcalcifications. IEEE Transactions on Medical Imaging 24: 1278-1285.

Western, A. W., T. R. Green, and R. B. Grayson. 1997. Hydrological Modelling of the Tarrawarra Catchment; Use of Soil Moisture Patterns. In: A. D. McDonald, and M. McAleer (Eds.), Proc. MODSIM 97 International Congress on Modelling and Simulation, Hobart, Australia. P. 409-416.

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

Wigneron, J. P., T. Schmugge, A. Chanzy, J. C. Calvet, and Y. Kerr. 1998. Use of Passive Microwave Remote Sensing to Monitor Soil Moisture. Agronomie 18(1): 27-43.

Wilcox, B., W. Rawls, D. Brakensiek, and J. Wight. 1990. Predicting runoff from rangeland catchments: A comparison of two models. Water Resources Research 26: 2401-2410.

Willmott, C. 1984. On the evaluation of model performance in physical geography. Spatial Statistics and Models 443–460.

Wipf, D., and S. Nagarajan. 2007. Beamforming using the relevance vector machine, Proceedings of the 24th International Conference on Machine learning, June 20-24, 2007, Corvalis, Oregon. p. 1023-1030.

Wong, S., and R. Cipolla. 2005. Real-time adaptive hand motion recognition using a sparse Bayesian classifier. Lecture notes in computer science 3766:170.

Xiao W., Z. Zhang, and W. Tan. 2005. Using temperature/vegetation index to assess surface soil moisture status. International Geoscience and Remote Sensing Symposium 6: 4493-4496.

Xu, H., Y. Ying, X. Fu, and S. Zhu. 2007. Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. Biosystems Engineering 96(4): 447-454.

Xu, W., B. Wu, J. Huang, Y. Zhang, and Y. Tian. 2004. A segmentation and classification approach of land cover mapping using Quick Bird image. In: International Geoscience and Remote Sensing Symposium (IGARSS). Anchorage, AK, United States p. 162-171.

Xue, Y., and A. Cracknell. 1995. Advanced thermal inertia modelling. International Journal of Remote Sensing 16 (3): 431-446.

Yang, S., and Y. Huang. 2009. Application of support vector machine based on time series for soil moisture and nitrate nitrogen content prediction. Computer and Computing Technologies in Agriculture II 3: 2037-2045.

Young, D., and P. Nobel. 1986. Predictions of soil-water potentials in the North-Western Sonoran desert. The Journal of Ecology 74(1): 143-154.

Zhang, H., and J. Malik. 2005. Selecting shape features using multi-class relevance vector machine. Technical Report UCB/EECS-2006-6, Electrical Engineering and Computer Sciences, University of California at Berkeley, USA.

APPENDICES

Appendix A. Geo-location of the soil moisture sampling points for Walnut Creek Watershed, Ames, Iowa.

Site	LowerLeft Latitude	Lower Left Longitude	UpperRight Latitude	UpperRight Longitude	Sampling Location Latitude	Sampling Location Longitude
WC01	41.9653	-93.7662	41.9725	-93.7566	41.9688	-93.7613
WC03	41.9798	-93.7566	41.9869	-93.7471	41.9833	-93.7499
WC04	41.9725	-93.7467	41.9796	-93.7372	41.9742	-93.7431
WC05	41.9578	-93.7466	41.9649	-93.7371	41.9622	-93.7412
WC06	41.928	-93.756	41.9361	-93.7502	41.9312	-93.752
WC08	41.9218	-93.7274	41.9288	-93.7178	41.9252	-93.7228
WC09	41.9216	-93.7078	41.9324	-93.6983	41.9255	-93.703
WC10	41.974	-93.6925	41.9811	-93.6878	41.9755	-93.6904
WC11	41.9713	-93.6963	41.9757	-93.6925	41.9612	-93.6878
WC12	41.9578	-93.6924	41.9648	-93.678	41.9733	-93.6944
WC13	41.9506	-93.6925	41.9576	-93.6828	41.9547	-93.6876
WC14	41.9434	-93.6979	41.9504	-93.6926	41.9469	-93.6956
*WC15	41.9362	-93.6739	41.9433	-93.659	41.939	-93.6643
*WC16	41.9326	-93.6697	41.9361	-93.6594	41.9341	-93.6656
WC17	41.9576	-93.6582	41.9647	-93.6489	41.9608	-93.654
WC18	41.9434	-93.6587	41.9503	-93.6536	41.9461	-93.656
WC19	41.9289	-93.6494	41.9363	-93.6399	41.9315	-93.6446
WC20	41.9217	-93.6494	41.929	-93.6399	41.9241	-93.6446
WC21	41.9648	-93.639	41.972	-93.6295	41.9686	-93.6345
WC22	41.9436	-93.6329	41.9503	-93.6219	41.9473	-93.6256
*WC23	41.9868	-93.5406	41.994	-93.5312	41.9908	-93.5372
*WC24	41.9868	-93.5299	41.994	-93.5261	41.991	-93.5276
WC25	41.9395	-93.5409	41.9459	-93.5346	41.9416	-93.5369
WC26	41.9723	-93.5115	41.9795	-93.5021	41.9764	-93.5066
WC27	41.9579	-93.463	41.9647	-93.4543	41.9609	-93.4582
WC28	41.9212	-93.4591	41.9282	-93.4481	41.9248	-93.4523
WC29	41.9829	-93.4347	41.9936	-93.4252	41.9869	-93.4319
WC30	41.9649	-93.4252	41.9704	-93.4161	41.9678	-93.4215
WC31	41.9652	-93.4155	41.9719	-93.4061	41.9679	-93.4105
WC32	41.9761	-93.6584	41.9793	-93.6423	41.978	-93.6466
WC33	41.9676	-93.6583	41.9762	-93.6395	41.9722	-93.6466

*Test sites

Appendix B. List of Symbols

Symbol	Description
P(A|B;C)	Probability of A given B and C
$\mathbf{x}$	Input matrix
y	Modeled value of soil moisture
$\mathbf{t}$	Observed values of soil moisture
$\{\mathbf{x}_n, y_n\}_{n=1}^{N}$	Inputs and modeled output pairs, where N is the number of training samples.
$\boldsymbol{y}(\mathbf{x}_n; \boldsymbol{w})$	Output vector as a function of inputs and weights
y_n (n=1,2,..,N)	Output vector, where N is the number of training samples.
ε_n	Deviation or generalization error bound
t_n (n=1,2,…, N)	n^{th} observed value of soil moisture
X ~ N(μ, σ^2)	is used to signify that X is normally distributed with mean μ and variance σ^2
$\boldsymbol{\varphi}(\mathbf{x}_n)$	Mapping function or basis function used to transform input vector $\mathbf{x}$
$\boldsymbol{\Phi}$	Matrix whose rows contain the response of all basis functions to the inputs $\boldsymbol{\varphi}(\mathbf{x}_n)$
$\boldsymbol{\Phi}^{\mathbf{T}}$	Transpose of $\boldsymbol{\Phi}$ matrix
$\boldsymbol{w}$	Weight vector
$k(\mathbf{x}_n, \mathbf{x}_{m-1})$	Kernel function
σ^2	Variance
$\boldsymbol{\alpha}$	Hyperparameter – linear expansion coefficient of the weight vector $\boldsymbol{w}$
α_i	i^{th} hyperparameter
β	Inverse of variance - 1/ σ^2
μ	Mean of the distribution
∑	Covariance of the distribution
α_{MP}	Most probable value of the hyperparameter α
σ^2_{MP}	Most probable value of the variance, σ^2
$\mathbf{I}$	Identity matrix
$\boldsymbol{A}$	Diagonal matrix with non-zero elements given by the vector of hyper-parameters α denoted by diag($\alpha_0,\alpha_1,\ldots,\alpha_N$)
$\mathbf{x}_{N+1}$	New input
y_{N+1}	New modeled output
t_{N+1}	New observed output
σ^2_{N+1}	Variance of the modeled output
μm	Micrometer
nm	Nanometer
f(x)	Function
$\mathbf{R}$	Real value
γ	Kernel parameter
$\langle \boldsymbol{w}, \mathbf{x} \rangle$	Inner product of weight vector and input matrix
C	Cost parameter
λ	1/C
b	Bias
ξ_i and ξ_i^*	Slack variables
α and α^*	Lagrange multipliers

CURRICULUM VITAE

Bushra Zaman
PhD Candidate
Utah State University
Utah Water Research Laboratory,
1600 Canyon Road,
Logan, UT, 84322-8200, USA
UMC 8200
Office: (435) 797-3149, Cell: (435) 764-2674
b.zaman@aggiemail.usu.edu
Visa status: F1

EDUCATION

Ph.D., Civil and Environmental Engineering (Water Resources), (Jan 2007 - May 2010)
Utah State University, Logan, Utah, USA.
Dissertation title: Fusion of Remotely Sensed Data for Water Resources Management using Relevance Vector Machines

M.Tech., Civil Engineering (Water Resources), (August 2000 - Jan 2002)
Indian Institute of Technology, Delhi, India.
Thesis title: Software Development for Ground Water Contamination Problem Using FEM (Finite Element Method) Technique

B.Tech., Civil Engineering, (November 1995- May 1999)
Bihar Institute of Technology, Sindri, Bihar, India.
Thesis title: Air pollution due to effluents from a coal-based thermal power plant.

CAREER OBJECTIVE: To utilize extensive work experience and research to develop decision support system tools for water resources management using remote sensing applications and learning machines.

PROFESSIONAL SUMMARY

- Graduate Coursework and emphasis in Water management
- Experience in developing decision support system models for water resources management.
- Over 4 years of professional experience in consulting and construction engineering.
- Extensive domestic/international experience in the field of civil engineering.
- Project management experience and training in underground rail system.
- Excellent interpersonal and communication skills; fluent in English, Urdu and Hindi.

PROFESSIONAL EXPERIENCE

Envision Utah – June 2008

- Developed a water distribution system design model for growth cost modeling. The design is being implemented on newly developed areas of Hyrum city.

Section Engineer - Delhi Metro Rail Corporation – October 2004 –December 2006

- Project monitoring & scheduling using PRIMAVERA and MS Project software packages. Preparation of Bar charts and milestone charts. Preparation of quarterly, monthly and weekly progress reports.
- Preparation of tender documents, pre-qualification documents, Tender Evaluation, Tender Policy, GCC, SCC etc.

Design Engineer - Tata Consulting Engineers, Mumbai –August 2002 –July 2003

- Dealt with analysis and design of water distribution and supply; and network modeling using WaterCAD; pipeline design, water hammer problems (Using WaterHAM software package) etc.

Trainee Engineer - Kirti Consultants, New Delhi - August 1999 – June 2000

- Dealt with structural analysis and design of concrete structures. Analysis and design of Beams, slabs etc. Detailing of reinforcement drawing. Estimation of civil & structural quantities, preparation of BBS (Bar bending schedule)

RESEARCH EXPERIENCE

Graduate Research Assistant, Utah Water Research Laboratory - May 2007-Present

- Developing Bayesian decision support system models (Relevance Vector Machine) for water resources management by data fusion of remotely sensed data. ArcGIS and ERDAS Imagine were also used.
- Working on the development of models related to processing and building models for landcover recognition. The purpose is to build Bayesian statistical routines by using inexpensive and readily available data and making use of the cutting edge technology that is being developed at USU for acquiring airborne imagery using drones. These unmanned aircrafts will acquire hyperspectral data.

Research Assistant - Indian Institute of Technology, Delhi –July 2003–October 2004

- Weir design using Khosla's theory – Used NISA (Finite element software for analyzing the flow problems)
- Optimal Barrage Design based on Subsurface Flow Considerations
  Used VENSIM Software Package- For developing a dynamic feedback model for Integrated River Basin Management for the seven major River Basins in India.

CONFERENCE PRESENTATION

- Conference presentation at AWRA 2008 Annual Water Resources Conference. New Orleans, Louisiana. *Fusion of remotely sensed data for Profile Soil Moisture Retrieval using RVMs and SVMs.*
- Conference presentation at 2009 AWRA Summer Specialty Conference. Snowbird, Utah. *Fusion of remotely sensed data for Landcover classification using Multiclass relevance vector machine.*

INDUSTRIAL TRAINING

- "Management Development Program in Advanced Computerized Project Management", at National Institute of Construction Management and Research, Pune.
- "Concrete Technology for Water Resources Structures" at Central Soil and Materials Research Institute, New Delhi.
- "Water and Energy Conservation" organized by Construction Industry Development Council (CIDC), at Radisson Hotel, New Delhi.

AFFILIATIONS

- American Water Resources Association (*Student Member*)
- American Geophysical Union (*Student Member*)
- Golden Key International Honor Society (*Invited Member*)
- Women's Center (Utah State University) – Volunteered for fundraising and awareness events.
- Community Abuse Prevention Services Agency (CAPSA) – Provided fundraising assistance.

LANGUAGES

Hindi: (Native Language)
English: Speaking, Reading, Writing (Excellent)

TECHNICAL SKILLS

- Project Management – PrimaVera, MS Project
- Spatial Analysis Applications: ARCGIS, ERDAS Imagine.
- Water Resources Engineering Applications: HEC-1/HEC HMS, HEC-RAS, DAMBRK, HEC-ResSim , SWMM, TOPMODEL; VENSIM; WaterCAD, H2ONet; Waterham (Water Hammer analysis software package)
- Programming Languages: R, Matlab.
- Design Software: AutoCAD, STAAD III.