8-12-2016

Methodology to Predict Daily Groundwater Levels by the Implementation of Machine Learning and Crop Models

Sandra Milena Guzman Gutierrez

Methodology to predict daily groundwater levels by the implementation of machine learning and crop models

By

Sandra Milena Guzman Gutierrez

A Dissertation
Submitted to the Faculty of Engineering
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Biological Engineering
in the Department of Agricultural & Biological Engineering

Mississippi State, Mississippi

August 2016

Methodology to predict daily groundwater levels by the implementation of machine learning and crop models

By

Sandra Milena Guzman Gutierrez

Approved:

Joel O. Paz
(Major Professor)

Mary Love M. Tagert
(Committee Member)

Jonathan W. Pote
(Committee Member)

Andrew E. Mercer
(Committee Member)

Fei Yu
(Graduate Coordinator)

Jason M. Keith
Dean
College of Engineering

Name: Sandra Milena Guzman Gutierrez

Date of Degree: August 12, 2016

Institution: Mississippi State University

Major Field: Biological Engineering

Major Professor: Joel O. Paz

Title of Study: Methodology to predict daily groundwater levels by the implementation of machine learning and crop models

Pages in Study 127

Candidate for Degree of Doctor of Philosophy

The continuous decline of groundwater levels caused by variations in climatic conditions and crop water demands is an increased concern for the agricultural community. It is necessary to understand the factors that control these changes in groundwater levels so that we can better address declines and develop improved conservation practices that will lead to a more sustainable use of water. In this study, two machine learning techniques namely support vector regression (SVR) and the nonlinear autoregressive with exogenous inputs (NARX) neural network were implemented to predict daily groundwater levels in a well located in the Mississippi Delta Region (MDR). Results of the NARX model indicate that a Bayesian regularization algorithm with two hidden nodes and 100 time delays was the best architecture to forecast groundwater levels. In another study, the SVR and the NARX model were compared for the prediction of groundwater withdrawal and recharge periods separately. Results from this study showed that input data classified by seasons lead to incremental improvements in the model accuracy, and that the SVR was the most efficient machine learning model with a Mean Squared Error (MSE) of 0.00123 m for the withdrawal season. Analysis of

input variables such as previous daily groundwater levels (Gw), precipitation (Pr), and evapotranspiration (ET) showed that the combination of Gw+Pr provides the optimal set for groundwater prediction and that ET degraded the modeling performance, especially during recharge seasons. Finally, the CROPGRO-Soybean crop model was used to simulate the impacts of different volumes of irrigation on the crop height and yield, and to generate the daily irrigation requirements for soybean crops in the MDR. Four irrigation threshold scenarios (20%, 40%, 50% and 60%) were obtained from the CROGRO-Soybean model and used as inputs in the SVR to evaluate the predicted response of daily groundwater levels to different irrigation demands. This study demonstrated that conservative irrigation management, by selecting a low irrigation threshold, can provide good yields comparable to what is produced by a high volume irrigation management practice. Thus, lower irrigation volumes can have a big impact on decreasing the amount of groundwater withdrawals, while still maintaining comparable yields.

DEDICATION

I dedicate this work to God and to Mother Nature who walked me through this exciting journey. To Paolita who is my little angel and always take care of me. To my lovely parents, Gerardo Guzmán and Nubia Gutiérrez, who tirelessly worked day and night to give us the opportunities they did not have. I am who I am because of you, thank you for believing in me and for reminding me that there is always time for everything. To my sister Alejandra (baby) and my brother Cristian who are two of the most beautiful gifts I have received. To my beloved Rene who has been my coach during this time. Your love, support, encouraging words, and the nice academic discussions we had were my source of energy when I felt low. I look forward to sharing more adventures with you. And to the memory of my dear uncle whom I left back home and never thought I will never see again “lo voy a extrañar tio”.

Finally to you reader, thanks for spend your time reading my work. Thank you and muchas gracias!

ACKNOWLEDGEMENTS

I would like to thank my mentor and friend Dr. Joel Paz. Thanks Dr. Paz for sharing your skills and expertise with me and for giving me the opportunity to work with you. It has been a great pleasure and honor, and I can only hope to live up to the example you have set for me, and to my Co – Major professor Dr. Mary Love Tagert, who has been always willing to help me in my research, and helped me to improve the quality of my work. Also to my committee members, Dr. Jonathan Pote who has offered me his wisdom, support and friendship along the way, and to Dr. Andrew Mercer for his continuous advising and training in the implementation of support vector machines, and for that energy that just made me run after each class or meeting to investigate every equation or theory that I use. I truly feel inspired.

There have been many people who walked alongside me since I came to the US four years ago. They have guided me, opened new opportunities for me to keep dreaming, and showed me the path to go further and push harder. I would like to extend special thanks to Dr. Jairo Diaz. Mil y mil gracias, it is an honor to say I am a friend of such a complete person, to Dr. Tim Schauwecker I don't have words to thank you enough for receiving me and opening the doors of MSU to follow my dreams, and meet my dear advisor Dr. Joel Paz. To Dr. James Martin who has been the best father-in-law I could ask for. To the Colombian crew: Germania and Jairo, Gina, Erick and Sergio (the adopted Colombians), Juan, and Laurens who were my family outside Colombia. Thanks for being there when I needed the most. To my fellow grad students and in particular, Juan Perez-Gutierrez, with whom I always shared a cup of Colombian coffee, and discussed over and over research problems and new ideas with me. To all the beautiful people I have met in this last four years, every talk, every correction, every proofreading, every cup of coffee and cheering words helped me improve my English communication skills and make me be a better researcher, thank you so much.

Finally, I would like to thank the funding provided by the National Institute of Food and Agriculture, U.S. Department of Agriculture, Hatch project under accession number 227246, the MAFES Director's Fellowship program, and the support of the Bagley College of Engineering.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Mississippi River Valley Alluvial Aquifer (MRVA) is the major source of water for irrigation in the Mississippi Delta Region (MDR), and it is the second most heavily pumped aquifer for agriculture (Arthur, 2001). Groundwater from this aquifer is the primary source of water for irrigation of corn, cotton, soybeans and rice, and also for catfish production (Wax et al., 2009). The region receives between 1000 and 1500 mm of precipitation annually (NCDC, 2005), but only five percent of this annual precipitation recharges the aquifer due to the 10-20 m thick impermeable silt-clay soil layer that covers most of the MDR (Barlow and Clark, 2011). In addition, most of the precipitation falls from September to April, outside the critical time for crop production. As a result, irrigation is necessary to satisfy the plant water requirements and to maintain crop production.

Despite the humid subtropical climate and abundant precipitation in the MDR, agricultural production in the region could potentially be affected by the continuous declining groundwater levels. Irrigation is required to meet the water demand of 70% of cropland in the MDR as well as the projected demand of increasing crop acreage in the region (NASS, 2003). The rapid groundwater level declines are currently a major concern for farmers and stakeholders in the region. Soybean, which is one of the top commodities and has the highest crop acreage in the MDR, could be affected by a shortage of groundwater. Thus, developing strategies for the efficient management of groundwater for irrigation and understanding the impacts of different pumping rates on the sustainability of the aquifer will generate solutions and help establish groundwater conservation plans to better maintain high agricultural productivity in the region.

To expand the analysis of the complex relationships between crop production and groundwater sustainability, it is necessary to obtain reliable information regarding the crop water demands based on soil, environmental and plant physiological conditions, and also to determine the impact of those water demands at local and regional scales. However, the monitoring and collection of field data is difficult and expensive. In addition, groundwater fluctuations are complex and dynamic in response to surface-groundwater hydrologic interactions, and interfacing aquifer water exchanges. This makes it difficult to directly assess, based on observations alone, the hydrologic interaction between groundwater and surface water and the aquifer responses to water extractions for crop production. The development of innovative modeling approaches that are able to provide an assessment of irrigation demands and groundwater supply at a local scale is then required.

Crop models are a potential solution to evaluate crop water requirements in the region. These models evaluate the crop physiological and yield variability from different sets of proposed environmental and management conditions, while also providing an estimate of the volume of water required for each management scenario.

In addition, machine learning techniques are a subfield of artificial intelligence that focuses on the use of data-driven algorithms for the extraction of patterns and information from complex datasets. In the field of hydrology, these algorithms have gained popularity for their efficiency in predicting water resource variables without the complete knowledge of all the underlying physical processes (Maier and Dandy, 2000; Raghavendra and Deka, 2014). Groundwater fluctuations are random and dynamic in its natural structure. Machine learning techniques such as artificial neural network (ANN) and support vector regression (SVR) can provide predictions of how the subsurface system responds to different environmental and management conditions and generate critical information that can be used by decision makers.

Although several studies demonstrate the efficiency of crop models for irrigation management and machine learning for groundwater modeling, there are no studies that integrate the crop irrigation requirements and its effects on the groundwater level variability at a local scale. Furthermore, a linked crop model – machine learning approach is necessary to better understand and predict the sustainability of the subsurface system. The general objective of this dissertation was to implement a novel methodology for the evaluation of daily groundwater levels at a local scale.

Dissertation organization

This dissertation is a compilation of journal manuscripts submitted or intended for submission to refereed scientific journals. Each manuscript addresses a specific objective. Chapter 2 outlines an implementation of a nonlinear autoregressive with exogenous inputs (NARX) neural network to forecast daily groundwater levels. Chapter 3 presents an evaluation of NARX and SVR to identify the most reliable method for the analysis of daily groundwater levels for recharge and withdrawal periods based on input time series divided by seasons. Chapter 4 examines the predictive performance of an SVR model using different arrangements of input variables related to the groundwater system.

Chapter 5 presents the application of a linked crop-SVR model to analyze the impacts of various irrigation thresholds on crop yield and daily changes in groundwater levels. Finally, Chapter 6 presents a summary of all the major conclusions found in each of the journal manuscripts.

References

Arthur, J.K., 2001. Hydrogeology, model description, and flow analysis of the Mississippi River alluvial aquifer in northwestern Mississippi. 2001-4035.

Barlow, J.R.B., Clark, B.R., 2011. Simulation of Water-Use Conservation Scenarios for the Mississippi Delta Using an Existing Regional Groundwater Flow Model. 5019, USGS.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software, 15(1): 101-124. DOI:http://dx.doi.org/10.1016/S1364-8152(99)00007-9

NASS, U., 2003. United States Department of Agriculture-National Agricultural Statistics Service. Cropland Data Layer.

NCDC, 2005. National Climatic Data Center.

Raghavendra, N.j., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Applied Soft Computing, 19(0): 372-386. DOI:http://dx.doi.org/10.1016/j.asoc.2014.02.002

Wax, C.L., Pote Jonathan W, L., M.T., 2009. Climatological and Cultural Influences on Annual Groundwater Decline in the Mississippi Delta Shallow Alluvial Aquifer, Mississippi Water Resources Conference, Jackson. Mississippi.

CHAPTER II

THE USE OF NARX NEURAL NETWORKS TO FORECAST DAILY GROUNDWATER LEVELS

A paper submitted to the Water Resources Management Journal

Sandra M. Guzman, Joel O. Paz, and Mary Love M. Tagert

Abstract

The lack of information for predicting groundwater levels at farm level makes it difficult to assess models in locations where availability of data is limited. Artificial neural networks (ANN) are modeling tools used to predict groundwater dynamic processes requiring a reduced number of inputs and parameters to generate efficient predictions. We present a novel application of a nonlinear autoregressive with exogenous inputs (NARX) recurrent neural network (RNN) to simulate daily groundwater levels for a well in the Mississippi River Valley Alluvial (MRVA) aquifer in the southeast US. The network was trained using Levenberg-Marquardt (LM) and Bayesian Regularization (BR) algorithms, and the results were compared to identify an optimal ANN architecture for the forecasting of daily groundwater levels over time. The training algorithms were implemented using different hidden layer combinations and delays (5, 25, 50, 75, and 100) until the optimal network was found. Eight years of daily precipitation and groundwater level data were used as inputs to the network. The results showed that BR with two hidden nodes and 100 time delays provided the most accurate prediction of

groundwater levels with an error of $\pm$ 0.00119 m. The comparison between LM and BR showed that NARX-BR is superior in forecasting daily levels based on the measurement of statistical performance with coefficients of determination higher than 0.80 for periods up to three months. The study of new models and empirical methods for predicting groundwater levels is important for mitigating risks for farmers, understanding dynamic groundwater processes, and implementing conservation practices and water use policies.

Introduction

The sustainability of water for irrigation in the southeastern United States relies primarily on the availability of groundwater resources (Hook et al. 2009). A warmer climate and an increased food demand have affected the availability of water for agricultural production in the last few decades (Ojha et al. 2013). The most recent drought reports around the nation show a growing necessity for information to measure the effects of reduced irrigation on crop yield and to develop solutions that will result in a sustainable groundwater supply (Asefa et al. 2007; Beigi and Tsai 2015; Kişi 2007; Kong-A-Siou et al. 2015; Konikow 2015; Kovacs et al. 2015). An understanding of the interaction between groundwater and other freshwater systems is required to provide a comprehensive assessment of the actual groundwater conditions. The variation of groundwater levels at farm scale affects directly the availability of water for irrigation in the crop seasons, thus short time reliable information is necessary to understand these groundwater punctual systems, especially on farms located in regions where groundwater sources are critical such as the Mississippi Delta. Governmental agencies such as the United States Geological Service (USGS) and the Yazoo Mississippi Delta Joint Water Management District (YMD) provide annual and seasonal information for groundwater levels in certain areas of Mississippi. However, this information is collected for a limited number of wells, and it is insufficient for studies at a short time scale. Thus, the use of models is a useful option when the availability of data is a constraint. Physically-based models are the main methods used to predict water resource variables, but they have practical limitations. For the farmers and decision makers, the capacity to provide an accurate and fast prediction is often more important than the understanding of the physical basis behind the model, especially when information is limited. Data-based models (DBMs) are particularly useful in these cases when it is not possible to fully describe all of the physical and mathematical formulations behind the processes and the computational sources are limited (Daliakopoulos et al. 2005; Tóth 1963). One of the most recognized DBMs in hydrology is artificial neural network. ANNs are capable of representing nonlinear systems with fewer initial conditions in a "black box" environment. There are different types of ANN structures for the time series prediction of hydrologic variables: feed–forward networks, RNNs, radial basis function (RBF), back–propagation, conjugate gradient, cascade correlation, input delay (IDNN), etc. (Anderson 2005; Jain et al. 1996; Jayawardena 2014; Zhang et al. 2003).

Several studies have examined the capabilities of diverse ANN structures in the estimation of groundwater levels. Coulibaly et al. (2001) evaluated three ANN models using limited historical records for groundwater level to predict monthly groundwater levels of shallow and deep wells in Gondo Plain, Burkina Faso. The analysis of RNN, Input Delay Neural Network (IDNN) and radial basis function (RBF) network showed that although all networks have good predictions, RNN provided the most robust architecture for the simulation of groundwater levels. Nayak et al. (2006) studied an

ANN model to forecast monthly groundwater levels in a shallow aquifer in Godavari, India and found that a feed-forward ANN was able to forecast monthly levels between two and four months ahead. Asefa et al. (2007) examined three types of training algorithms for the prediction of field-scale groundwater levels and found that feed-forward backpropagation was one of the most efficient predictors based on forecasting lead time accuracy despite of the difficulty to implement the method and the time requirements. They concluded that training algorithms should be evaluated in terms of efficiency of the prediction and training time, especially when computational resources are not available.

ANN models have been used in the prediction of rainfall (French et al. 1992), evapotranspiration (Kumar et al. 2002), river flow (Jayawardena and Fernando 2001), water quality (Maier and Dandy 1996; Sarkar and Pandey 2015), and other water resource variables (Bowden et al. 2005; Dawson and Wilby 2001; Kong-A-Siou et al. 2015; Pandey and Srinivas 2015). Although several structures and training methods have been applied previously, feed-forward networks are the most common approach used for the estimation of water resource variables (Maier and Dandy 2000). The correct selection of the ANN type and training architecture is fundamental to properly configure the model structure. In the field of hydrology, some of the most promising ANN training functions have not been studied and compared with the traditional feed-forward networks. The American Society of Civil Engineers (ASCE) task committee on neural networks (Govindaraju 2000a; Govindaraju 2000b) analyzed the role of ANNs in hydrology and water resources showing the process, strengths, limitations and applications for hydrologists and water experts. There is still some ANN modeling approaches that need to be evaluated to determine optimal procedures for the prediction of surface and groundwater variables over time.

The non–linear autoregressive with exogenous inputs (NARX) modeling approach is a special type of recurrent, dynamic ANN that describes the modeled process based on lagged input - output variables and prediction errors. NARX became popular in the last years for its good performance in the prediction of time series that has a seasonality component. It has been shown that NARX networks can provide optimal predictions without computational losses in comparison with the conventional RNNs, and that they are at least as powerful as Turing machines (Siegelmann et al. 1997; Tsung-Nan et al. 1997). The embedded memory of NARX networks provides shorter paths to optimally propagate the information and backpropagate the error signal, reducing the model's long–term dependencies. NARX neural networks have been applied in the prediction of time series (Diaconescu 2008; Siegelmann et al. 1997), with a powerful performance that can approximate almost every non–linear function.

Recent studies that focused on the application of neural networks for determining changes in water quality concentration (Chang et al. 2015), water yield (Gharun et al. 2015), sediment concentration (Singh and Chakrapani 2015), and flood levels (Ruslan et al. 2013) have shown the advantages of the NARX model for hydrological applications. However, the use of this modeling network has not been explored in the forecasting of daily groundwater levels. In this paper, we present a novel implementation of NARX networks, including the evaluation of the Bayesian Regularization (BR) and Levenberg-Marquardt (LM) training algorithms, to determine the most efficient training architecture for the forecasting of daily groundwater levels in a farm well located in northwest

Mississippi. The application of models such as a NARX network is important for reliable forecasting of daily values at a small spatial scale, and for the management of groundwater resources in the Mississippi Delta region.

Materials and Methods

Study Area

The study area is part of the MRVA aquifer that covers roughly 82,800 square kilometers in the states of Arkansas, Kentucky, Louisiana, Tennessee, Missouri, Illinois and Mississippi. This aquifer is the primary source of groundwater for irrigation in the region with withdrawals around 406.94 $m^3 s^{-1}$ (Maupin and Barber 2005). There is a "cone of depression" in the portion of the MRVA covering the central MS Delta, and this area is characterized its substantial groundwater level declines and its concurrent high crop production levels. Annual precipitation in the region ranges between 1,143 to 1,448 mm per year, distributed mostly in winter and spring seasons (Snipes et al. 2005). However, only 5% of the total annual rainfall has the potential to recharge the aquifer due to the reduced soil infiltration capacity and the fine soil layer overlaying the superficial portion of the aquifer (Arthur 2001; Barlow and Clark 2011; Welch et al. 2011).

For this study, groundwater level data was obtained from a U.S. Geological Survey (USGS) well (USGS M0038) located between 33º 28' 25" latitude and 90º 44' 22" longitude in part of the cone of depression in Sunflower County (Figure 2.1). This well was chosen from a limited number of wells in the Mississippi Delta that have daily groundwater records, because of its location within the cone of depression, and the long length of daily records available. Eight years of daily groundwater level data from 1987 to 1994 were collected from the study well through the USGS National Water

Information System (NWIS) website (USGS 2001; USGS 2014). In addition, daily mean precipitation data were obtained from a weather station, located 32 km east of the groundwater well in Moorhead, Mississippi (Menne et al. 2012).

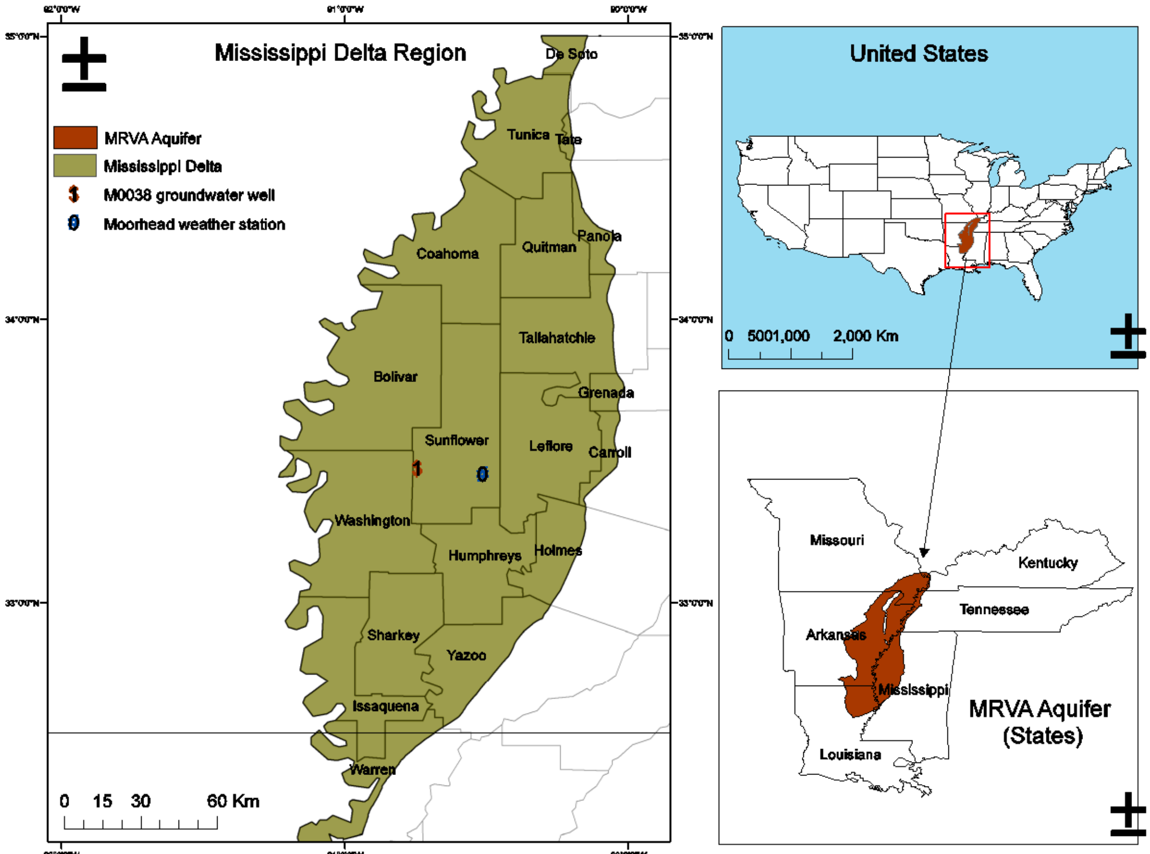

Figure 2.1 Study area and Sunflower well location

The last week of February and the first week of March in 1991 had incomplete groundwater data records, and therefore this gap was used to partition the data for training (March 1987- February 1991) and testing (April 1991- June 1994).

Normalization

For this study, the input data were selected to represent the climatic and groundwater conditions of the region. Data was partitioned in two sets: the first five years for calibration and the next three years for validation following the conventional method as suggested by Asefa et al. (2007) and Lohani and Krishan (Lohani and Krishan 2015). Preprocessing was performed to clean and manage the input and feedback data series for the training step. A normalization function was applied to the input values, whereby the data were transformed into a common range to better fit the training process. Input and feedback values were normalized between -1 to 1 range, and back transformed in the testing step.

NARX Model Architecture

The nonlinear autoregressive network with exogenous inputs (NARX) is a dynamic recurrent neural network (RNN) used in time series prediction with efficient results and demonstrated capabilities in finding long time patterns (Seidl and Lorenz 1991; Siegelmann et al. 1997). The common definition for the NARX model is given by:

$$y(t) = f\big(y(t-1), y(t-2), \dots, y(t-n_y), u(t-1), u(t-2), \dots$$
$$\dots, u(t-n_u)\big) \qquad (2.1)$$

where $u(t)$ and $y(t)$ represent the inputs and outputs of the network at a discrete time step t, respectively, n_u and n_y are the input and output layers of the network, H is the hidden layer and f is a nonlinear function. The outputs, $y(t)$, are regressed onto previous values of the independent or exogenous input signal, improving the convergence time of the network. The description of the mapping function, f, used in this study is shown in Figure 2.2.

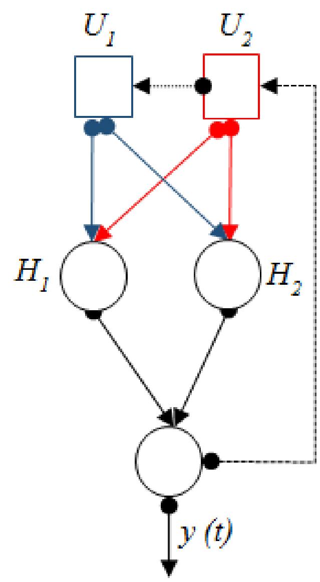

Figure 2.2 A NARX network with $n_u = n_y = 2$ and H = 2

First, the most efficient configuration to train the network was determined, to implement the NARX model. During the training of the network the true output, or measured value, is used as the regressive input in the network. This is represented by Figure 2.3a. Once the network is trained and used for prediction purposes, the calculated output is feedback to the network to obtain the estimation for the next prediction step. This loop is shown in Figure 2.3b. The configurations shown in Figure 2.3a and Figure 2.3b are commonly called series-parallel and parallel, respectively. For this study, a series-parallel architecture was implemented in the training step to include the true output forward in the network, and the parallel configuration was used for the multi-step ahead prediction. The use of a series-parallel architecture optimizes the training by reducing the iteration time. Typically, training comprises a specific number of hidden neurons and randomly selected values for the weights with fixed connections. These weights can make the connection between internal nodes stronger or weaker. For example, if the weight is zero, then there is no connection. In this study, the weighting process started as random, and it was calibrated along with the iteration in the series-parallel architecture.

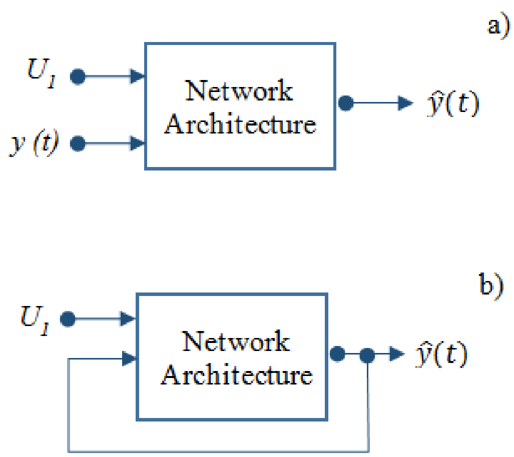

Figure 2.3 a) Series-parallel and, b) parallel architecture to train NARX networks

Training Algorithms

Levenberg – Marquardt (LM)

Levenberg – Marquardt is one of the most widely used functions for time series network prediction and training (Adeloye and De Munari 2006; Hagan and Menhaj 1994; Khaki et al. 2015; Kişi 2007). This method is a variation of the Gauss Newton algorithm that finds the function minima and optimizes the solution. It uses an approximation of the Hessian matrix as given below (adapted from Sahoo and Jha 2013):

$$\Delta w = [J^T J + \mu I]^{-1} J^T(w) e(w) \tag{2.2}$$

where w is the weight, $J^T J$ represents the Hessian matrix, J is the jacobian matrix, μ is the learning constant, J^T is the transpose of J, I is the identity matrix and e represents the vector of errors. The learning constant μ is adjusted based on the error in each iteration, to find the minima. In this study, the iteration process started with a random μ for the weight optimization with the LM algorithm. The function *trainlm* from MATLAB® was used to train the network with the LM method.

Bayesian Regularization (BR)

The Bayesian Regularization training function is a method used to reduce the negative effects of large weights in the training process. Regularization reduces the probability of overfitting the model by setting the optimal performance function to provide an efficient generalization based on Bayesian inference techniques (Foresee and Hagan 1997). The computation of the Hessian matrix is required to find the optimal regularization parameters in the BR function. A Gauss-Newton approximation of the Hessian matrix $J^T J$ is applied, following David MacKay's Bayesian techniques (MacKay 1992) to optimize regularization. For this study, the function *trainbr* that is part of the Neural Network Toolbox™ in MATLAB® 2014a was used to train the NARX model with the BR method.

Network Architecture

To identify the optimum NARX architecture, we used the common trial-and-error method to select the number of hidden nodes and the transfer function as outlined by Maier and Dandy (Maier and Dandy 2000). In this study, we found that two hidden nodes with a sigmoid transfer function and a single output node with linear function provided the most effective network architecture. The number of hidden nodes was evaluated for its ability to generate accurate model responses. The use of more than one hidden node is commonly used for the approximation of complex functions. However, the selection of the number of nodes depends on each case study. The sigmoid–linear transfer function combination has the ability to provide an efficient mathematical representation of the output as a function of the input signal. Maier and Dandy (Maier and Dandy 2000) suggest that the transfer function represents a considerable difference in learning speed and weight. Figure 2.4 shows a flow diagram of the steps implemented in this study to determine the network architecture and train the network.

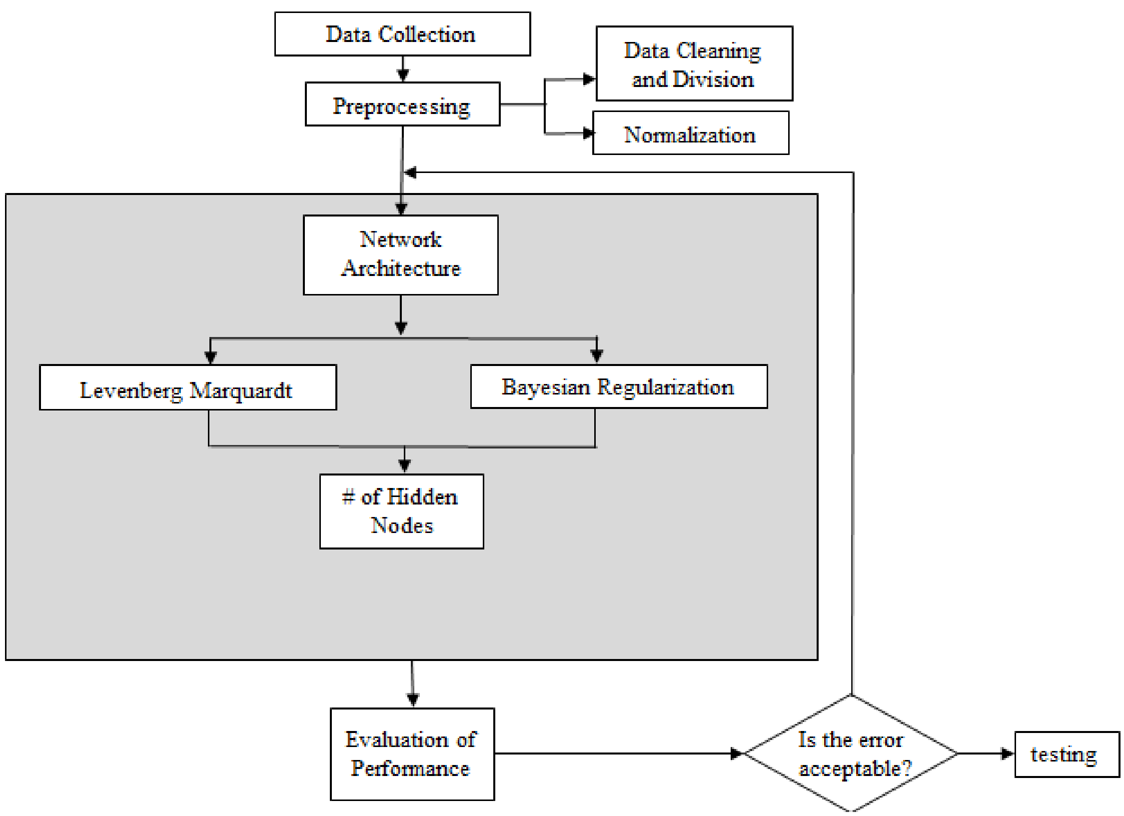

Figure 2.4 Diagram of the steps involved in the NARX network training process

Evaluation of Performance

The prediction performance of the two networks trained with LM and BR functions were compared using statistical equations of goodness of fit. The Mean Squared Error (MSE) evaluates the difference between observed and predicted values by:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2.3}$$

where y_i represents observed values, $\hat{y}_i$ is the predicted value and N equals the number of values. The lowest MSE provide the best prediction performance. Additionally, the coefficient of determination, R^2, was used to analyze the best linear fit between observed and predicted values. The best model fit is provided by an R^2 coefficient closest to 1. R^2 is given by:

$$R^2 = \frac{\sum(y_i - \hat{y}_i)^2}{\sum y_i^2 - \frac{\sum \hat{y}_i^2}{n}} \tag{2.4}$$

Results and Discussion

Input Structure

Figure 2.5 shows the input time series used for training and testing the network. Daily groundwater fluctuation follows a sinusoidal distribution with peaks in March – April and pronounced decreases in water depth between May – August due to seasonal pumping of groundwater for irrigation use during the growing season. Groundwater withdrawals were higher than the recharge capacity of the well. The annual recharge is noticeably lower over time. The minimum recharge occurred in 1994 compared to the peak in 1987, showing how the consumption of groundwater for irrigation is increasing over time. Although precipitation is uniformly distributed over the years, the volumes are not sufficient to maintain the aquifer levels and provide a sustainable resource for irrigation. The demand for irrigation will increase in the near future due to crop expansion and climatic variability (Cathcart et al. 2007; Kebede et al. 2014).

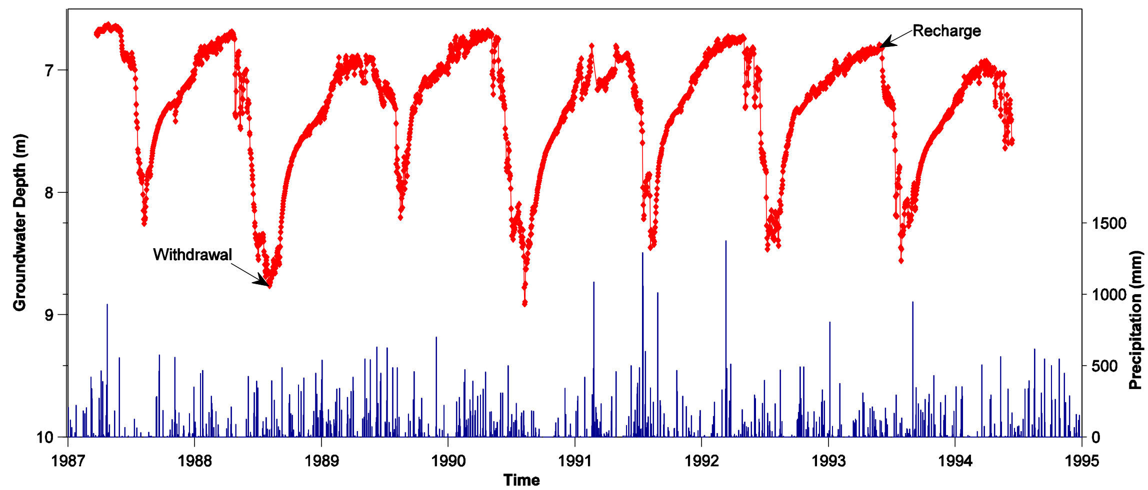

Figure 2.5 Daily precipitation and groundwater level series

Two input layer scenarios, namely groundwater and groundwater plus precipitation, were examined to identify which input variables had more effect on the network response. For this study, the scenario with an input variable combination of groundwater and precipitation provided the most accurate training performance. The results of the different scenario trials are not presented in the paper as they are not significant for the scope of this study.

Training and Testing

The comparison between BR and LM was performed with the same procedures and programming code developed in MATLAB® 2014a. In this study, two hidden layers were defined as optimal for the network. The use of two hidden layers was also reported by Coulibaly et al. (2001). The sigmoid activation function was used for the hidden layers, while the linear function was used for the processes in the output layer. Autocorrelation functions helped identify the training architecture that provided the highest model performance. An array of 5, 25, 50, 75 and 100 input time delays was evaluated to determine the best model architecture. Time delays of 50, 75, and 100 produced the three best architecture performances based on the results of the autocorrelation function (Figure 2.6). For a perfect training fit, the lagged points should be zero over the entire function, which means there is no prediction error. In general, the performance of BR was better than LM in all cases. The BR model learning performance improved as the number of delays increased, except for 50 time delays (Figure 2.6). Although the autocorrelation values did not significantly differ from zero, the MSE with 50 delays was higher than the observed at 100 delays, as is shown in Table 2.1. From these results, BR with 100 time delays was selected as the optimal training function for estimating groundwater fluctuations in the USGS monitoring well, as it provided the best combination of autocorrelation–MSE outputs.

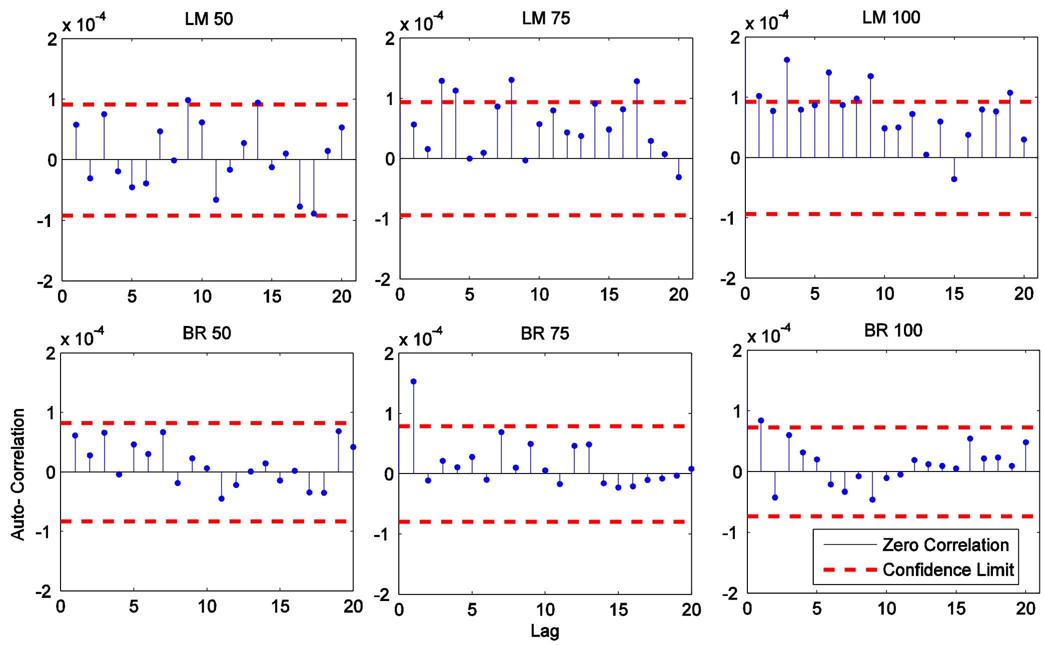

Figure 2.6 Autocorrelation of errors function for the daily groundwater training set

The MSE and computational speed for each number of delays and training architectures are summarized in Table 2.1. The differences between training algorithms were more evident as the number of delays increased, similar to the autocorrelation responses. The number of iterations in LM was less than BR, with a faster convergence. In terms of model performance, BR showed higher predictive capabilities. This training algorithm provides the most optimal prediction if the computational resources are available. However, LM is a good alternative and provides acceptable estimates in cases where the resources are not sufficient or the model management is more suitable with this method. The results for the two training algorithms were very promising, with a difference overall of less than 0.003 m between observed and predicted groundwater levels (Table 2.1). The advantage of NARX models is that they optimize the time performance in comparison with other neural network algorithms (Siegelmann et al. 1997). Thus, the use of BR for this study provides the fastest convergence (in minutes) if compared with other studies in the same field.

Table 2.1 Mean square error (MSE) and number of iterations for the evaluation of model performance

	LM		BR	
# of Delays	MSE	# of Iterations	MSE	# of Iterations
100	0.0194	38	0.00119	793
75	0.0272	15	0.00117	166
50	0.00188	44	0.00138	1000
25	0.00171	20	0.00144	203
5	0.002014	25	0.00174	185

Figure 2.7 shows the predicted time series for a NARX-RNN with two hidden layers, BR training algorithm and 100 time delays. The model was able to efficiently predict daily groundwater levels and the variability between withdrawal and recharge periods for a lead time up to three months. This study focuses on daily groundwater variations and on the availability of water for irrigation. The model efficiently predicted the yearly peaks of demand and the effect of precipitation in the recharge process.

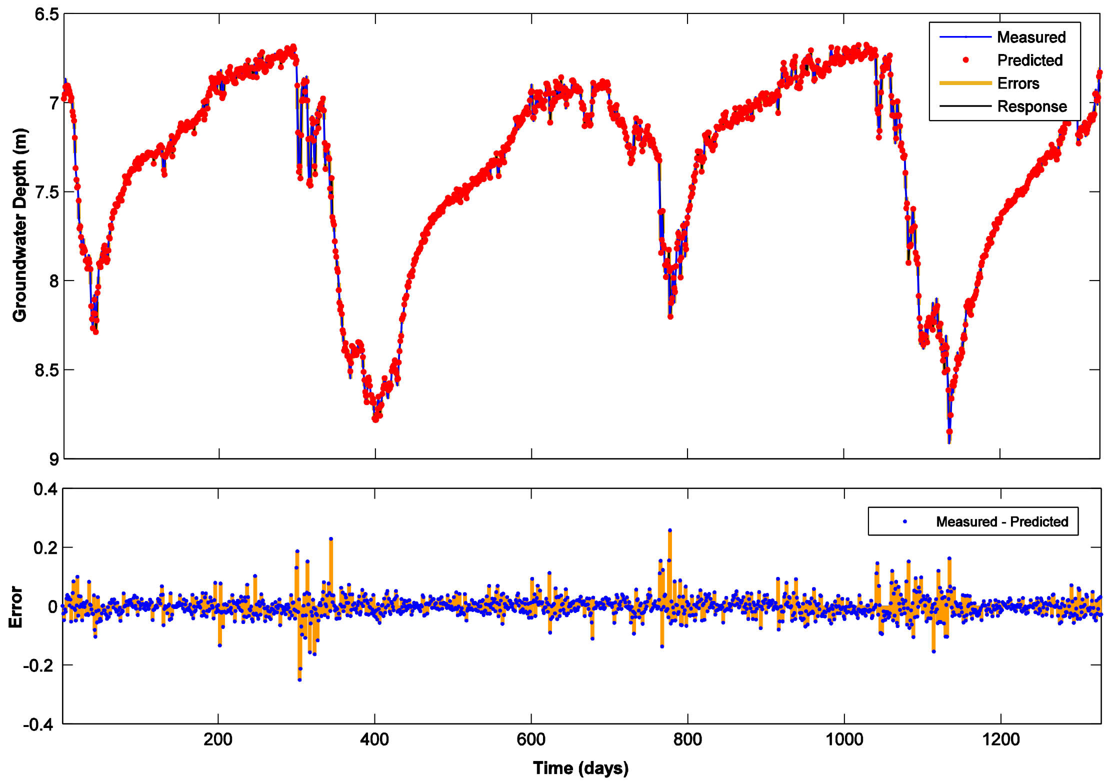

Figure 2.7 Time series prediction using Bayesian Regularization with 100 time delays

For the testing phase, additional values were introduced to the network to evaluate the differences between predicted and observed daily levels. The statistics of linear relation between observed and predicted values showed good responses as indicated in the forecasting graphs (Figure 2.8) and performance table (Table 2.2). The prediction becomes less accurate over time, based on the MSE results. The best performance is shown for the forecasting at 15 days ahead with a difference less than 0.0013 m between the observed and predicted values. It should be noted that the results shown in the forecasting section were trained with the same network architecture presented in section 3.2. One advantage of analyzing NARX models with daily values is the increased accuracy for the description of localized wells. Neural networks with Bayesian

Regularization have been applied to the prediction of groundwater levels with inputs at seasonal and monthly time intervals (Coulibaly et al. 2001; Daliakopoulos et al. 2005). However, they have not been evaluated for their predictive capabilities on a daily basis. The forecasting of daily groundwater levels at shorter time scales is advantageous for water management at specific wells, and provides useful information to evaluate groundwater plans for irrigation and efficient use of water.

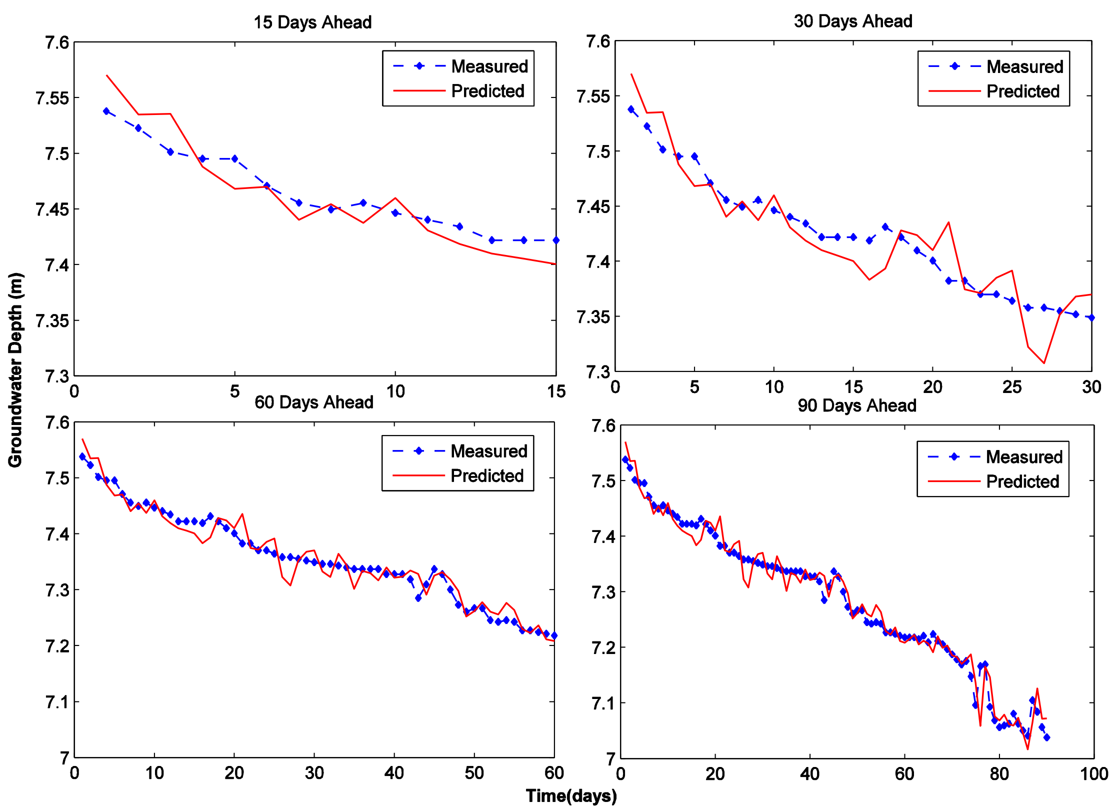

Figure 2.8 Observed and forecasted groundwater depths at various lead times (USGS well M0038)

Table 2.2 Statistical performance of forecasted levels at different lead times

Lead Time (Days)	MSE	R^2
15	0.001262	0.801
30	0.001688	0.828
60	0.001595	0.935
90	0.001763	0.932

Summary and Conclusions

This study demonstrated the capabilities of a NARX neural network in forecasting daily groundwater levels for individual wells in the Mississippi Delta region at various lead times. We identified the most accurate and efficient training algorithm for a NARX RNN, using precipitation and daily groundwater level data from a well in the Delta cone of depression, as inputs to the network. The number of hidden neurons and time delays were varied to optimize the training algorithms. Based on the statistical performance criteria and training results, BR with 100 time delays and 2 hidden layers was the most accurate architecture (MSE = 0.00119) for forecasting groundwater levels up to three months ahead. The LM training algorithm required the least number of iterations for the model convergence. The reduction in time is approximately 12% compared with BR. However, the general prediction performance of BR was more robust.

Neural networks are efficient modeling tools, with proven capabilities for the prediction of different water resource variables based on empirical analysis. The results showed that the use of a NARX network with BR algorithm can be a useful tool for the prediction of daily groundwater level time series despite the presence of strong seasonal trends. However, since this type of network relies heavily on the availability of training data, the prediction depends on the quality of input values provided for the training process. Thus, it is recommended that the network be retrained with more recent data to reflect the constant changes in groundwater processes over time. Future studies will include the analysis of the network performance by the addition of variables such as irrigation requirements and pumping rates.

References

Adeloye A, De Munari A (2006) Artificial neural network based generalized storage–yield–reliability models using the Levenberg–Marquardt algorithm. J Hydrol 326:215-230.

Anderson MG (2005) Encyclopedia of hydrological sciences vol 1. Chichester, West Sussex, England ; Hoboken, NJ : J. Wiley .

Arthur JK (2001) Hydrogeology, model description, and flow analysis of the Mississippi River alluvial aquifer in northwestern Mississippi U.S. Geological Survey Water Resources Investigations Report.

Asefa T, Wanakule N, Adams A (2007) Field-Scale Application of Three Types of Neural Networks to Predict Ground-Water Levels. JAWRA J Am Water Resour As 43:1245-1256

Barlow JRB, Clark BR (2011) Simulation of Water-Use Conservation Scenarios for the Mississippi Delta Using an Existing Regional Groundwater Flow Model: U.S. Geological Survey Scientific Investigations Report, 2011–5019

Beigi E, Tsai FC (2015) Comparative study of climate-change scenarios on groundwater recharge, southwestern Mississippi and southeastern Louisiana, USA. Hydrogeol J 23:789-806 doi:10.1007/s10040-014-1228-8

Bowden GJ, Dandy GC, Maier HR (2005) Input determination for neural network models in water resources applications. Part 1—background and methodology. J Hydrol 301:75-92

Cathcart TP, Wax CL, Pote JW, Triyono S (2007) A climatological basis for conserving groundwater and reducing overflow in aquaculture ponds in the Southeast United States. Aquacult eng 36:225-232

Chang F-J, Tsai Y-H, Chen P-A, Coynel A, Vachaud G (2015) Modeling water quality in an urban river using hydrological factors–Data driven approaches. Journal of environ manage 151:87-96

Coulibaly P, Anctil F, Aravena R, Bobée B (2001) Artificial neural network modeling of water table depth fluctuations .Water Resour Res 37:885-896 doi:10.1029/2000wr900368

Daliakopoulos IN, Coulibaly P, Tsanis IK (2005) Groundwater level forecasting using artificial neural networks. J Hydrol. 309:229-240 doi:http://dx.doi.org/10.1016/j.jhydrol.2004.12.001

Dawson C, Wilby R (2001) Hydrological modelling using artificial neural networks. Progress in physical Geography 25:80-108

Diaconescu E (2008) The use of NARX neural networks to predict chaotic time series WSEAS. Trans Comp Res 3:182-191

Foresee FD, Hagan MT (1997) Gauss-Newton approximation to Bayesian learning. In: International Conference on Neural Networks. IEEE, pp 1930-1935

French MN, Krajewski WF, Cuykendall RR (1992) Rainfall forecasting in space and time using a neural network. J Hydrol 137:1-31

Gharun M, Azmi M, Adams MA (2015) Short-Term Forecasting of Water Yield from Forested Catchments after Bushfire: A Case Study from Southeast Australia. Water 7:599-614

Govindaraju RS (2000a) Artificial neural networks in hydrology. I: Preliminary concepts. J Hydrol Eng 5:115-123

Govindaraju RS (2000b) Artificial neural networks in hydrology. II: hydrologic applications. J Hydrol Eng 5:124-137

Hagan MT, Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks 5:989-993

Hook JE, Hoogenboom G, Paz J, Risse M, Bergstrom J, Mullen J Assessing agricultural groundwater needs for the future: Identifying irrigated area and sources. In: Proceedings of the Georgia Water Resources Conference Athens, GA, 2009.

Jain AK, Mao J, Mohiuddin KM (1996) Artificial neural networks: A tutorial. IEEE computer 29:31-44 doi:10.1109/2.485891

Jayawardena A (2014) Environmental and hydrological systems modelling. CRC Press,

Jayawardena A, Fernando T (2001) River flow prediction: an artificial neural network approach. IAHS Publication:239-246

Kebede H, Fisher DK, Sui R, Reddy KN (2014) Irrigation methods and scheduling in the delta region of Mississippi: current status and strategies to improve irrigation efficiency. American Journal of Plant Sciences 5:2917

Khaki M, Yusoff I, Islami N (2015) Application of the Artificial Neural Network and Neuro-fuzzy System for Assessment of Groundwater Quality . CLEAN–Soil, Air, Water 43:551-560

Kişi Ö (2007) Streamflow forecasting using different artificial neural network algorithms. J Hydrol Eng 12:532-539

Kong-A-Siou L, Jourde H, Johannet A (2015) Neural Networks Model as Transparent Box: Toward Extraction of Proxies to Better Assess Karst/River Interactions (Coulazou Catchment, South of France). In: Hydrogeological and Environmental Investigations in Karst Systems. Springer, pp 353-360

Konikow LF (2015) Long-Term Groundwater Depletion in the United States. Groundwater 53:2-9

Kovacs KF, Mancini M, West G (2015) Landscape irrigation management for maintaining an aquifer and economic returns Journal of Environ Manage160:271-282

Kumar M, Raghuwanshi N, Singh R, Wallender W, Pruitt W (2002) Estimating evapotranspiration using artificial neural network. J Irrig Drain E-ASCE 128:224-233

Lohani A, Krishan G (2015) Application of Artificial Neural Network for Groundwater Level Simulation in Amritsar and Gurdaspur Districts of Punjab, India. J Earth Sci Clim Change 6:2

MacKay DJ (1992) Bayesian interpolation. Neural computation 4:415-447

Maier HR, Dandy GC (1996) The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. Water Resour Res 32:1013-1022 doi:10.1029/96WR03529

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Modell Softw 15:101-124 doi:http://dx.doi.org/10.1016/S1364-8152(99)00007-9

Maupin MA, Barber NL (2005) Estimated withdrawals from principal aquifers in the United States, 2000 [electronic resource] / by Molly A. Maupin and Nancy L. Barber. Circular: 1279. Reston, Va. : U.S. Dept. of the Interior, U.S. Geological Survey ; Denver, CO

Menne M et al. (2012) Global Historical Climatology Network-Daily (GHCN-Daily), Version 3. 2 NOAA National Climatic Data Center

Nayak P, Rao YRS, Sudheer KP (2006) Groundwater Level Forecasting in a Shallow Aquifer Using Artificial Neural Network Approach. Water Resour Manage 20:77-90 doi:10.1007/s11269-006-4007-z

Ojha R, Ramadas M, Govindaraju R (2013) Current and Future Challenges in Groundwater. I: Modeling and Management of Resources. J Hydrol Eng 20:A4014007 doi:10.1061/(ASCE)HE.1943-5584.0000928

Pandey A, Srinivas V (2015) Use of Data Driven Techniques for Short Lead Time Streamflow Forecasting in Mahanadi Basin. Aquatic Procedia 4:972-978

Ruslan FA, Samad AM, Zain ZM, Adnan R (2013) Flood prediction using NARX neural network and EKF prediction technique: A comparative study. In: IEEE 3rd International Conference on, System Engineering and Technology (ICSET). IEEE, pp 203-208

Sahoo S, Jha MK (2013) Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeol J 21:1865-1887

Sarkar A, Pandey P (2015) River Water Quality Modelling Using Artificial Neural Network Technique. Aquatic Procedia 4:1070-1077

Seidl DR, Lorenz RD (1991) A structure by which a recurrent neural network can approximate a nonlinear dynamic system. In: IJCNN-91-Seattle International Joint Conference on Neural Networks. IEEE, pp 709-714

Siegelmann HT, Horne BG, Giles CL (1997) Computational capabilities of recurrent NARX neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 27:208-215

Singh N, Chakrapani G (2015) ANN modelling of sediment concentration in the dynamic glacial environment of Gangotri in Himalaya. Environmental monitoring and assessment 187:1-14

Snipes CE, Nichols S, Poston D, Walker T, Evans L, Robinson H (2005) Current agricultural practices of the Mississippi Delta.

Tóth J (1963) A theoretical analysis of groundwater flow in small drainage basins. Journal of Geophysical Research 68:4795-4812 doi:10.1029/JZ068i016p04795

Tsung-Nan L, Giles CL, Horne BG, Sun-Yuan K (1997) A delay damage model selection algorithm for NARX .IEEE Transactions on neural networks Signal Processing 45:2719-2730 doi:10.1109/78.650098

USGS USGS (2001) National Water Information System data available on the World Wide Web (Water Data for the Nation). http://waterdata.usgs.gov/nwis/. Accessed Jun 15 2013

National Water Information System: Web Interface (2014) http://waterdata.usgs.gov/nwis/gw. Accessed June/15/2014

Welch HL, Green CT, Coupe RH (2011) The fate and transport of nitrate in shallow groundwater in northwestern Mississippi, USA. Hydrogeol J 19:1239-1252

Zhang Q-J, Gupta KC, Devabhaktuni VK (2003) Artificial neural networks for RF and microwave design-from theory to practice Microwave Theory and Techniques, IEEE Transactions on 51:1339-1350 doi:10.1109/TMTT.2003.809179

CHAPTER III

EVALUATION OF SEASONALLY DIVIDED INPUTS FOR THE PREDICTION OF DAILY GROUNDWATER LEVELS: ARTIFICIAL NEURAL NETWORKS VS SUPPORT VECTOR MACHINES

A paper submitted to the Journal of Hydrology
Sandra M. Guzman, Joel O. Paz, Mary Love M. Tagert, and Mercer A.

Abstract

Farmers and stakeholders who use groundwater for irrigation need efficient and cost-effective techniques to sustain their use. Especially as the demand for water continues to increase, farmers need better and more current information about the variability of groundwater levels in their wells. However, the prediction of groundwater levels is difficult and very dynamic under traditional modeling approaches, and manual monitoring of individual wells is costly and time-consuming. We have studied two machine learning models for predicting daily groundwater levels by comparing the nonlinear autoregressive with exogenous inputs (NARX) artificial neural network (ANN), and support vector regression (SVR) for an irrigation well located in a highly productive agricultural region in the southeastern United States. Multiple years of daily input time series were preprocessed and divided between the summer and winter seasons to predict withdrawal and recharge periods separately. The results show that SVR has a better modeling performance based on its mean squared error (MSE) and prediction

trend. In addition, it has demonstrated that the prediction of daily levels with input time series divided by seasons provides higher accuracy than yearly input time series with all seasons included. Results also indicate that the recharge season becomes a linear problem, which reduces the SVR modeling computational requirements. The application of these data driven models for the management of water resources for irrigation provides important information for decision making and development of future regional scale analysis.

Introduction

The most recent advances in technology and information processes have provided new techniques for the analysis of complex natural systems. In the field of hydrology, the use of data based models (DBM's) such as artificial neural networks (ANN), fuzzy logic, support vector machines (SVM), and genetic algorithms has become more popular for the prediction of processes where input data is insufficient, the modeled processes have random components, and the model parameters and conditions are unknown (Behzad et al., 2009; Dibike et al., 2001; Govindaraju and Rao, 2013; Lima et al., 2015; Nayak et al., 2004). One of the most used DBMs for water resources applications are ANNs. Inspired by the architecture of a biological neuron, ANN consists of a series of nodes, layers and functions that "learn" the data behavior based on the information that is given to the model. The capacity to reproduce highly complex non-linear functions and to generalize the time series trend makes ANN advantageous over other modeling approaches. Although the idea of ANNs was proposed in the 1940's (McCulloch and Pitts, 1943), their application in the field of hydrology is relatively new with an increased interest since the early nineties (ASCE, 2000). Since that time these models have been implemented to predict several water-related variables such as rainfall/runoff, evapotranspiration, flow, water quality and groundwater table (Adamowski and Chan, 2011; Daliakopoulos et al., 2005; French et al., 1992; Jayawardena and Fernando, 2001; Kumar et al., 2002; Maier and Dandy, 1996; Yoon et al., 2011).

Although ANNs have been successfully applied in a wide range of studies with exceptional results, the definition of the training parameters and architecture are still under discussion, requiring further analysis regarding the definition of input structures and standard methods to generalize the training functions to compare them with other modeling approaches. Maier and Dandy (2000) examined 43 studies where ANN was implemented in different water resources applications, and the training, parameterization and modeling efficiency was discussed. From these papers, most of the parameter optimization methods were performed by trial and error or not reported within the study. This omission makes it difficult to reproduce and compare the ANN architecture between different studies and locations. The discussion regarding the application of ANNs in hydrology opens the window for alternative procedures to evaluate and compare the effectiveness of different machine learning techniques for each hydrological process.

Among other learning methods, support vector machines (SVMs) have emerged as a new technique for the prediction of hydrologic variables. SVMs are based on the structural risk minimization inductive principle, which reduces the empirical error and model complexity, compared with the empirical risk minimization used by most of the ANN training algorithms (Basak et al., 2007; Yoon et al., 2011). Several authors have examined the capabilities of SVMs for the prediction of lake surface water levels (Khan and Coulibaly, 2006), runoff modeling (Bray and Han, 2004), nitrate concentration, groundwater sources (Arabgol et al., 2016), and for hybrid models in the forecasting of droughts at a range of lead times (Ganguli and Reddy, 2014). However, for the forecasting of daily groundwater levels on a local scale, the number of studies reported is very limited.

In the field of groundwater level predictions, the implementation of techniques such as ANN and SVM are practical options that provide useful information for decision making. In order to make an efficient evaluation of the water system dynamics on a small scale, it is necessary to have continuous real-time information on farm level water table fluctuations. The groundwater system is very dynamic, both spatially and temporally, and the factors surrounding the system are characterized by their complexity and non-linearity. The relationship between surface and subsurface water movement presents high randomness, especially when recharge sources, such as precipitation, are included. Several studies demonstrated the applicability of these techniques in the prediction and forecasting of groundwater levels. Coulibaly et al. (2001) evaluated three ANN models using a limited number of groundwater inputs to predict monthly levels of shallow and deep wells in Gondo Plain, Burkina Faso. Daliakopoulos et al. (2005) evaluated different ANNs to determine a proper architecture design for the forecasting of monthly groundwater levels up to 18 months ahead in Messara Valley, Greece, and Nayak et al. (2006) studied an ANN model to forecast monthly groundwater levels up to four months ahead in a shallow aquifer in Godavari, India. These studies concluded that using recurrent neural networks (RNN) in the forecasting of groundwater levels is advantageous to obtain the most accurate estimations. However, all of these studies used input time series with monthly and seasonal time steps and a limited amount of data to train the network. In a recent study, Guzmán et al. (2014) evaluated two RNN training algorithms for predicting daily groundwater levels for a northwest Mississippi well and found that a non-autoregressive with exogenous inputs (NARX) RNN is the most suitable ANN for the forecasting of daily levels in that location. This study was the first approach using machine learning techniques for the prediction of groundwater levels in the region.

The study of groundwater levels by using machine learning techniques is very recent, and additional research is needed to determine more efficient methods to generate optimal groundwater predictions. In comparison with ANN papers, only a few studies have used SVR to forecast groundwater levels. Asefa et al. (2004), for example, applied SVRs to identify the most influential monitoring wells in a regional groundwater network located in northwestern Washington State, USA. The authors reported that SVR is a useful technique to develop procedures for the management of subsurface water at a regional level. Similarly, Shiri et al. (2013) evaluated the applicability of different methods, including SVR, for the forecasting of groundwater levels. Almost all the methods used for the study showed better estimations in comparison with the auto-regressive moving average (ARMA) technique. Yoon et al. (2011) reviewed the performance of SVR when trained with different input values and compared it with the prediction of an ANN in a costal aquifer in Korea. The author found that, besides historical groundwater levels, precipitation and tidal levels should be included for the evaluation of models in coastal aquifers.

For the agricultural community, it is important to determine the effects of irrigation practices and changes in environmental conditions on the availability of water sources for future crop seasons (Dakhlalla et al., 2016; Karamouz et al., 2004; Scanlon et al., 2012). A better understanding of groundwater level fluctuations in agricultural wells helps provide integral management of water resources on the farms, especially to evaluate the annual recharge and withdrawal difference in a crop season. In most of the previous studies, groundwater fluctuations have been predicted from historical input data that compounds summer and winter periods, but there are a few studies that make a contrast between machine learning methods from data arranged by season. The objective of this study is to evaluate the influence of input time series divided by season in the forecasting of daily groundwater levels for a local well. Additionally, ANN and SVR methods are compared to identify a machine learning technique that can efficiently describe the complexities of groundwater level variability for a case study in the southeastern United States.

Materials and Methods

Case Study

Groundwater is the most important water source for crop production in the state of Mississippi. The Mississippi River Valley Alluvial (MRVA) aquifer supplies the majority of water for irrigation and catfish production in the region. In addition, the region has fertile soils, average annual rainfall around 1,300 mm, and appropriate environmental conditions that make agricultural production in the region competitive. However, over the past few decades, the use of groundwater has increased due to the expansion of irrigated hectares, requirements for higher crop yields, and lack of timely precipitation during the growing season. Although environmental conditions are favorable for agriculture, the aquifer levels have decreased steadily. Groundwater withdrawals are higher than the aquifer's recharge capacity by approximately 37 ha-m per year, with steady annual declines of 150 to 600 mm (Dyer et al., 2015). In addition, due to the water requirements to optimize crop production in the region, the number of permitted wells for irrigation is expanding continuously. The Mississippi Department of Environmental Quality (MDEQ) reports an increasing number of permits granted for irrigation purposes, especially in the central part of the MRVA-Mississippi. Figure 3.1 shows the extension of the MRVA in Mississippi and the volume of wells currently active for irrigation. The peak of withdrawals and number of active wells corresponds with the aquifer cone of depression that is expanding in the region. It is expected that the crop land area as well as the number of permitted irrigation wells will continue to increase in the near future.

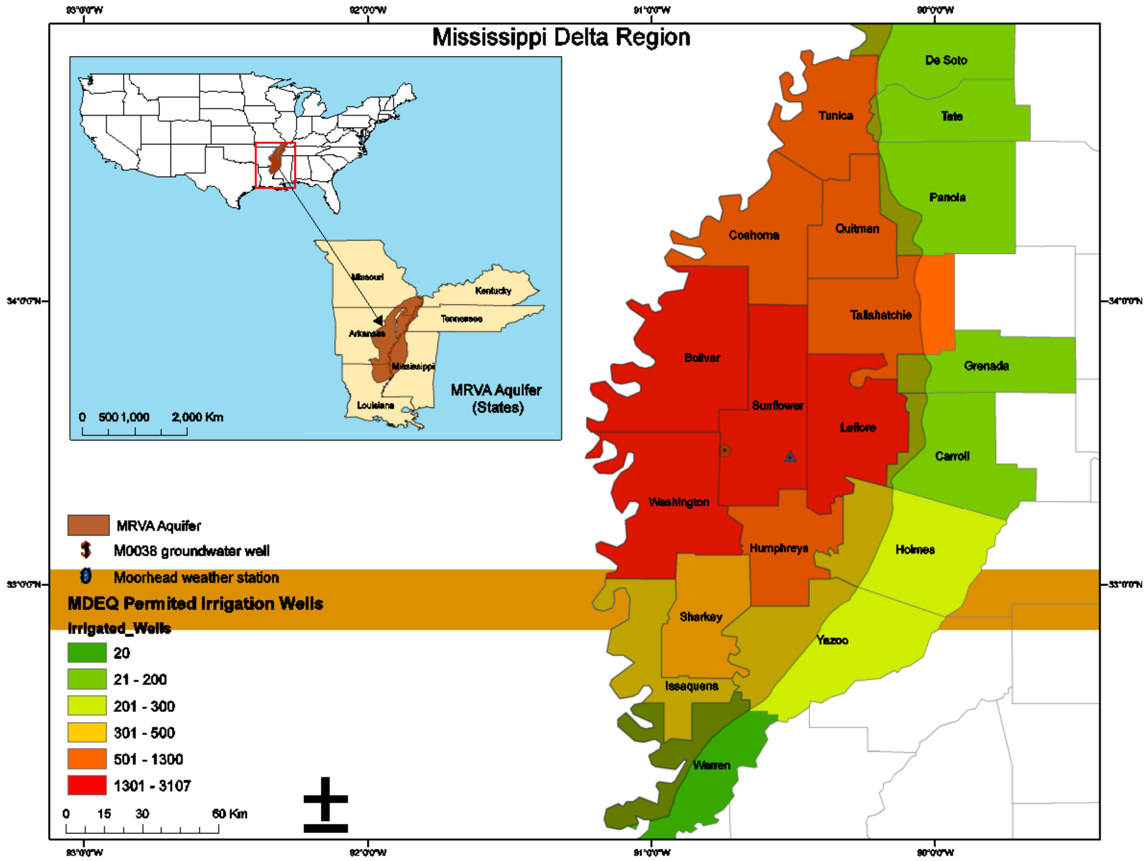

Figure 3.1 Study area with distribution of irrigation wells in the northwest Mississippi Shaded counties correspond to areas within the MRVA.

There is an increased urgency among farmers, stakeholders, and water management agencies to evaluate new methods to measure and track groundwater levels so that the impacts of water conservation practices can be determined. However, one drawback to monitoring groundwater levels is that the collection of daily information at a local scale is expensive and not feasible to maintain in the long term. In this study, nine years of groundwater level input data collected from June 1985 to September 1994 were used for the prediction of groundwater levels in northwest Mississippi. The USGS groundwater well (USGS M0038) located in Sunflower County between 33º 28’ 25” latitude and 90º 44’ 22” longitude was selected for its availability of continuous daily data over multiple years Figure 3.1. Groundwater information was extracted from the USGS-National Water Information System (NWIS) groundwater web-database (USGS, 2001). Similarly, daily precipitation from the same period of time was obtained from the National Oceanic and Atmospheric Administration (NOAA) climate data online website for a location near the USGS well (Figure 3.1). The weather station is located 32 km east of the well in Moorhead, Mississippi (GHCND:USC00226009) and provides daily precipitation summaries, max-min temperature, solar radiation, and evaporation for the region (NCDC, 2005). However, precipitation and temperature are the only parameters with continuous information. Thus, evapotranspiration values were calculated using the Priestley Taylor (PT) method, and solar radiation (SR) was calculated from the WP method (Woli and Paz, 2012). The WP method was developed to provide the most accurate estimation of SR in the Mississippi Delta, thus this method is the most reliable for this study. The nine years of historical groundwater level values were divided between two periods: April 1 and September 30 (hereafter designated as withdrawal season) and from October 1 to March 31 (hereafter designated as recharge season). By following this approach, the models were calibrated for periods with high influence of irrigation versus periods with reduced or no irrigation. Two calibrations for each model were generated from seasonal subsets of the annual groundwater level input information to evaluate possible changes on the parametrization by season.

Training Setup

Normalization

Input data was preprocessed to reduce noise in the calibration process. The multiple years dataset divided by seasons were analyzed to identify the presence of missing values. The groundwater time series has scattered discontinuities between February and March, 1991. Thus, this period was selected as a cutoff point to divide the input time series for training (70%) and testing (30%). In addition, the input datasets were normalized in a range between -1 to 1 in Matlab 2014®a (MathWorks, 2013). These preprocessed values were used to train the models and then rescaled to their original values after the training step was done. By performing this step model, efficiency is improved because the input variables are ranged at the same scale, which makes the parameterization faster and prevents the training from being dominated by extreme values.

Parameter and Input Selection

The models' training architectures were established by selecting the appropriate parameter and parameter ranges based on their impact on model performance. For ANN, a combination of parameters such as number of hidden layers, number of delays, and training methods were tested. The selection of the training architecture was made by trial and error until an optimum performance was found. More detailed information about the selection of parameters and establishment of the general ANN model architecture can be found in Guzmán et al. (2014). For SVR, parameters such as gamma, epsilon, and cost were tuned to obtain the optimal parameter combination. The parameters with the lowest training error were selected as optimal from the tuning process. In addition, three SVR kernel functions were evaluated, namely polynomial, radial basis function, and sigmoid. Every input combination was evaluated for each kernel function, and the selection of kernels was based on reports from previous studies that evaluated similar hydrologic processes using SVR (Cherkassky and Ma, 2004; Raghavendra. N and Deka, 2014; Yu et al., 2006).

To evaluate the predictive capabilities of ANN and SVR under similar conditions, lag time was selected to be the same for both models in this study. The optimal lag time was determined to be 100 days by trial and error. In addition, input combinations of lagged groundwater levels (Gw), precipitation, and evapotranspiration (ETo) were arranged in seven permutations: Gw, Pr, ETo, Gw+Pr, Gw+ETo, Pr+ETo, and Gw+Pr+ETo. After the optimal parameters were found, the ANN and SVR training functions were evaluated for each scenario to find the set of inputs that provides the best performance. An evaluation of the interdependence of input data layers found that correlations were fairly low. The correlation between groundwater level and precipitation was 0.076, while groundwater and evapotranspiration was 0.336, and finally precipitation and evapotranspiration was 0.112. For this study, the Mean Squared Error (MSE) was used as statistical measurement of the training performances. MSE provides the difference between observed and predicted values given by:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3.1}$$

where e_i is the error per value, y_i is the observed value, $\hat{y}_i$ is the predicted value and N is the total number of values. The permutation with the lowest MSE provides the best prediction performance.

Although ANN and SVR have different model approaches, both models are part of the same group of DBMs called machine learning. Thus, the generalities for training and testing the models require the same sequenced structure: 1) data is preprocessed and normalized to reduce the errors, 2) the model parameters are calibrated and trained to find an initial prediction, 3) training performance is evaluated by the selection of the lowest MSE and correlation coefficient, and 4) the model is tested with a new dataset to define the predictive efficiency and evaluated based on its error. Figure 3.2 shows the general structure used in this study to process the two machine learning models, and the correspondence between terms in ANN and SVR.

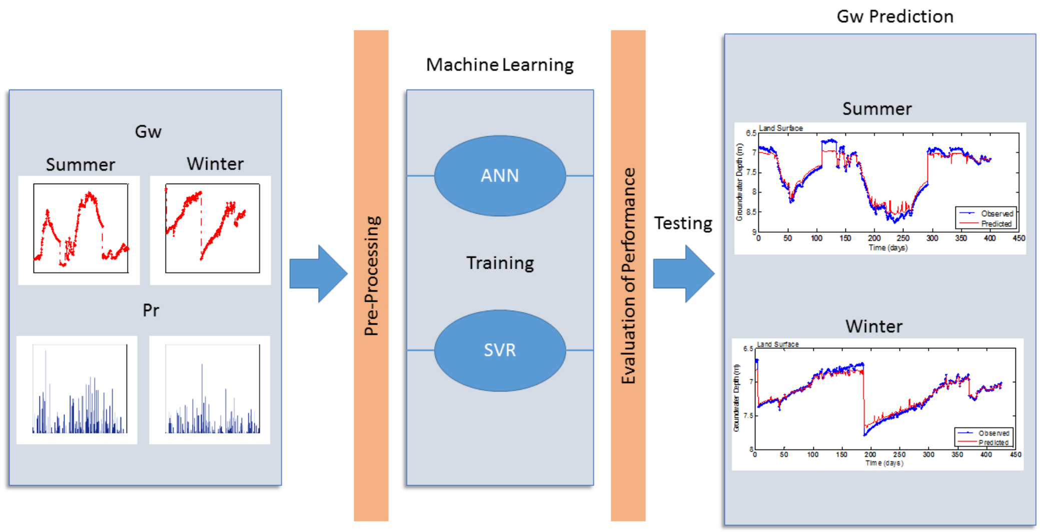

Figure 3.2 General description of the modeling process followed in this study

Artificial Neural Networks

ANN is a machine learning technique designed to evaluate processes with high complexities and reduced availability of information for the prediction. The technique is similar to the way human neurons classify and process information in the brain. ANN is one of the most commonly used machine learning methods for the estimation of hydrological variables in the last decades, and its applicability is very popular for the estimation of non-linear functions with efficient results. A general diagram of the ANN structure is presented in Figure 3.3. For this study, the ANN was implemented with a NARX neural network function:

$$y(t) = f\big(y(t-1), y(t-2), \dots, y\big(t-n_y\big), u(t-1), u(t-2), \dots, u(t-n_u)\big) \quad (3.2)$$

where $\mathrm{u(t)}$ and $\mathrm{y(t)}$ are the inputs and outputs of the network, respectively, t is the time step, $\mathrm{n_u}$ and $\mathrm{n_y}$ are the input and output layers of the network, respectively, for a hidden layer in a f nonlinear function. This function uses a dynamic recurrent neural network that takes the input data and feeds back from the output in to the function until convergence is reached. It has been shown that NARX networks are much faster than other ANN functions, and they also perform better when the process being predicted involves long term dependencies (Siegelmann et al., 1997). Thus, we selected the NARX networks as the most suitable function for the purposes of this study. The model was evaluated to determine the best algorithm and parameters that generate an optimal performance. The Levenberg – Marquardt and Bayesian Regularization training algorithms were tested by trial and error with different combinations of time delays (5, 25, 50, 75 and 100) and hidden layers until the optimal performance was reached. From this procedure, the optimal architecture was a Bayesian Regularization algorithm with 100 time delays and two hidden layers. For more details on the training process and the modeling procedure, the reader can refer to Guzman et al. (2014).

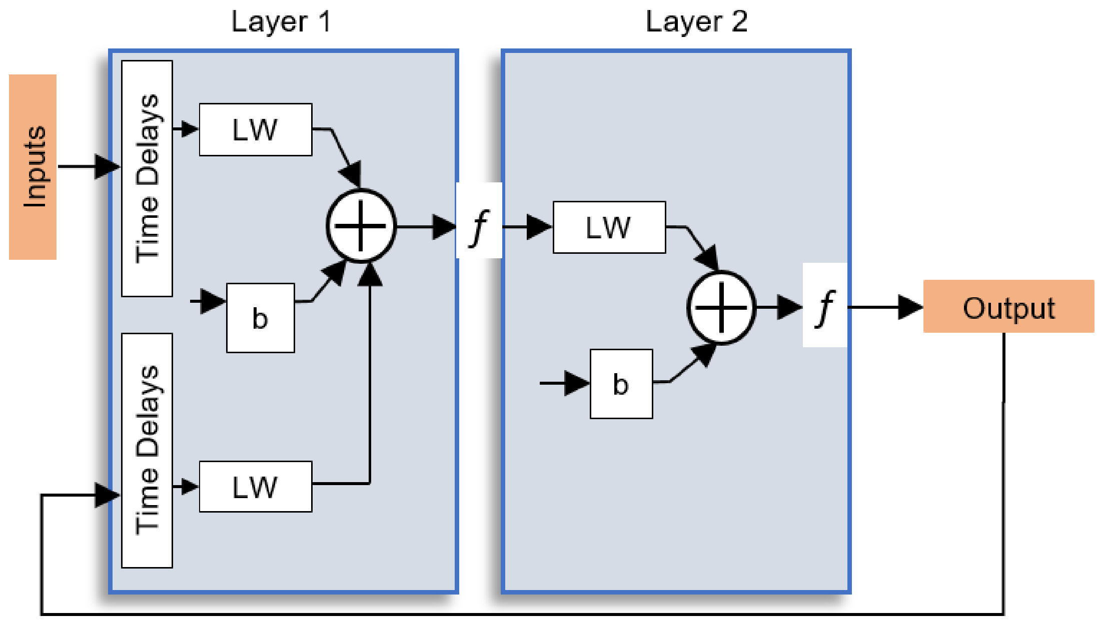

Figure 3.3 A NARX neural network with two hidden layers.

(IW: input weights, LW: layer weights, b: biases, f: function approximation).

Support Vector Regression

SVR is a relatively new machine learning technique introduced by Cortes and Vapnik (1995) in an effort to optimize a problem subject to parametrization constraints. This technique has gained popularity in the last several years for its strong capabilities in predicting and generalizing complex problems. One of the advantages of SVR is that it uses structural risk minimization instead of the empirical risk used by other machine learning models such as ANN (Raghavendra and Deka, 2014). SVR consists of a series of input vectors that "support" the training architecture for the estimation of non-linear time series. The training function is performed in a hyperplane where the series trend is transformed to be treated as a simple linear function (Figure 3.4). The objective is to find a function f(x) that can predict the target values y_i with an admissible error no higher than ε. A simple linear expression of an SVR function is given by:

$$f(x) = \langle w, x \rangle + b \tag{3.3}$$

where $w \epsilon x$, x is the input space, $b \epsilon R$ is the function bias and $\langle w, x \rangle$ is the dot product between vector w and x. For this problem, the ideal is to have the smallest w possible so that the norm $\|w\|^2 = \langle w. w \rangle$ is minimized. This problem can be written as a convex optimization problem:

$$\text{minimize } \frac{1}{2}\|\mathrm{w}\|^2 \tag{3.4}$$

$$\text{subject to} \begin{cases} \mathrm{y_i} - \langle \mathrm{w}, \mathrm{x_i} \rangle - \mathrm{b} \leq \varepsilon \\ \langle \mathrm{w}, \mathrm{x_i} \rangle + \mathrm{b} - \mathrm{y_i} \leq \varepsilon \end{cases}$$

The assumptions in equation 3.3 should be feasible, meaning that there is a function which is able to approximate all pairs of observed x_i and predicted y_i data with ε precision. In reality, it is not always possible to find a feasible function, in which case, it becomes necessary to include the ξ, ξ_i^* error variables to handle equation 3.3. This step transforms the initial problem into:

$$\frac{1}{2}\|\mathrm{w}\|^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{l}}(\xi_\mathrm{i} + \xi_\mathrm{i}^*) \tag{3.5}$$

$$\text{subject to} \begin{cases} \mathrm{y_i} - \langle \mathrm{w}, \mathrm{x_i} \rangle - \mathrm{b} \leq \varepsilon + \xi_\mathrm{i} \\ \langle \mathrm{w}, \mathrm{x_i} \rangle + \mathrm{b} - \mathrm{y_i} \leq \varepsilon + \xi_\mathrm{i}^* \\ \xi_\mathrm{i}, \xi_\mathrm{i}^* \qquad\qquad\qquad \geq 0 \end{cases}$$

For a non-linear problem, equation 3.3 can be solved by the use of Lagrangian multipliers, that solves the dual optimization problem, by:

$$\frac{1}{2}\|\mathrm{w}\|^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{l}}(\xi_\mathrm{i} + \xi_\mathrm{i}^*) \tag{3.6}$$

$$\text{subject to} \begin{cases} \mathrm{y_i} - \langle \mathrm{w}, \mathrm{x_i} \rangle - \mathrm{b} \leq \varepsilon + \xi_\mathrm{i} \\ \langle \mathrm{w}, \mathrm{x_i} \rangle + \mathrm{b} - \mathrm{y_i} \leq \varepsilon + \xi_\mathrm{i}^* \\ \xi_\mathrm{i}, \xi_\mathrm{i}^* \qquad\qquad\qquad \geq 0 \end{cases}$$

Equation 3.5 is reformulated in to:

$$f(x) = \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) k \langle x_i, x \rangle + b \quad (3.7)$$

where α_i, α_i^* are Lagrangian multipliers and $k\langle x_i, x \rangle$ is a kernel function that evaluates the nonlinearity between two instances of the input variables for l input values. The Lagrangian multipliers kernel function maps the non-linear function in a high dimensional feature space and transforms it into a linear problem to be used in the standard SVR. Figure 3.4 shows a graphic description of a nonlinear SVR regression problem where ξ_i and ξ_i^* represent the confidence interval of the kernel function f(x) and ε represents the admissible error. The values that are outside of the confidence interval are the function prediction errors.

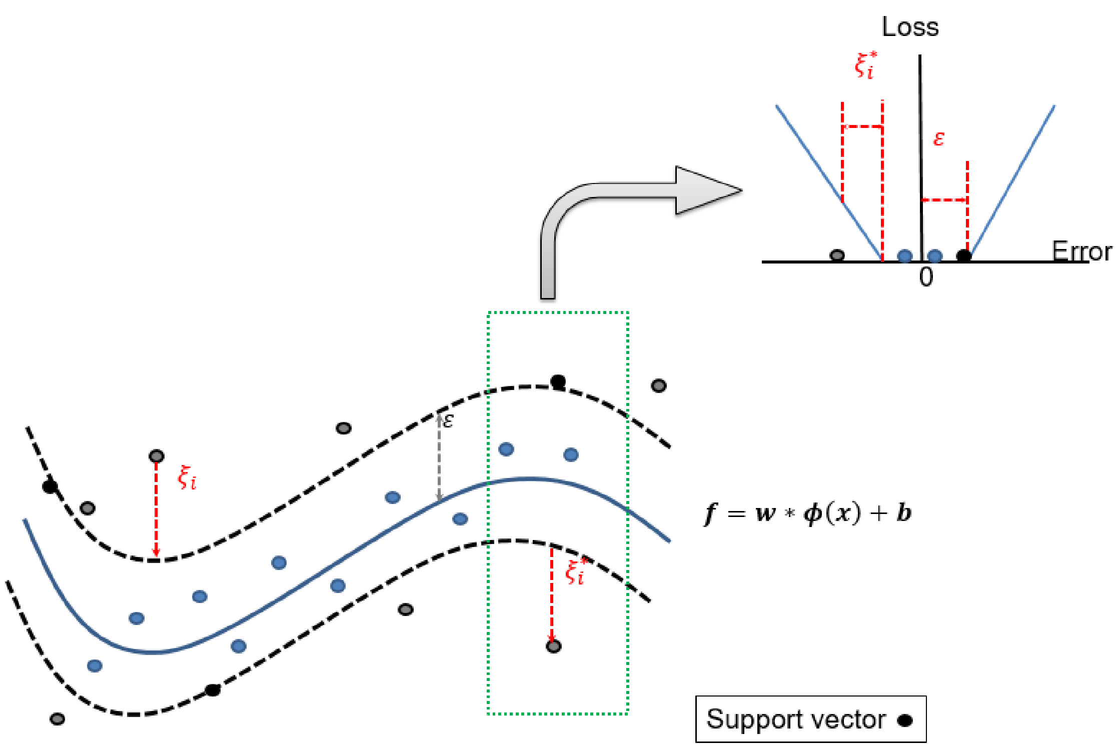

Figure 3.4 Diagram of the SVR Vapnik's architecture

Adapted from Yu et al. (2006)

Results and discussion

Preprocessing and Training

In this study, MSE was used as a measurement of performance to evaluate the best combination from three input variables – lagged daily groundwater level, precipitation, and evapotranspiration data series - for each model as shown in Table 3.1. The Gw + Pr scenario provides the optimal combination with MSEs of 0.01277 and 0.00123 meters for ANN and SVR, respectively. Results reported by Coulibaly et al. (2001) show similar findings with a significant influence from precipitation and previous groundwater levels combined into the neural network model performance. Yoon et al. (2011) also showed that the two input variables combined are sensitive and essential for the prediction of groundwater levels. Our results show the high influence of the combined input variables of precipitation and previous groundwater levels for the daily level prediction in the study well. The results indicate that the processes of recharge are highly related to the precipitation patterns and the historical variability of groundwater levels. It is important to note that the selection of the most sensitive input values is case dependent, but in general for groundwater prediction, previous groundwater levels and precipitation have a significant influence on the model performance. Table 3.1 shows the differences between MSEs for all permutations using ANN and SVR. For scenarios that did not include ETo, both models had acceptable errors between predicted and observed values ranging from 0.001277 to 0.21058 m. However, the addition of ETo input data had a negative impact on the performance of the SVR model as shown in the Gw+ETo and Gw+Pr+ETo results, with errors of 6.70807 and 4.70044, respectively. The SVR model with Gw+Pr as inputs provides the overall lowest MSE.

Table 3.1 Mean squared error (MSE) for different input variable combinations

	Gw	Pr	ETo	Gw+Pr	Gw+ETo	Pr+ETo	Gw+Pr+ETo*
SVR	0.21058	0.10572	0.22274	0.00123	6.70807	0.39035	4.70044
ANN	0.02922	0.15830	0.71580	0.01277	0.70844	0.52426	0.03103

*Gw = Groundwater, Pr = precipitation, ETo = Evapotranspiration

SVR Architecture

Below is a summary of selected parameter combinations for the analysis of the best SVR training architecture. The results from scenarios of kernel and kernel parameters are shown in Table 3.2. For the entire year time series model, the best performance - or lowest MSE_y - is shown by a radial basis function (RBF), with a gamma (γ) of 0.01, a cost function (C) of 100, and epsilon (ε) of 0.1, followed by a RBF kernel with a γ of 0.0001 C of 1 and ε of 0.1. Similarly, the best parameter performance for summer and winter season is shown by the RBF kernel function with a gamma (γ) of 0.01, a cost function (C) of 100, and epsilon (ε) of 0.1 respectively. For all kernels and parameter combinations, the winter time series has the smallest training error (MSE_w) followed by the entire year (MSE_y) and summer (MSE_s). These results show the applicability of using time series groundwater level data divided by season for the evaluation of recharge and withdrawal levels. In this case, the prediction of the recharge period becomes a linear problem, and the withdrawal period gains performance efficiency.

Table 3.2 Summary of selected results from SVR parameter estimation

Kernel Type	Gamma	Cost	Epsilon	*MSE_y	MSE_s	MSE_w
RBF	0.01	100	0.1	0.00072	0.00123	0.00011
RBF	0.001	1	0.1	0.00161	0.00578	0.00135
RBF	0.3	100	0.5	0.00522	0.01921	0.00181
RBF	0.01	100	0.51	0.02453	0.04043	0.00548
RBF	0.001	91	0.51	0.02693	0.04298	0.00590
Polynomial	0.001	91	0.1	0.10524	0.00160	0.00022
Polynomial	0.001	91	0.51	0.28806	0.06972	0.07990
Sigmoid	0.001	1	0.1	1.59339	0.07288	0.01629
Polynomial	0.5	91	0.1	2.39758	0.13467	0.02294

MSEy: Error for entire year time series calibration, MSEs: Error for summer time series calibration, and MSEw: Error for winter time series calibration

The evaluation of kernel parameters in Table 3.2 shows that gamma and epsilon are the most sensitive parameters for the calibration of the training architecture. Figure 3.5 shows the model response of gamma and epsilon adjusted one at a time for an RBF kernel function, with C (100) and γ (0.01) or ε (0.1) were kept constant for each trial. Figure 3.5a presents the changes in the model efficiency when gamma is increased from 0.01 to 0.5. Summer and the entire year time series have low peaks for a gamma of 0.05 and 0.4, respectively. It is clear that γ is the most sensitive parameter for the training function. In general, a gamma between 0 and 0.1 provides good performance for this case study. In contrast, Figure 3.5b shows the error variation for epsilon. For this parameter selection, the training function has a good performance between 0 and 0.2 for all of the seasonal combinations. The error increases exponentially after epsilon is increased for the three trials. The MSE response to variations is more evident for the epsilon parameter, with a more consistent response. The trial for C is not shown, as the response is almost linear and is the least sensitive parameter. Although the C, γ, and ε are interdependent, the importance of gamma is apparent for identifying an efficient training architecture.

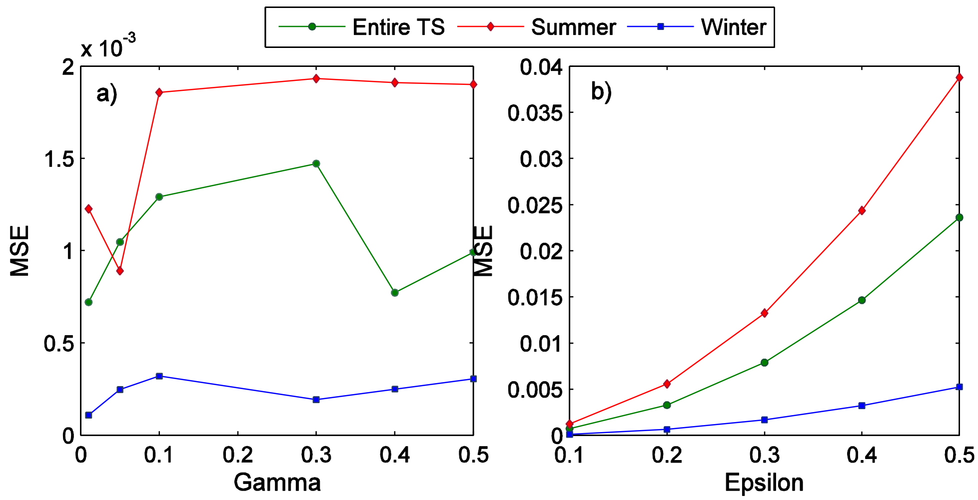

Figure 3.5 Seasonal variation of Mean squared Error with different a) Gamma and b) Epsilon for a Radial Basis function and cost = 100

Table 3.3 shows a parallel of the optimal training architecture determined for ANN and SVR in this study. It is important to recognize that this set up is case dependent and that it can be variable. The number of support vectors, for example, is the result of the parameterization and kernel function selected for this study.

Table 3.3 Training architecture for ANN and SVR models

ANN	SVR
• Number of Hidden Layers: 2 • Training algorithm: Bayesian Regularization $C(k) = \beta * E_d + \alpha * E_w$	• Support Vectors: 62 • Kernel Function: Radial Basis Function $exp(-gamma * \|u - v\|\text{^}2)$

Seasonal Prediction

Summer Period

Time series of observed and predicted daily groundwater levels with summer inputs for SVR and ANN are shown in Figure 3.6 and Figure 3.7.The observed time series shows two strong withdrawal periods between 20-100 days (first summer period) and 200-275 days (second summer period). The first summer period is characterized by its gradual and comparatively reduced volume of water extracted in contrast with the second summer period. Between 200 and 250 days, the gradual withdrawals with depths greater than 8.4 m show a second summer period with more critical conditions for the well under study. The predicted SVR time series was able to appropriately describe and predict the seasonal trend level for the well. As illustrated in Figure 3.6, the SVR modeled time series initializes by underpredicting groundwater levels between 0 and 25 days. When the model reaches the lower withdrawal peak, the prediction gains performance and shifts to overpredicting after the first critical summer period. After 100 days, the SVR model underpredicts again until 170 days with an error less than 0.025 m. In the first summer irrigation season, the model efficiently predicts the trend of water demands. The model shows a small reduction in performance and higher variability for the second summer irrigation period (200 – 250 days). However, the SVR model is able to adequately represent the daily withdrawal trends.

Figure 3.7 shows the prediction results from the ANN model for the withdrawal season. Similar to SVR, the ANN model is able to predict the groundwater trends for summer periods. However, the model has a delayed response after 150 days and a reduction on the predictive performance after the second withdrawal period. The predictive function underpredicts the observed values until it has reached the first withdrawal decline period. After this period, the prediction is less efficient in capturing the small drops between 100 and 170 days. For the second withdrawal season, the ANN predictive capacity is affected by a delayed response of approximately 50 days, especially after 240 days. When the model passes 350 days, ANN is not able to represent the daily variability for the groundwater well. Although both machine learning techniques can provide satisfactory predictions, SVR shows better predictive performance for the critical periods of water demand and small recharge drops from 100 - 170 days and 300 – 400 days. Overall, SVR is superior to ANN in terms of generalization, performance (based on MSE), and predictive capacity. These findings are important in providing more efficient level estimations for the summer irrigation periods when information regarding the availability of water to supply crop water demand is important to ensure profitable crop yields.

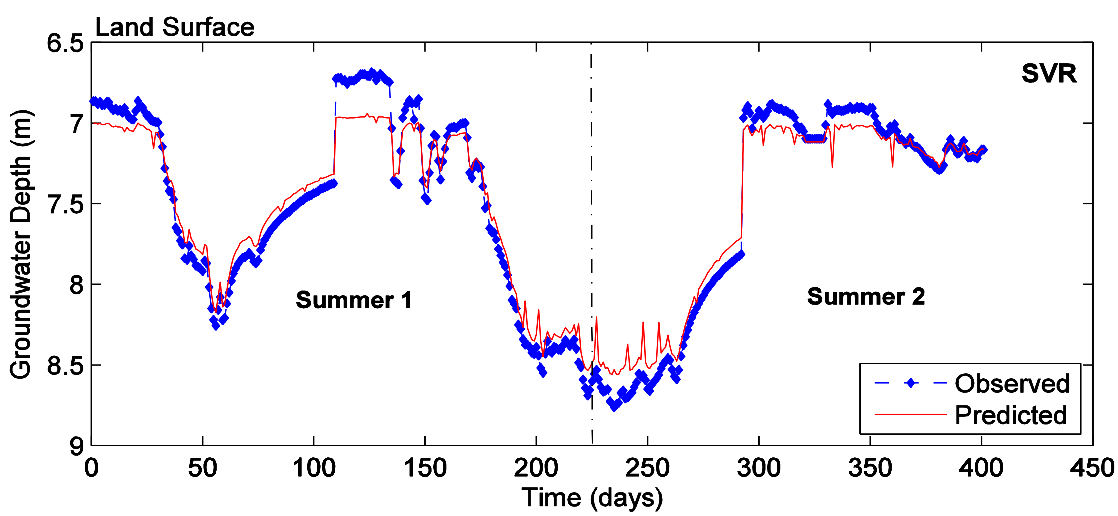

Figure 3.6 SVR daily groundwater prediction with summer input data

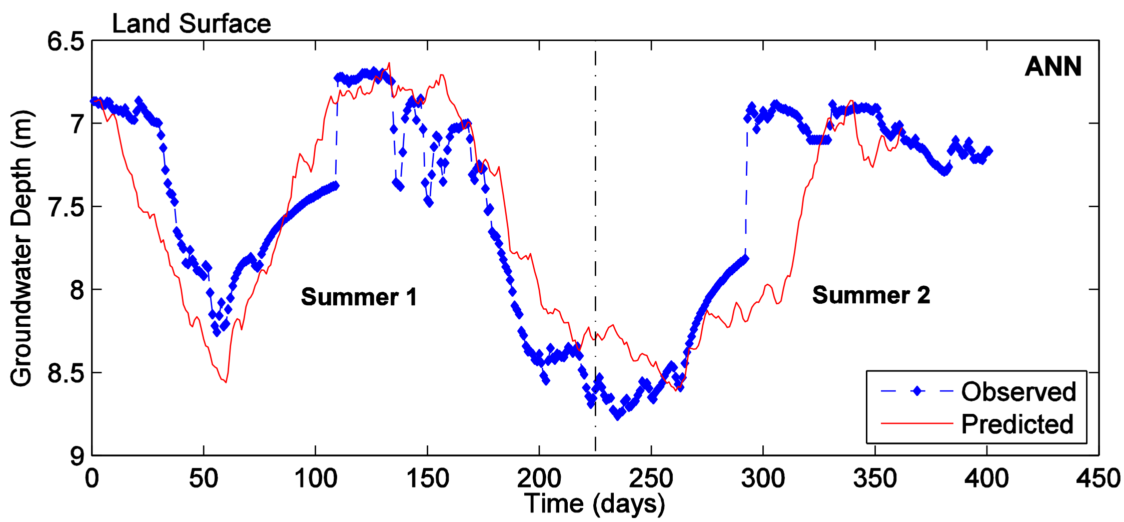

Figure 3.7 ANN daily groundwater prediction with summer input data

Winter Period

The results of daily groundwater level predictions for the recharge season using the SVM and ANN models are shown in Figure 3.8 and Figure 3.9, respectively. The observed winter time series shows a continuous trend of linear groundwater level increases for the two main recharge periods. Similar to the summer period, the winter period is divided into two periods, the first recharge period (between 0 to 190 days) with a level increase from 7.4 m to 6.6 m below the land surface, and a second period (between 190 to 360 days) from 7.8 m to 7 m. The SVR daily prediction agrees well with the observed groundwater levels as is shown in Figure 3.8, whereas the discrepancies become larger when leading time is between 150 and 220 days. However, these discrepancies are minimal, and the SVR successfully predicts the rising and falling trends. In contrast, the accuracy from the ANN prediction is much lower, especially between 0 and 160 days (Figure 3.9). The difference between observed and predicted groundwater levels is around ±0.4m during the first recharge season with more stable results for the second recharge season. However, compared to the SVR performance, the ANN prediction is not efficient. The good SVR performance for the winter recharge period can be attributed to the simplification of the time series general shape. As is shown in the figures, groundwater trends for recharge periods become linear, thus the model does not require using high computations to linearize the function in a high dimensional feature space. On the other hand, the ANN model is trained with a NARX function that is designed to analyze non-linear time series, hence it becomes less effective for the winter recharge season.

The objective of this study was to evaluate the performance of two machine learning techniques for the prediction of daily groundwater levels. For comparative purposes, the ANN architecture was unchanged in each seasonal dataset, but for future studies alternative methods to describe the recharge season should be evaluated. One of the most remarkable findings of this study is that the recharge time series can be treated as a linear problem, which implies less computational requirements and provides faster solutions.

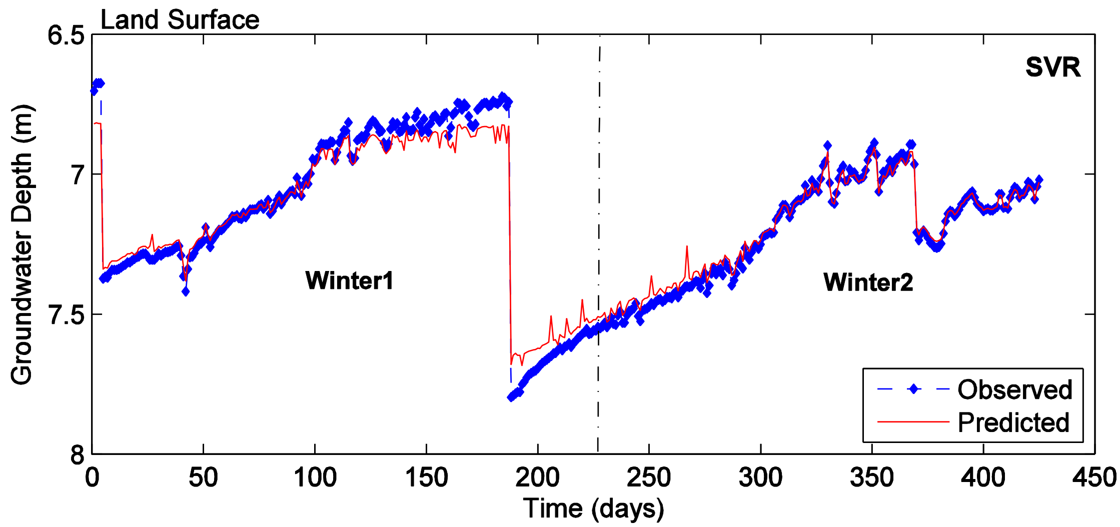

Figure 3.8 SVR daily groundwater prediction with winter input data

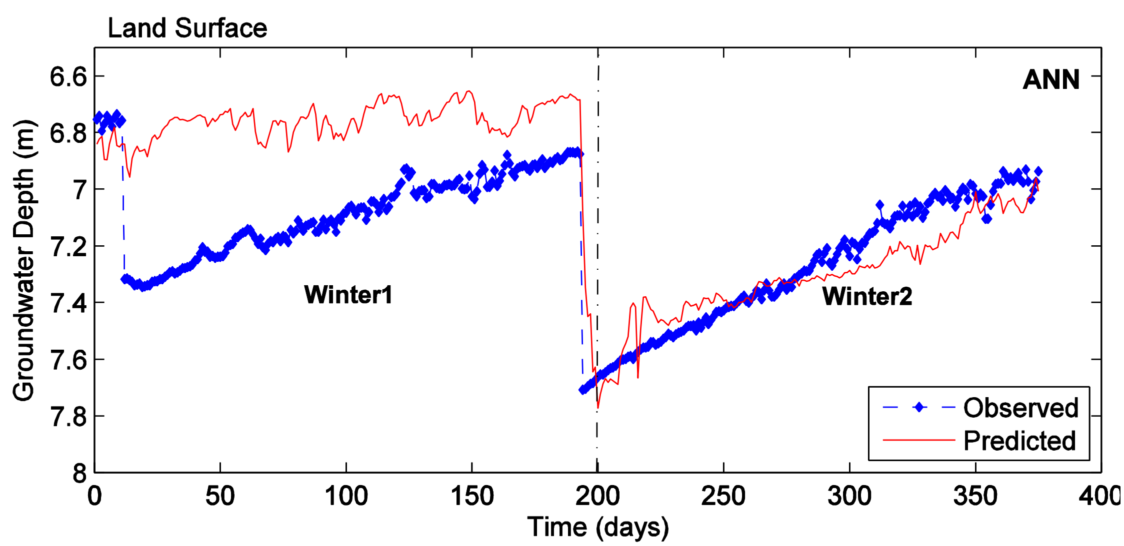

Figure 3.9 ANN daily groundwater prediction with winter input data

Previous work performed by Guzman et al. (2014) for the same study area indicated that the performance of ANN was more efficient in comparison with SVR when the yearly time series are used as inputs for the model. However, for this study (time series input divided by seasons), ANN has lower predictive capabilities, and it is difficult for the model to capture the winter groundwater trends. The entire year input data has 1428 values per input for the training step. In contrast, the number of values for the summer period is 732, and for the winter period, it is 695. Partitioning the input data into two seasons reduced the performance of NARX ANN. However, using separate seasonal data, as opposed to using the whole time series, favored the SVR. In fact, the best overall performance was provided by the SVR model for the winter season.

The modeling results presented in this section show the comparative performances of SVR and ANN for the prediction of daily groundwater levels during withdrawal and recharge seasons. This underscores the applicability of machine learning techniques for the analysis of groundwater levels at the farm level when the availability of measured data is a constraint. Although the minimum number of values for the model calibration and validation was not established, our study used nine years of daily values in which 1642 were used for training and 950 were used for testing. The volume of input information utilized in this study is notably higher compared with the number of values reported by authors using similar techniques (Coulibaly et al., 2001; Daliakopoulos et al., 2005; Nayak et al., 2006; Yang et al., 2009). Based on the MSE, the results indicate that SVR estimates provide the best performance for a well located in the Mississippi Delta cone of groundwater depression. Data preprocessing and parameter calibration is a subject of special attention because of their role in determining an accurate prediction. The common rules to train these techniques in the field of hydrology have not been established. Thus, it is important to report which parameters and input variables were selected and the methods employed. This study provides a baseline for the implementation of ANN and SVR in forecasting groundwater levels in the Mississippi region. However, is important to note the necessity of finding the optimal parameters for each case study because the models are based on data. The application of new techniques such as SVR is important for the future sustainability of groundwater resources in regions where the availability of short term measured data is not sufficient to provide information for decision making.

Conclusions

This study has demonstrated that skillful predictions of groundwater levels at farm wells can be provided by machine learning approaches with input variables divided by season to generate information for decision making. Both ANN and SVR are capable of efficiently predicting the lower and higher withdrawal trends of the summer withdrawal season. However, for the winter season, ANN underpredicted the groundwater level trend. SVR is superior to ANN in terms of prediction performance and capacity to reproduce the seasonal groundwater. An SVR with a RBF kernel function, γ = 0.01, C=100, and ε =0.1 provided the best architecture for both the summer withdrawal and winter recharge seasons. The results also demonstrated that SVR training had reduced computational requirements and fast iteration responses. Although finding the parameters is time consuming, SVR is still the most efficient approach for predicting groundwater levels based on its predictive performance and estimation of the general trend. Because the techniques used in this study are based on data, it also offers a less costly and efficient alternative compared with process-based models. The proposed methodology for simulating and predicting future groundwater levels is a novel approach to help farmers and stakeholders effectively manage and plan for the efficient use of groundwater resources. This new method can also provide input information for studies at the regional scale. The implementation of this modeling approach will complement efforts to manage groundwater levels in the MRVA aquifer by generating daily level predictions that are not available from monitoring wells or current groundwater modeling approaches. Future work includes evaluating the SVR architecture after tuning parameters for multiple wells in the region, proposing a standard for model parametrization to compare results with multiple studies under similar conditions, and evaluating the crop physiological changes in response to variable groundwater levels.

References

Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology, 407(1–4): 28-40. DOI:http://dx.doi.org/10.1016/j.jhydrol.2011.06.013

Arabgol, R., Sartaj, M., Asghari, K., 2016. Predicting Nitrate Concentration and Its Spatial Distribution in Groundwater Resources Using Support Vector Machines (SVMs) Model. Environmental Modeling & Assessment, 21(1): 71-82.

ASCE, 2000. Artificial Neural Networks in Hydrology. I: Preliminary Concepts. Journal of Hydrologic Engineering, 5(2): 115-123. DOI:10.1061/(ASCE)1084-0699(2000)5:2(115)

Asefa, T., Kemblowski, M.W., Urroz, G., McKee, M., Khalil, A., 2004. Support vectors–based groundwater head observation networks design. Water Resources Research, 40(11).

Basak, D., Pal, S., Patranabis, D.C., 2007. Support vector regression. Neural Information Processing-Letters and Reviews, 11(10): 203-224.

Behzad, M., Asghari, K., Eazi, M., Palhang, M., 2009. Generalization performance of support vector machines and neural networks in runoff modeling. Expert Systems with Applications, 36(4): 7624-7629. DOI:http://dx.doi.org/10.1016/j.eswa.2008.09.053

Bray, M., Han, D., 2004. Identification of support vector machines for runoff modelling. Journal of Hydroinformatics, 6(4): 265-280.

Cherkassky, V., Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1): 113-126.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach Learn, 20(3): 273-297. DOI:10.1007/BF00994018

Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. Water Resources Research, 37(4): 885-896. DOI:10.1029/2000wr900368

Dakhlalla, A.O., Parajuli, P.B., Ouyang, Y., Schmitz, D.W., 2016. Evaluating the impacts of crop rotations on groundwater storage and recharge in an agricultural watershed. Agricultural Water Management, 163: 332-343. DOI:http://dx.doi.org/10.1016/j.agwat.2015.10.001

Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. Journal of Hydrology, 309(1–4): 229-240. DOI:http://dx.doi.org/10.1016/j.jhydrol.2004.12.001

Dibike, Y., Velickov, S., Solomatine, D., Abbott, M., 2001. Model Induction with Support Vector Machines: Introduction and Applications. Journal of Computing in Civil Engineering, 15(3): 208-216. DOI:doi:10.1061/(ASCE)0887-3801(2001)15:3(208)

Dyer, J., Mercer, A., Rigby, J.R., Grimes, A., 2015. Identification of recharge zones in the Lower Mississippi River alluvial aquifer using high-resolution precipitation estimates. Journal of Hydrology, 531, Part 2: 360-369. DOI:http://dx.doi.org/10.1016/j.jhydrol.2015.07.016

French, M.N., Krajewski, W.F., Cuykendall, R.R., 1992. Rainfall forecasting in space and time using a neural network. Journal of hydrology, 137(1): 1-31.

Ganguli, P., Reddy, M.J., 2014. Ensemble prediction of regional droughts using climate inputs and the SVM–copula approach. Hydrological Processes, 28(19): 4989-5009.

Govindaraju, R.S., Rao, A.R., 2013. Artificial neural networks in hydrology, 36. Springer Science & Business Media.

Guzmán, S.M., Paz, J.O., Tagert, M.L.M., 2014. A neural network framework to estimate groundwater levels in the Mississippi river valley shallow alluvial aquifer ASABE Annual International Meeting. ASABE, Montreal, Quebec Canada.

Jayawardena, A., Fernando, T., 2001. River flow prediction: an artificial neural network approach. IAHS Publication: 239-246.

Karamouz, M., Kerachian, R., Zahraie, B., 2004. Monthly Water Resources and Irrigation Planning: Case Study of Conjunctive Use of Surface and Groundwater Resources. Journal of Irrigation and Drainage Engineering, 130(5): 391-402. DOI:doi:10.1061/(ASCE)0733-9437(2004)130:5(391)

Khan, M.S., Coulibaly, P., 2006. Application of support vector machine in lake water level prediction. Journal of Hydrologic Engineering, 11(3): 199-205.

Kumar, M., Raghuwanshi, N., Singh, R., Wallender, W., Pruitt, W., 2002. Estimating evapotranspiration using artificial neural network. Journal of Irrigation and Drainage Engineering, 128(4): 224-233.

Lima, M.L., Romanelli, A., Massone, H.E., 2015. Assessing groundwater pollution hazard changes under different socio-economic and environmental scenarios in an agricultural watershed. Science of The Total Environment, 530–531: 333-346. DOI:http://dx.doi.org/10.1016/j.scitotenv.2015.05.026

Maier, H.R., Dandy, G.C., 1996. The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. Water Resources Research, 32(4): 1013-1022. DOI:10.1029/96WR03529

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software, 15(1): 101-124. DOI:http://dx.doi.org/10.1016/S1364-8152(99)00007-9

MathWorks, I., 2013. MATLAB R2013a. The MathWorks Inc.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4): 115-133.

Nayak, P., Rao, Y.R.S., Sudheer, K.P., 2006. Groundwater Level Forecasting in a Shallow Aquifer Using Artificial Neural Network Approach. Water Resour Manage, 20(1): 77-90. DOI:10.1007/s11269-006-4007-z

Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., 2004. A neuro-fuzzy computing technique for modeling hydrological time series. Journal of Hydrology, 291(1–2): 52-66. DOI:http://dx.doi.org/10.1016/j.jhydrol.2003.12.010

NCDC, 2005. National Climatic Data Center.

Raghavendra. N, S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Applied Soft Computing, 19(0): 372-386. DOI:http://dx.doi.org/10.1016/j.asoc.2014.02.002

Scanlon, B.R. et al., 2012. Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. Proceedings of the national academy of sciences, 109(24): 9320-9325.

Shiri, J., Kisi, O., Yoon, H., Lee, K.-K., Hossein Nazemi, A., 2013. Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques. Computers & Geosciences, 56: 32-44. DOI:http://dx.doi.org/10.1016/j.cageo.2013.01.007

Siegelmann, H.T., Horne, B.G., Giles, C.L., 1997. Computational capabilities of recurrent NARX neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 27(2): 208-215.

United States Geological Survey USGS, 2001. National Water Information System data available on the World Wide Web (Water Data for the Nation).

Woli, P., Paz, J.O., 2012. Evaluation of Various Methods for Estimating Global Solar Radiation in the Southeastern United States. Journal of Applied Meteorology and Climatology, 51(5): 972-985. DOI:10.1175/JAMC-D-11-0141.1

Yang, Z.P., Lu, W.X., Long, Y.Q., Li, P., 2009. Application and comparison of two prediction models for groundwater levels: A case study in Western Jilin Province, China. Journal of Arid Environments, 73(4–5): 487-492. DOI:http://dx.doi.org/10.1016/j.jaridenv.2008.11.008

Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. Journal of Hydrology, 396(1–2): 128-138. DOI:http://dx.doi.org/10.1016/j.jhydrol.2010.11.002

Yu, P.-S., Chen, S.-T., Chang, I.F., 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology, 328(3–4): 704-716. DOI:http://dx.doi.org/10.1016/j.jhydrol.2006.01.021

CHAPTER IV

APPLICATION OF SUPPORT VECTOR REGRESSION FOR GROUNDWATER LEVEL FORECASTING: SELECTION OF INPUT VARIABLES

A paper to be submitted to the Journal of Expert Systems with Applications
Sandra M. Guzman, Joel O. Paz, Mary Love M. Tagert, and Andrew E. Mercer.

Abstract

The availability of groundwater for irrigation plays an important role in agricultural sustainability due to the high crop water demands in regions where subsurface water is the main source of water for irrigation. This continuous demand of water in addition to the expected variations on climate requires the implementation of innovative modeling techniques to determine alternative solutions for the management of groundwater sources. Support Vector Regression (SVR) is one of the machine learning techniques that has gained popularity in hydrological studies over the last decade. The exceptional generalization properties, the use of structural risk minimization instead of empirical risk, and the capacity to avoid local minima during the optimization process are some of the main advantages of this technique. However, there is limited research regarding the estimation of input variables to predict groundwater levels using SVR. In this paper, a sensitivity analysis was conducted to determine the optimal set of input variables and variable arrangement to obtain a predictive SVR model for groundwater levels in a monitoring well located in Northwest Mississippi. Data on daily groundwater

levels (Gw), precipitation (Pr), and evapotranspiration (ET) were used as inputs to the SVR model. The combination of Gr+Pr provided the optimal input set for predictive SVR modeling of daily groundwater levels. The model performed poorly with the addition of ET as the third input variable (Gr+Pr+ET), and the degraded performance was particularly evident during the recharge periods.

Introduction

Groundwater is an important source of water for agriculture. The continual expansion of crop land areas and increasing temperatures have raised the demand on groundwater for irrigation in the last twenty five years (Dyer et al., 2015; Kebede et al., 2014; Mainuddin et al., 1997; Scanlon et al., 2012; Wax Charles L. et al., 2009). The Mississippi River Valley Alluvial (MRVA) aquifer is the second most pumped aquifer in the United States for irrigation with daily withdrawals of approximately 406.94 $m^3\ s^{-1}$ (Maupin and Barber, 2005). The area of the MRVA aquifer that covers the Mississippi Delta Region (MDR) has experienced a concerning decline in the last fifteen years where water levels have receded over six meters, especially in the central area, where irrigated agriculture is highly extensive (Byrd, 2011). Because of the importance of the aquifer for agriculture in the MDR, it is critical to design strategic management practices starting from a local scale or farm level to optimize water withdrawals while ensuring a sustainable use of the aquifer for the long term. To achieve this purpose, it is necessary to better understand the variability of groundwater levels at the farm scale through increased monitoring and by implementing innovative modeling tools that can be easily transferred to and used by managers and stakeholders.

The challenge of increasing agricultural production in the MDR while maintaining the viability of the MRVA requires the adoption of new techniques and methods to predict daily changes in groundwater levels which are necessary to design appropriate water management plans. Machine learning techniques such as SVR have gained popularity within the water resources community due to their efficacy and efficiency to predict hydrologic time series, and also because of the reduced amount of data and parametrization required. Although SVR is relatively new, it has become more popular in hydrologic studies for its increased performance when compared with other methods such as artificial neural networks (ANN) (Asefa et al., 2006; Raghavendra and Deka, 2014; Yu et al., 2006). The ability of SVR to reduce both the empirical risk and the predictive function confidence interval and its similarity with a physical model, make this method one of the most robust for the prediction of hydrologic processes.

The use of SVR requires the modeler to have a good understanding of the inputs and their relationship with the variable under prediction. A set of inputs that has not been revised and selected can generate large prediction errors and provide inaccurate estimations. Preprocessing the data and performing a sensitivity analyses to evaluate how different input variables influence the SVR modeling performance can ensure strong causal relationships between the inputs and outputs of interest. Some of the benefits of preprocessing and selecting the input variables include: 1) more evident relationships between the inputs and predicted variable, thereby reducing the storage - machine requirements, 2) minimized training time, and 3) improved prediction performance (Hwang et al., 2012; Wang et al., 2006). The main goal is to create a set of predictors good enough to generate an accurate response.

Several studies have explored the application of SVR for groundwater level predictions (Asefa et al., 2004; Behzad et al., 2009; Jin et al., 2009; Shiri et al., 2013; Sudheer et al., 2011; Yoon et al., 2011). However, there are still many research gaps in applying these machine learning techniques to determine the set of input variables that provides the most efficient groundwater prediction. Although there are studies that evaluate the effect of input variables in hydrologic applications, the assessment of these inputs for groundwater level prediction is very limited. Noori et al. (2011) used 18 input variables for the prediction of monthly streamflow with an SVR approach, and found that only six input variables were the most relevant for the SVR prediction based on principal component analysis (PCA) and gamma test (GT) calculations. Also, Huang and Dun (2008) evaluated the particle swarm optimization method to optimize the selection of inputs and kernel parameters with SVR approaches. The authors emphasized the importance of selecting input variables and parameters for an efficient SVR model prediction.

For the study of groundwater levels, in which the number of input variables is reduced, it is important to determine the adequate set that provides the most reliable prediction. Although the selection of inputs should be performed for each case study separately, the initial assessment of variables that influence the model prediction can be generalized for each hydrologic process. A study on groundwater level forecasting conducted by Nayak et al. (2006) found that precipitation, groundwater level, and canal releases were the most important variables for predicting groundwater levels. For this case, the expert selection of inputs played a fundamental role in the estimation of an optimal model performance. However, not all input variables available to include in the model have a significant effect on the final prediction. Thus, evaluating the set of inputs that provides the highest SVR performance will contribute to model optimization by reducing the training process and cost of data collection.

To that end, the objective of this study was to assess the most efficient input variable arrangement for the prediction of daily groundwater levels with an SVR technique. This novel study aims to provide a guide for the use of input variables when groundwater levels are predicted by SVR in the southeast region of the United States.

Support Vector Regression

SVR is a machine learning technique that uses robust methods to predict complex trends in the input data. The model structure creates a transfer function that maps the inputs into a high-dimensional feature space, called kernel function, in which a linear regression can be performed to determine any nonlinear distribution (Basak et al., 2007). The simple linear SVR is given by:

$$f(x) = \langle w, x \rangle + b \tag{4.1}$$

Applied to a non-linear function, the SVR function is given by:

$$f(x) = \langle w, \psi(x) \rangle + b \tag{4.2}$$

where x is a vector of input values, w is a vector of weights associated with the inputs in x, b is the function bias, $\langle w, x \rangle$ is the dot product between vectors w and x, and $\psi(\mathrm{x})$ the kernel function. For this non-linear problem, it is possible to minimize the norm $\|\mathrm{w}\|^2 = \langle \mathrm{w}.\mathrm{w} \rangle$ to optimize the function:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{4.3}$$

$$subject\ to \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \qquad\qquad\qquad \ge 0 \end{cases}$$

Equations 4.1 to 4.3 highlight how the SVR model performance is highly dependent on the set of input variables and underscore the importance of evaluating the most influential inputs to optimize the model architecture.

For this study, the SVR structure was set-up by implementing the LIBSVM function as part of the R statistical software package (R Core Team 2012). Within LIBSVM, there are five SVR types and four kernels (linear, polynomial, radial basis function, and sigmoid) that can be selected based on the type of data and the purpose of the model. The selection of the SVR architecture and parameters was based on a review of the literature (Asefa et al., 2004; Bray and Han, 2004; Raghavendra and Deka, 2014; Shiri et al., 2013), and by trial and error until the highest performance was reached. The efficiency of the SVR parameter combinations were evaluated based on the mean squared error (MSE) coefficient of performance and the model's ability to represent the time series general trends. The final architecture was determined by tuning different combinations of kernels and kernel parameters until the lowest MSE was found. Detailed information about the steps and procedures to determine the SVR architecture can be found in Guzman et al. (2015). The final SVR parameter selection and architecture is listed in Figure 4.1.

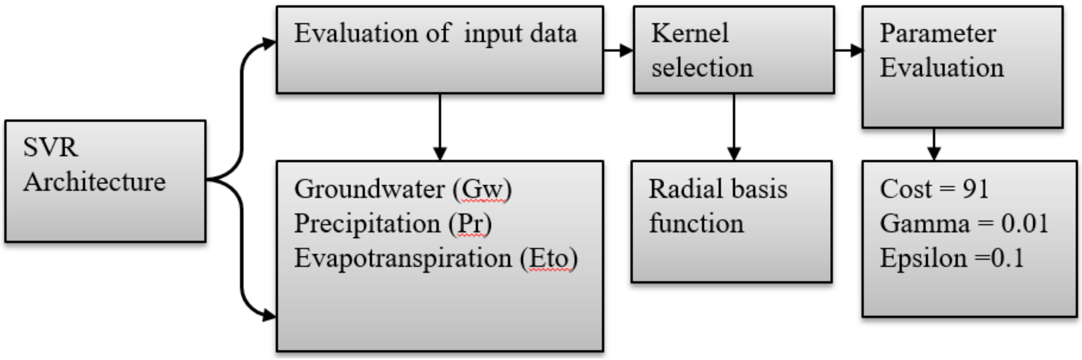

Figure 4.1 Process to determine SVR architecture (upper level) and structure selected for this study (lower level).

Case Study

A U.S. Geological Survey (USGS) monitoring well (M0038) was selected for this study. The well, which is used to monitor groundwater levels of the MRVA, is located in the central part of the Mississippi Delta Region (MDR) in Sunflower county, Mississippi (Figure 4.2). This county is subject to special attention because of the steady decline in groundwater levels over the past several years caused by water withdrawals to support agricultural production. Daily groundwater depths were obtained from the USGS national water information system website (USGS, 2014), and daily mean precipitation was used from a weather station located in Moorhead Mississippi (Menne et al., 2012). In addition, daily evapotranspiration was calculated using the Priestly-Taylor method (Priestley and Taylor, 1972). The daily data collected for this study was available during the time period from 1987 to 1994. The input time series was divided into two periods: from 1987 - 1990 for training and from 1991 - 1994 for testing. The selection of the method to partition the training and testing data was determined from previous studies in groundwater modeling (Coulibaly et al., 2001; Nayak et al., 2006).

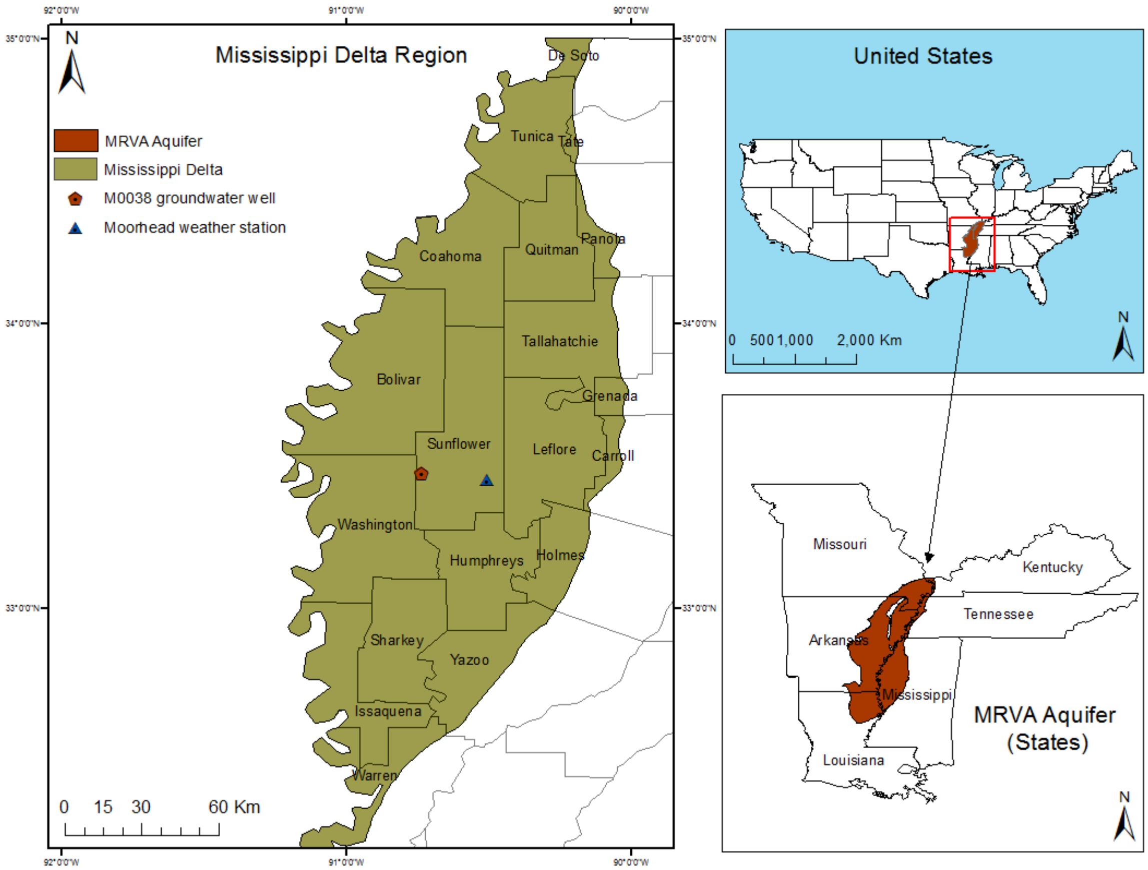

Figure 4.2 Study area and Sunflower well location.

Input Data Description

Data on Daily Groundwater Level

Presently, the evaluation of groundwater levels in the MDR is performed at a seasonal scale. However, various studies have demonstrated that the variability of groundwater levels can range from months to days (Asefa et al., 2007; Coulibaly et al., 2001; Daliakopoulos et al., 2005; Nayak et al., 2006; Yang et al., 2009). This study used daily data which are critical in understanding changes in groundwater levels, especially during periods of high volume withdrawals.

The daily groundwater level time series collected for this study has a sinusoidal behavior that varies from 6.5 m from the ground surface during recharge seasons up to 8.5 – 9 m during times of peak withdrawals. The highest points represent periods of groundwater recharge that usually occur during October and March, and the low sections are the periods when the maximum groundwater demand occurs and usually ranges between March and the end of July. The decline in groundwater levels during the winter months (recharge period) from 1987 to 1994 indicates a reduction in the annual recharge capacity of the well (Figure 4.3a). The highest withdrawals occurred for the summer months in 1990 which shows levels around 9 m under the soil surface. In contrast, for the summer of 1989, the well had one of the lowest withdrawals with levels around 8 m under the soil surface.

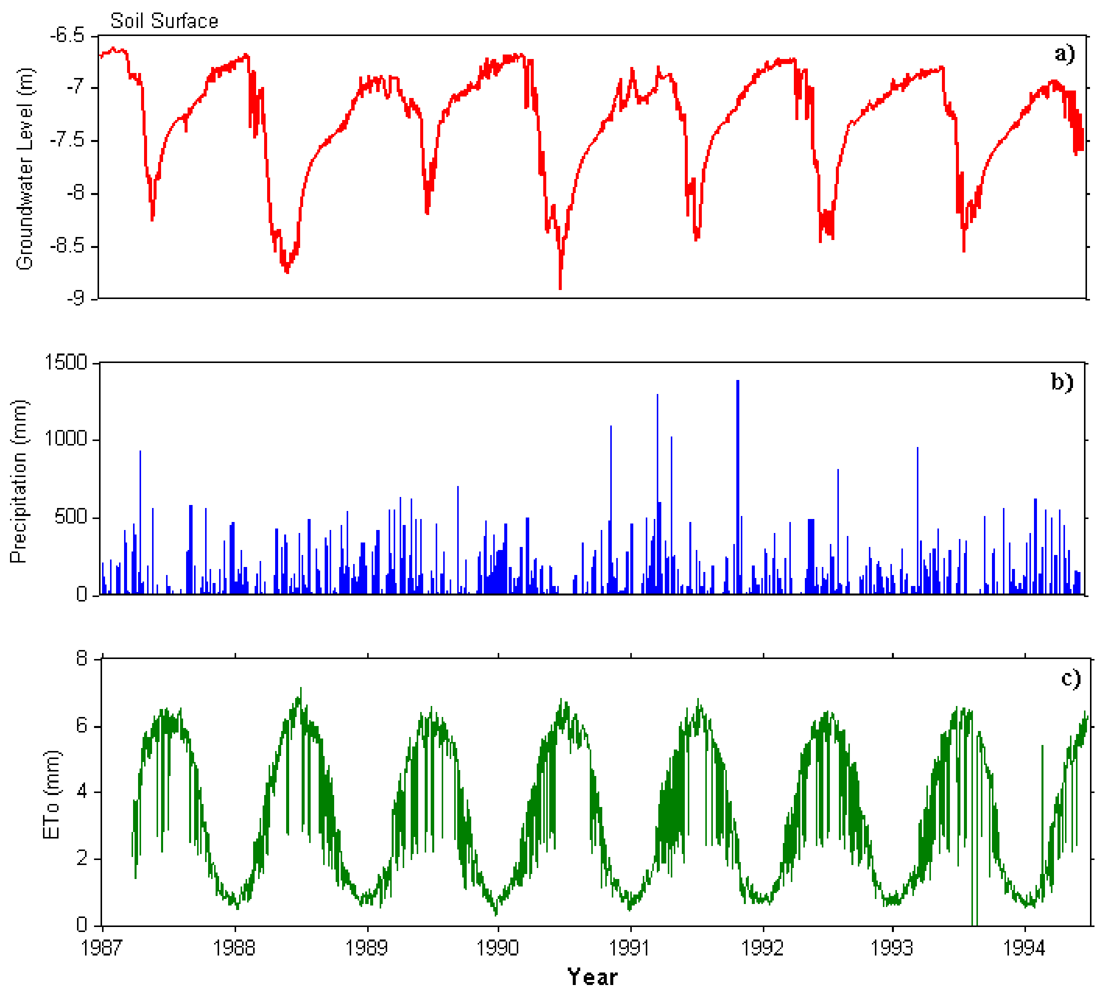

Figure 4.3 Daily data showing a) groundwater level, b) precipitation, and c) calculated evapotranspiration (ET) using Priestly-Taylor method.

Evapotranspiration

Evapotranspiration affects the soil-plant-atmosphere water balance, especially when high temperatures are present. Changes in available soil moisture, crop water demand, and atmospheric conditions necessary for water movement from plants and soils influence the variability of daily evapotranspiration. One of the most efficient methods for evaluating the amount of daily groundwater required for irrigation is through ET-based irrigation scheduling. The method is based on how ET varies during the growing season, and water lost due to ET is replaced by irrigation to satisfy the plant water requirements. Measured ET values were not available from the Moorhead weather station. Thus, daily ET values were calculated using the Priestly-Taylor method (Priestley and Taylor, 1972). This method is one of the most commonly used to calculate evapotranspiration for its reduced number of required parameters and its efficiency when measured data and more complex parameters are not available. The Priestly Taylor function is described by:

$$\Delta ET = \alpha \frac{\Delta}{\Delta + \gamma}(R_n - G) \tag{4.4}$$

where α= 1.26 is a correction factor empirically determined, Δ is the slope of the saturation vapor pressure versus temperature curve, γ is the psychrometric constant, R_n is net radiation, and G is the soil heat flux. This empirical equation relies on the assumption that ET is a function of only solar radiation and temperature. However, continuous daily measurements of solar radiation are difficult to find in the region. Thus, solar radiation was calculated using the WP method (Woli and Paz, 2012), which was designed to generate the most exact estimations of solar radiation for locations within the state of Mississippi. In this study, solar radiation values were used as an input to the Priestly Taylor function to generate the ET time series shown in Figure 4.3c. The average ET values for Sunflower County range between 1.5 mm in winter to 4.8 mm in summer periods with a uniform seasonal trend. However, as shown in Table 4.1, the summer seasons of 1987, 1988 and 1994 have slightly higher average ET values around 5.25 mm.

Table 4.1 Seasonal summary of the average and range of precipitation and ET input variables used in this study.

Year	Statistic	Summer		Winter	
		Pr	ET	Pr	ET
1987	Max	0.93	6.53	0.57	3.75
	Min	0.00	1.88	0.00	0.67
	Avg	0.03	5.27	0.03	1.86
1988	Max	0.42	7.11	0.49	4.32
	Min	0.00	1.72	0.00	0.50
	Avg	0.03	5.21	0.04	1.52
1989	Max	0.63	6.55	0.70	4.32
	Min	0.00	1.12	0.00	0.35
	Avg	0.05	4.83	0.04	1.54
1990	Max	0.50	6.79	1.09	4.10
	Min	0.00	1.76	0.00	0.47
	Avg	0.02	5.17	0.05	1.57
1991	Max	1.29	6.77	1.09	4.26
	Min	0.00	1.48	0.00	0.70
	Avg	0.06	4.76	0.04	1.59
1992	Max	1.29	6.40	1.37	3.97
	Min	0.00	1.83	0.00	0.61
	Avg	0.05	4.86	0.03	1.40
1993	Max	1.09	6.55	1.37	4.21
	Min	0.00	0.10	0.00	0.58
	Avg	0.04	4.56	0.03	1.31
1994	Max	1.29	6.43	1.37	5.40
	Min	0.00	2.21	0.00	0.58
	Avg	0.04	5.24	0.04	1.78

Input Data Arrangements

The SVR model was forced with different arrangements of the groundwater (Gw), precipitation (Pr), and evapotranspiration (ET) input variables for a period of eight years. The simplest setup represents the forcing of the SVR with only one input variable (e.g. only Pr), while the most complex setup represents the forcing of the SVR with all input variables. The setups evaluated in this investigation are summarized in Table 4.2. Each model setup was trained individually, and its performance was evaluated based on the MSE to reproduce observed groundwater levels. Before the time series of input variables was used for the SVR, they were normalized between -1 and 1. The purpose of the normalization is to give equal weight to the input variables regardless of their units.

Evaluation of Performance

To evaluate the model performance, we used a common architecture and training routine for all the setups evaluated in this study. Each setup was evaluated based on its MSE values and the overall capacity to predict the groundwater recharge/withdrawal trends. The performance results for each model setup are shown in Table 4.2. The most efficient input variable arrangement was obtained by the combination of daily groundwater levels and precipitation with an MSE of 0.00123 m, followed by precipitation only and groundwater only with an MSE of 0.10572 m and 0.21058 m, respectively. Based on the results of model setups 3, 5, 6 and 7, the effect of ET on the modeling performance was negligible for the SVR training process (Table 4.2). For all the input arrangements with ET included, the MSE was higher, especially for the Gw+ET and Gw+Pr+ET setups. The results of this study show that ET is not a significant input in the prediction of daily groundwater levels for the humid subtropical climate of the southeastern United States, and the ET input variable does not influence changes in groundwater levels. Cooper et al. (2006) and Taormina et al. (2012) demonstrated the significance of ET as an input variable for studies in arid regions and discussed the importance of evaluating this input for different soil types and climatic conditions.

Table 4.2 SVR model performance under different input arrangements.

Model	Arrangement	MSE
Setup 1	Gw	0.21058
Setup 2	Pr	0.10572
Setup 3	ET	0.22274
Setup 4	Gw+Pr	0.00123
Setup 5	Gw+ET	6.70807
Setup 6	Pr+ET	0.39035
Setup 7	Gw+Pr+ET	4.70044

Prediction of Groundwater Trends

The SVR model predictions of daily groundwater levels for a selected group of arrangements that provided the highest modeling performance are shown in Figure 4.4 to Figure 4.7. The figures present the testing results of more than 1200 days from 1991 to 1994. There are four drawdowns which correspond to periods of groundwater withdrawal, and three main highs which describe the recharge periods. In general, the SVR model prediction with Gw as the only input variable (Figure 4.4) was able to capture the variations in seasonal and yearly groundwater levels. The model was able to predict the groundwater withdrawals, especially for the first and third drawdown (around 100 and 900 days, respectively). However, the model was unable to adequately reproduce the second drawdown (around 400 days) with a difference of approximately 0.4 m between observed and predicted levels. Additionally, the prediction has a slight delay of approximately 10 days that could indicate the need for a different modeling architecture. In contrast, the output response of the SVR trained with Pr as the sole input variable showed difficulties capturing the withdrawal seasons for the entire testing period (Figure 4.5). Although the MSE for the Pr only input is relatively small, the model is not able to capture the withdrawal trends for the testing period. The prediction trend shows the effect of Pr in the variability of groundwater levels. As mentioned above, the highest amounts of Pr occur during the winter periods, and the lowest during the summer, which is visible in the SVR prediction.

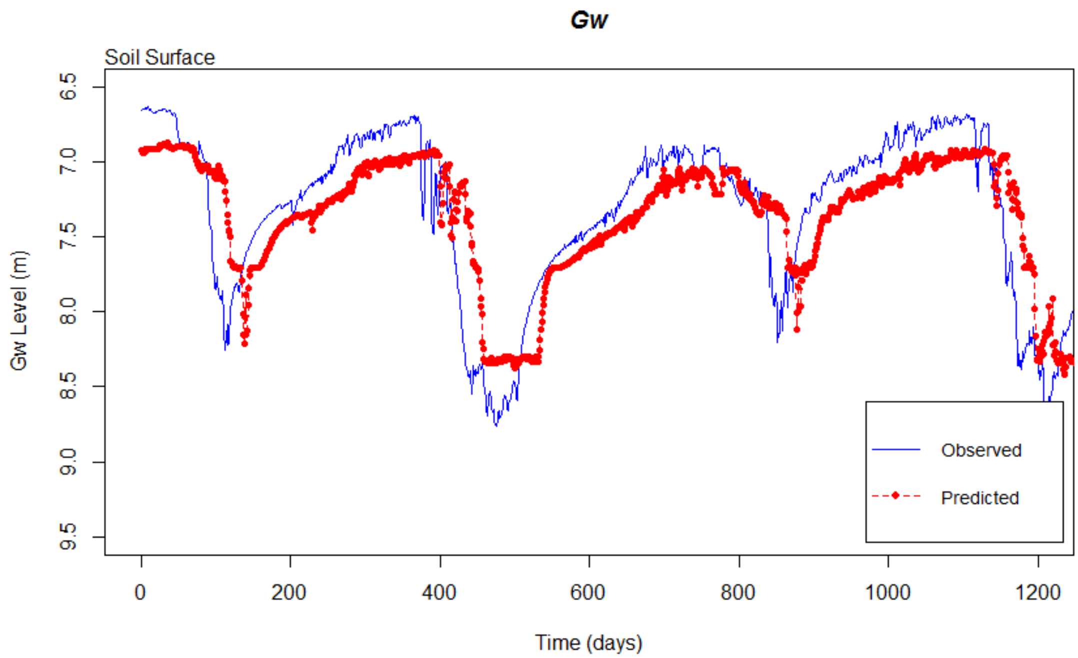

Figure 4.4 SVR daily groundwater level prediction with historical daily groundwater levels (Gw) as input.

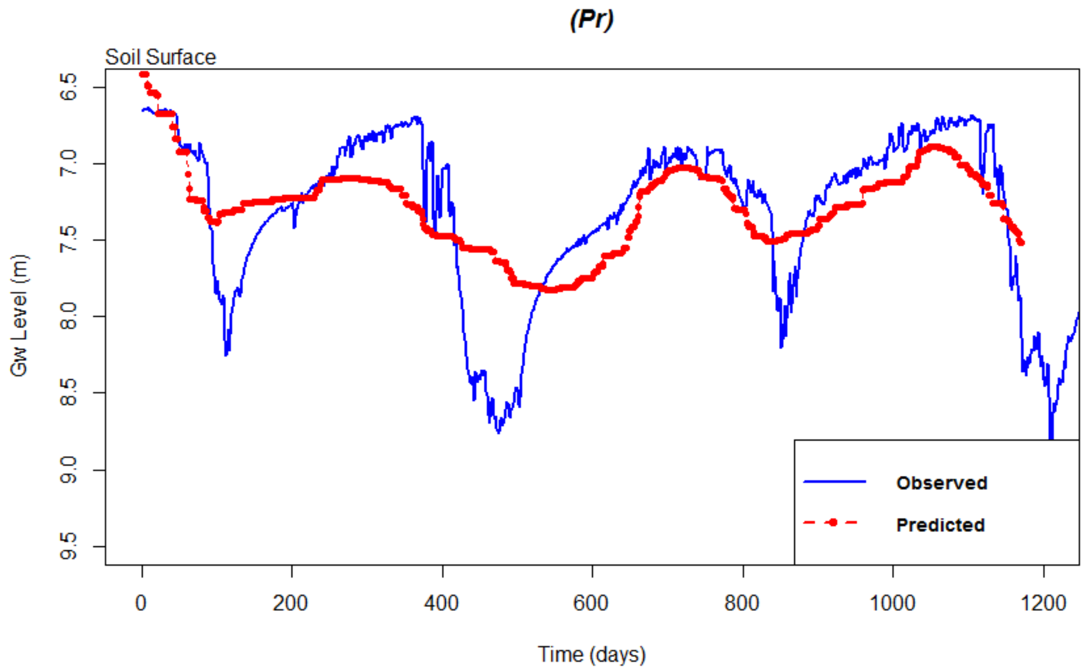

Figure 4.5 SVR daily groundwater level prediction with daily precipitation (Pr) as input.

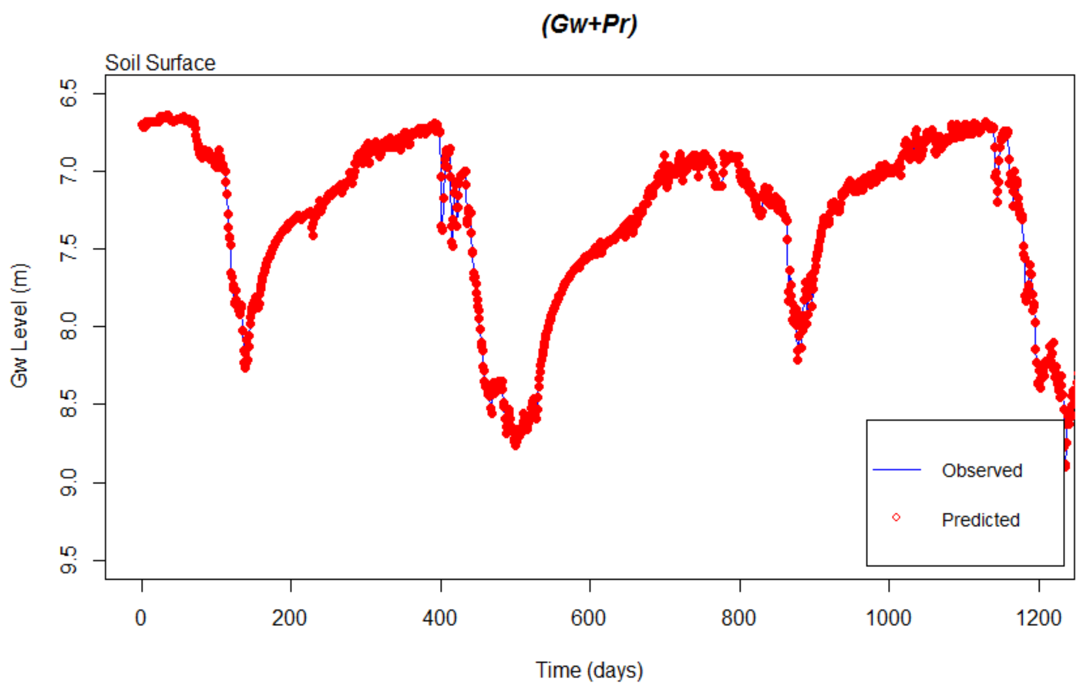

Figure 4.6 SVR daily groundwater level prediction with daily groundwater and precipitation (Gw+Pr) as inputs.

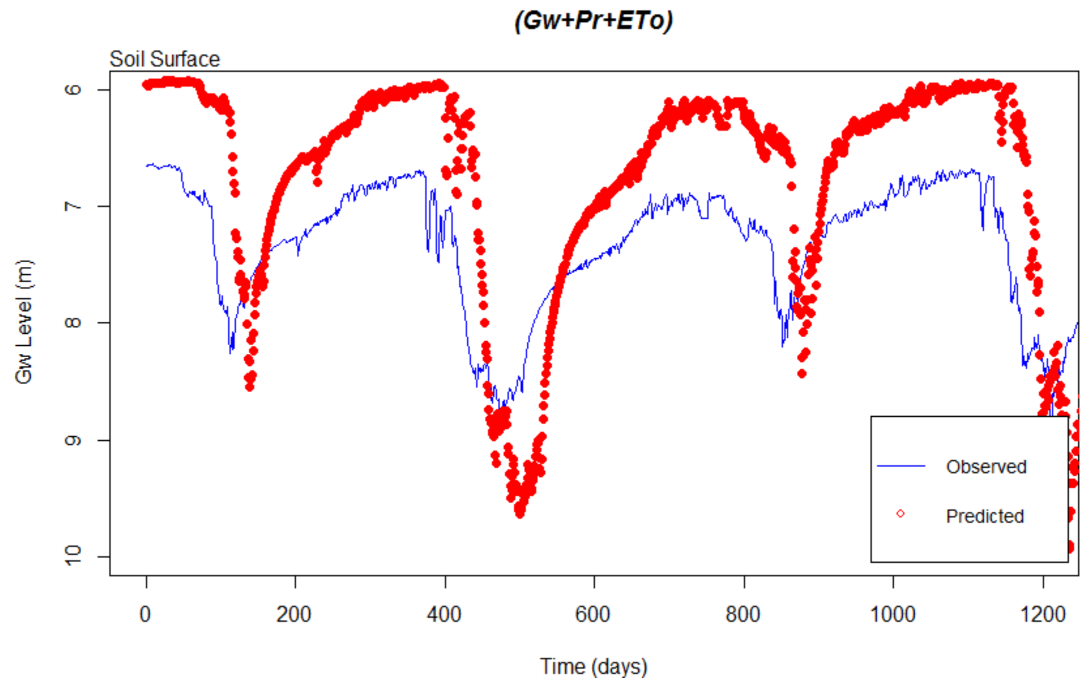

Figure 4.7 Groundwater, precipitation and evapotranspiration (Gw+Pr+ET).

Discussion

The evaluation of input data layers is important in any modeling approach that attempts to optimize a prediction of hydrological processes. It is always necessary to preprocess the data to understand the relationship of the input variables with the variable under prediction and to detect possible errors that can affect the model estimation. The management of input variables is especially important for data-dependent models such as SVR. In this study, daily groundwater levels were predicted with seven arrangements of input variables in which three of them, Gw+Pr, Pr, and Gw, provided good predictions based on the MSE performance. The model performed poorly with the addition of ET as the third input variable (Gw+Pr+ET). ET had a reduced effect on the recharge and withdrawal trends of groundwater for this well. However, it is possible that ET may have a different impact on the SVR model performance if combined with other variables such as crop water requirement or irrigation pumping rates.

From the results shown in Table 4.2 and Figure 4.4, it is evident that the arrangement with Gw+Pr provides the most reliable SVR prediction of daily groundwater levels for the USGS monitoring well in Sunflower county, Mississippi. The model's capability to predict the withdrawal and recharge patterns throughout the year are advantageous for the development of water management and irrigation plans at the local scale. The results from this study can be used as a baseline to construct cost effective data-driven models to forecast groundwater levels in individual wells around the Delta. However, it is important to highlight that SVR should be calibrated for each individual case study for further analysis of multiple wells around the region, as the model is based on data and does not account for the regional aquifer dynamics.

Figure 4.4 shows the groundwater level predictions of the SVR model under different input arrangements. The results show the applicability of this study for different prediction scenarios when the availability of input variables is limited. Therefore, the decision maker has the flexibility to choose between different input arrangements and evaluate which scenario is most suitable to generate a prediction for practical purposes, taking into account the errors associated with each input option. This study shows that the model with Gw+Pr input arrangement gave the best prediction of groundwater levels. However, when Pr is not available for the model, the user can still use Gw as an input variable and obtain acceptable predictions with an error of 0.21058 m. This means an additional error of 0.20935 m as compared to the MSE = 0.00123 m for Gw+Pr prediction, which is still relatively small and will provide practical information to evaluate groundwater withdrawal plans to minimize severe level declines. The results provided in this study contribute to the understanding of the groundwater level variation at the local scale, which is useful for the development of conservation and management strategies for water use at the farm level.

Conclusions

The simulation of groundwater processes is a difficult task due to the dynamic interactions with other natural processes such as surface water, geology, and the uncertainty of water movement in the subsoil. For management purposes, it is increasingly important to have new tools to predict the dynamic behavior of groundwater levels, and to evaluate future impacts on the aquifers. SVR is an important modeling tool that provides an efficient alternative to predict daily levels at the local scale. This tool can be used to make predictions without the complete knowledge of all processes driving the groundwater movement. Additionally, after the parameters are optimized, the function can be easily run for other locations with fast and accurate results. The purpose of this study was to evaluate different arrangements of input variables in the SVR model performance for the prediction of daily groundwater levels in a well that is part of the MRVA aquifer in the Mississippi Delta region. The results of this study suggest that the arrangement of Gw+Pr time series provide the optimal input set for predictive SVR modeling of groundwater levels. The addition of ET as a third input variable reduced the predictive performance of the groundwater levels. The effect of ET on SVR modeling predictions may have a different outcome if the input variable can be combined with additional input variables other than those shown in this paper, such as crop water requirement or daily irrigation pumping rates.

If the decision maker has not availability of the necessary inputs for the SVR prediction, for practical purposes, it can be selected individual setups such as Pr only or Gw only to generate a prediction of groundwater levels considering the reduction of modeling performance that produces the use of individual inputs. Although the MSE increase when Pr and Gw are used individually, in terms of the evaluation of groundwater level variability these errors are acceptable to generate general conclusions about the management of the groundwater well.

The SVR was able to predict the peaks and lows of groundwater level and the trends of the time series data, considering the complexity of the nonlinear groundwater system. Although SVR is a relatively new machine learning technique, it shows efficient results for groundwater daily level prediction in the MDR. The approach presented in this study can be used as a baseline for further studies in the region and to replicate the SVR modeling procedure other wells with similar conditions. For further applications of SVR in groundwater studies, it is recommended that an evaluation be conducted on the sensitivity of input variable prior to model training. This study offers an initial guideline on the selection of input datasets to reproduce groundwater levels. The results provide meaningful insights on the general processes required in the training of SVR and parameters that can be compared with other cases.

References

Asefa, T., Kemblowski, M., McKee, M., Khalil, A., 2006. Multi-time scale stream flow predictions: The support vector machines approach. Journal of Hydrology, 318(1–4): 7-16. DOI:http://dx.doi.org/10.1016/j.jhydrol.2005.06.001

Asefa, T., Kemblowski, M.W., Urroz, G., McKee, M., Khalil, A., 2004. Support vectors–based groundwater head observation networks design. Water Resources Research, 40(11).

Asefa, T., Wanakule, N., Adams, A., 2007. Field-Scale Application of Three Types of Neural Networks to Predict Ground-Water Levels1. JAWRA Journal of the American Water Resources Association, 43(5): 1245-1256.

Basak, D., Pal, S., Patranabis, D.C., 2007. Support vector regression. Neural Information Processing-Letters and Reviews, 11(10): 203-224.

Behzad, M., Asghari, K., Coppola Jr, E.A., 2009. Comparative study of SVMs and ANNs in aquifer water level prediction. Journal of Computing in Civil Engineering, 24(5): 408-413.

Bray, M., Han, D., 2004. Identification of support vector machines for runoff modelling. Journal of Hydroinformatics, 6(4): 265-280.

Byrd, C., 2011. Water resources in the Mississippi Delta. Mississippi Department of Environmental Quality Report, Mississippi Department of Environmental Quality, Water Research Scientists Meeting: Water and Fish for Farmers.

Cooper, D.J., Sanderson, J.S., Stannard, D.I., Groeneveld, D.P., 2006. Effects of long-term water table drawdown on evapotranspiration and vegetation in an arid region phreatophyte community. Journal of Hydrology, 325(1): 21-34.

CoreTeam, R., 2012. R: A Language and Environment for Statistical Computing.

Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. Water Resources Research, 37(4): 885-896. DOI:10.1029/2000wr900368

Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. Journal of Hydrology, 309(1–4): 229-240. DOI:http://dx.doi.org/10.1016/j.jhydrol.2004.12.001

Dyer, J., Mercer, A., Rigby, J.R., Grimes, A., 2015. Identification of recharge zones in the Lower Mississippi River alluvial aquifer using high-resolution precipitation estimates. Journal of Hydrology, 531, Part 2: 360-369. DOI:http://dx.doi.org/10.1016/j.jhydrol.2015.07.016

Guzman, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A., 2015. Artificial Neural Networks and Support Vector Machines: Contrast Study for Groundwater Level Prediction Paper 152181983, 2015 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, New Orleans, Louisiana.

Huang, C.-L., Dun, J.-F., 2008. A distributed PSO–SVM hybrid system with feature selection and parameter optimization. Applied Soft Computing, 8(4): 1381-1391. DOI:http://dx.doi.org/10.1016/j.asoc.2007.10.007

Hwang, S.H., Ham, D.H., Kim, J.H., 2012. Forecasting performance of LS-SVM for nonlinear hydrological time series. KSCE Journal of Civil Engineering, 16(5): 870-882.

Jin, L., Jian-xia, C., Wen-ge, Z., 2009. Groundwater Level Dynamic Prediction Based on Chaos Optimization and Support Vector Machine, Genetic and Evolutionary Computing, 2009. WGEC '09. 3rd International Conference on, pp. 39-43. DOI:10.1109/WGEC.2009.25

Kebede, H., Fisher, D.K., Sui, R., Reddy, K.N., 2014. Irrigation methods and scheduling in the delta region of Mississippi: current status and strategies to improve irrigation efficiency. American Journal of Plant Sciences, 5(20): 2917.

Mainuddin, M., Das Gupta, A., Raj Onta, P., 1997. Optimal crop planning model for an existing groundwater irrigation project in Thailand. Agricultural Water Management, 33(1): 43-62. DOI:http://dx.doi.org/10.1016/S0378-3774(96)01278-4

Maupin, M.A., Barber, N.L., 2005. Estimated withdrawals from principal aquifers in the United States, 2000 [electronic resource] / by Molly A. Maupin and Nancy L. Barber. Circular: 1279. Reston, Va. : U.S. Dept. of the Interior, U.S. Geological Survey ; Denver, CO.

Menne, M. et al., 2012. Global Historical Climatology Network-Daily (GHCN-Daily), Version 3. 2. NOAA National Climatic Data Center.

Nayak, P., Rao, Y.R.S., Sudheer, K.P., 2006. Groundwater Level Forecasting in a Shallow Aquifer Using Artificial Neural Network Approach. Water Resour Manage, 20(1): 77-90. DOI:10.1007/s11269-006-4007-z

Noori, R. et al., 2011. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. Journal of Hydrology, 401(3): 177-189.

Priestley, C., Taylor, R., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly weather review, 100(2): 81-92.

Raghavendra, N.j., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Applied Soft Computing, 19(0): 372-386. DOI:http://dx.doi.org/10.1016/j.asoc.2014.02.002

Scanlon, B.R. et al., 2012. Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. Proceedings of the National Academy of Sciences, 109(24): 9320-9325. DOI:10.1073/pnas.1200311109

Shiri, J., Kisi, O., Yoon, H., Lee, K.-K., Hossein Nazemi, A., 2013. Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques. Computers & Geosciences, 56: 32-44. DOI:http://dx.doi.org/10.1016/j.cageo.2013.01.007

Sudheer, C., Shrivastava, N., Panigrahi, B., Mathur, S., 2011. Groundwater Level Forecasting Using SVM-QPSO. In: Panigrahi, B., Suganthan, P., Das, S., Satapathy, S. (Eds.), Swarm, Evolutionary, and Memetic Computing. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 731-741. DOI:10.1007/978-3-642-27172-4_86

Taormina, R., Chau, K.-W., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Engineering Applications of Artificial Intelligence, 25(8): 1670-1676.

USGS, U.S.G.S., 2014. National Water Information System: Web Interface.

Wang, W., Van Gelder, P.H., Vrijling, J., Ma, J., 2006. Forecasting daily streamflow using hybrid ANN models. Journal of Hydrology, 324(1): 383-399.

Wax Charles L. , Pote Jonathan W, L., M.T., 2009. Climatological and Cultural Influences on Annual Groundwater Decline in the Mississippi Delta Shallow Alluvial Aquifer, Mississippi Water Resources Conference Jackson. Mississippi.

Woli, P., Paz, J.O., 2012. Evaluation of Various Methods for Estimating Global Solar Radiation in the Southeastern United States. Journal of Applied Meteorology and Climatology, 51(5): 972-985. DOI:10.1175/JAMC-D-11-0141.1

Yang, Z.P., Lu, W.X., Long, Y.Q., Li, P., 2009. Application and comparison of two prediction models for groundwater levels: A case study in Western Jilin Province, China. Journal of Arid Environments, 73(4–5): 487-492. DOI:http://dx.doi.org/10.1016/j.jaridenv.2008.11.008

Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. Journal of Hydrology, 396(1–2): 128-138. DOI:http://dx.doi.org/10.1016/j.jhydrol.2010.11.002

Yu, P.-S., Chen, S.-T., Chang, I.F., 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology, 328(3–4): 704-716. DOI:http://dx.doi.org/10.1016/j.jhydrol.2006.01.021

CHAPTER V

AN INTEGRATED SVR AND CROP MODEL TO ESTIMATE DAILY GROUNDWATER LEVEL IMPACTS OF SOYBEAN IRRIGATION DEMANDS

A paper to be submitted to the Agricultural Systems Journal
Sandra M. Guzman, Joel O. Paz, Mary Love M. Tagert, Mercer. A.

Abstract

As groundwater resources are used more intensively, the need to define appropriate strategies to plan and manage irrigation systems under diverse environmental conditions becomes increasingly important. To promote more efficient irrigation practices, accurate and optimal information regarding the interaction between crop water use and groundwater sustainability is needed. In this study, we outlined a modeling approach that combines the features of a crop growth model and a support vector regression (SVR) model for the comprehensive assessment of groundwater variability under different soybean (Glycine max. [L.] Merr) irrigation thresholds throughout the growing season. The 20%, 40%, 50% and 60% thresholds of water available were calibrated using the CROPGRO-Soybean model to simulate daily irrigation requirements of soybeans grown in the Mississippi Delta Region (MDR). The daily crop water requirements along with precipitation and daily groundwater levels from 1985 to 1994 were used as inputs in the SVR to evaluate the predicted response of daily groundwater

levels to different irrigation demands. We examined the performance of the SVR model based on the Mean Squared Error (MSE) and its ability to capture the seasonal variability in groundwater levels under different scenarios. Results demonstrate that an increase in the volume of water applied through irrigation did not translate into a significant increase in soybean yields. In addition, this increased volume has a short term effect on the changes in groundwater levels. Finally, we conclude that the linked crop-SVR model is able to assess the demands on groundwater supplies for irrigation and is able to provide useful information for decision making.

Introduction

In recent decades, the sustainability of water resources and the expanding drawdown in water tables, especially in regions affected by changes in climate, has been an increasing topic of concern (Hook et al., 2009). For the agricultural community that relies on groundwater sources, the availability of groundwater for irrigation is a subject that has received special attention, and the development of management strategies that provide a balance between groundwater sustainability and crop production is required. Over the years, various strategies have been evaluated to optimize irrigation water management practices (Bruns et al., 2003; Carruth et al., 2014; Graterol et al., 1993; Karam et al., 2005; Sassenrath et al., 2013). However, the combined evaluation of the relationship between crop yields and the potential impacts to groundwater sources is limited. In addition, reliable data that serves the decision making process is difficult to obtain, and the information collected through field data is restricted. Understanding and evaluating yield variability under different levels of water availability has become an important research area in the Southeast US, especially for highly commercialized crops such as soybean (Heatherly et al., 1990; Hook, 1994; Jones et al., 1998; Kebede et al., 2014; Salazar et al., 2012).

The integration of crop and data-driven models is a valuable approach for assessing the impacts of the irrigation demands on crop production and on the sustainability of groundwater sources (Holzworth et al., 2015; Ojha et al., 2013; Yang et al., 2014). These models serve the following three purposes: 1) to assess present and future scenarios of environmental and plant physiology conditions for the improvement of crop production, 2) to evaluate options for the integral management of water for irrigation, and 3) to help develop more comprehensive groundwater studies at the local scale and predict water variability. Crop and irrigation scheduling models have been used in Northwest Mississippi (Sassenrath et al., 2013). However, these models have been used individually and not as an integrated modeling tool to simulate crop production, groundwater variability, and water balance simultaneously at a local scale. There is a critical need for additional work on the integration of crop and data-based models to have more realistic tools to support water management activities in Mississippi and in the southeastern US.

In this study, we propose a parsimonious model that is able to integrate crop water needs and changes in groundwater levels at a local scale in response to irrigation management scenarios. The integrated model combines a crop and a support vector regression (SVR) model that can be used as a farm scale planning tool to better manage groundwater withdrawals and optimize crop production. The application of integrated machine learning approaches, such as SVR, and physical models is fairly well known in the atmospheric sciences (Chevalier et al., 2011; Mercer et al., 2008). However for agricultural studies, the application of SVR is relatively new. Navarro-Hellín et al. (2016) used machine learning techniques to develop a decision support system for the timing of irrigation sets. This model used soil and weather inputs to train and test an adaptive neuro fuzzy inference system (ANFIS) and a partial least square regression (PLSR) model to generate irrigation reports that were validated against the decisions that the irrigation manager took. Their research was based on determining the amount of water required by crops. Similarly, crop models have been used to analyze the interaction between water resources and crop production (Garcia y Garcia et al., 2010; Hook, 1994; MacRobert and Savage, 1998). Paz et al. (1998) used the CROPGRO-Soybean model to evaluate how different water stress conditions affected the yield of various soybean fields in Iowa. In addition, Salazar et al. (2012) implemented the DSSAT CERES-Maize model to predict the amount of water required to irrigate maize crops in south Georgia. Given the factors that could affect crop yields and the availability of groundwater sources for irrigation, it is clear that the implementation of new modeling approaches offers an advantage for the optimal use of limited water sources and more efficient irrigation scenarios.

The objective of this study is to provide a modeling tool that links the CROPGRO-Soybean and SVR models to assess the impacts of different irrigation management scenarios on the daily variability of groundwater levels at a local scale. We implemented the integrated model and applied it to two soybean farms and one groundwater well located in northwest Mississippi.

Materials and Methods

DSSAT Description

The Decision Support System for Agrotechnology Transfer (DSSAT v4.6) is a comprehensive modeling framework for biophysical modeling that integrates more than 28 cropping systems (Hoogenboom et al., 2015; Jones et al., 2003; Jones et al., 1998). DSSAT simulates crop growth, yield and water demands in response to physiological, climatic, soil, and management conditions. This modeling framework has been evaluated for diverse environments and crops to predict yields, water use, and crop decision strategies around the world (Batchelor et al., 2002; Liu et al., 2011; Salazar et al., 2012; Thorp et al., 2008; Yang et al., 2006).

DSSAT requires weather inputs, crop and soil data, and information related to agricultural practices such as timing of planting and harvest for the study area. The most important weather information required by the model includes solar radiation, temperature, and precipitation. The required crop and soil data include plant genetic characteristics, maturity group, root soil growth, soil type and physical properties such as field capacity, permanent wilting point, water content and soil layers (Jones et al., 1998). The data related to agricultural practices include planting and harvest dates, plant population, and crop configuration among others (Figure 5.1).

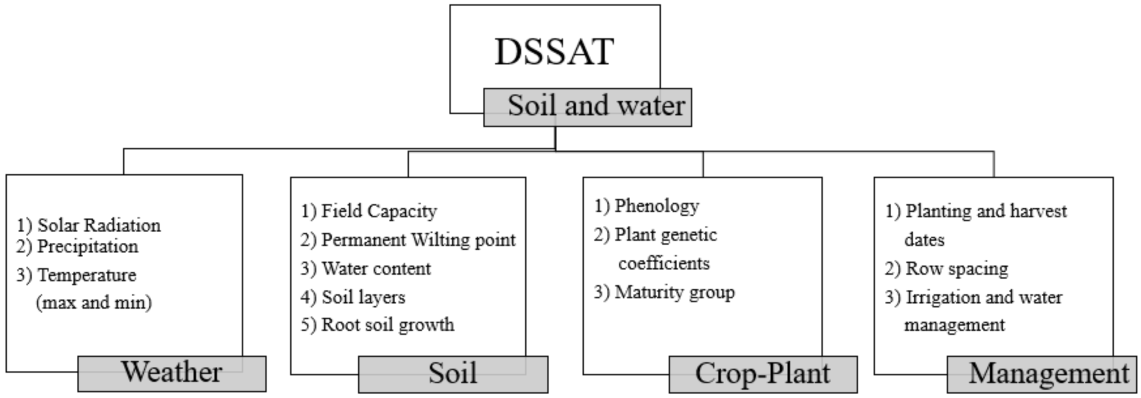

Figure 5.1 General DSSAT structure used for this study.

Available Water

DSSAT has a user-specified irrigation threshold to activate or deactivate irrigation periods. The water thresholds are determined based on the soil moisture conditions and reflect different levels of soil water available for plant development. . The main conditions include 1) field capacity (FC), which is the upper limit of water available to the plant after the excess water has drained; 2) permanent wilting point (PWP), which is the lower limit of soil moisture that a plant can absorb before complete physiological damage; and 3) available water (AW), which is the difference between FC and PWP. The AW represents the water that plants can absorb for optimal physiological development (Brouwer et al., 1985) and is defined by:

$$AW = (FC_v - PWP_v)\,{}^{D_r}/_{100} \tag{5.1}$$

where FC_v and PWP_v are field capacity and permanent wilting point, respectively, in volumetric units, D_r is the depth of the root zone, and AW is expressed as the depth of available water. The total AW represents the maximum amount (100%) of water that the soybean crop can take from the soil.

CROPGRO-Soybean Setup

For the crop simulations, soybean yield data and crop management information were obtained from the Mississippi State University (MSU) soybean variety trials conducted from 2010 to 2013 (Burgess et al., 2010; Burgess et al., 2011; Burgess et al., 2012; Burgess et al., 2013). Detailed information included soil characteristics, planting and harvest dates, rainfall patterns, plant population, crop configuration and management practices. The following inputs were included for each year in the simulations: planting rate equal to 27 seeds per meter of row for 0.76 m row spacing, and irrigated plots with three rows spaced 0.48 m apart. The crop management conditions collected for this study were assumed to be representative for the MDR.

CROPGRO-Soybean Model Calibration

The CROPGRO-Soybean model was calibrated by comparing four years of simulated versus measured values of crop yield and height from the Mississippi State University (MSU) - Delta Branch Experiment Station (DBES) state variety trials. Measured yield and plant height for two Roundup Ready commercial soybean varieties namely, Asgrow AG4730 and AG4831, were used to evaluate the model predictions. The calibration was conducted using an irrigation threshold of 50%, which represents the regular irrigation conditions present in the region. The results from this comparison were used to generate nine years of daily soybean water requirements that served as input variables for the groundwater SVR model.

The weather, soil, and plant genetic inputs were preprocessed and arranged to be included in CROPGRO-Soybean as the base conditions for crop model simulations (Jones et al., 2003). Daily precipitation, temperature, and solar radiation from the

Stoneville weather station were collected from the MSU Delta Research and Extension Center (DREC) Agricultural Weather website (DREC-MSU, 2015). The missing solar radiation values from the DREC site were determined using the WP method (Woli and Paz, 2012), which was designed to calculate solar radiation values for the state of Mississippi. Soil profile information was obtained from the USDA-NRCS web soil survey (USDA-NRCS, 2015).

The soybean genetic coefficients were selected from the default DSSAT database. Parameters such as soil root growth factor and the critical short day length below which reproductive development progresses with no day length effect, were adjusted to obtain the model configuration that could describe the processes observed on the field and minimize the error between measured and predicted soybean yield.

In this study, the CROPGRO-Soybean model used modules for initial conditions, soil analysis, cultivar, planting, and harvest. For the initial conditions, we set the start measurement date at planting and soybean as the crop grown the previous year. The Sharkey clay and Tunica clay loam were selected for the soil analysis module. These soil types had a set of layers with depths from 100 to 200 cm below the surface. Initial conditions for soil nutrient concentrations such as ammonium (NH4) and nitrate (NO3) content were calculated by default with values of 0.1 g [N] Mg^{-1}[soil] and 1.1 g [N] Mg^{-1} [soil], respectively. The cultivars tested for this study were late maturity group IV for the variety AG4730, and early maturity group V for the variety AG4831. The planting and harvest dates for each farm were included in the CROPGRO- Soybean planting module as shown in Table 5.1. Additional information specified in the module included dry seed as the planting method, planting distribution in rows, and row direction by default.

Table 5.1 Field conditions for the farms analyzed in this study.

	Field conditions	Stoneville	Clarksdale
2010	Soil type:	Sharkey clay	Tunica clay loam
	Previous crop:	soybean	soybean
	Planting date	28-Apr	07-May
	Harvest date	21-Sep	21-Sep
2011	Soil type:	Sharkey clay	Tunica clay loam
	Previous crop:	soybean	soybean
	Planting date	10-May	11-May
	Harvest date	21-Sep	03-Oct
2012	Soil type:	Sharkey clay	Forest dale silt loam
	Previous crop:	soybean	soybean
	Planting date	07-May	27-Apr
	Harvest date	26-Sep	05-Oct
2013	Soil type:	Sharkey clay	Forest dale silt loam
	Previous crop:	soybean	soybean
	Planting date	30-Apr	15-May
	Harvest date	11-Oct	14-Oct

Irrigation Thresholds

For this study, we used the calibrated CROPGRO-Soybean model to simulate soybean physiological and crop yield responses as well as daily cumulative irrigation requirements (IRRIC) based on four irrigation threshold (ITHRL) scenarios for a period of nine years (1985 to 1994).

To generate the IRRIC, we used four ITHRLs which represent water scarcity (20%), normal available water conditions (40% and 50%), and full water supply (60%). Each threshold represents the limit of available water for which CROPGRO-Soybean simulates an irrigation event. The crop model simulates water balance based on atmospheric conditions and crop water demand, and activates the automatic irrigation period if AW is less than or equal to a set ITHRL (Figure 5.2).

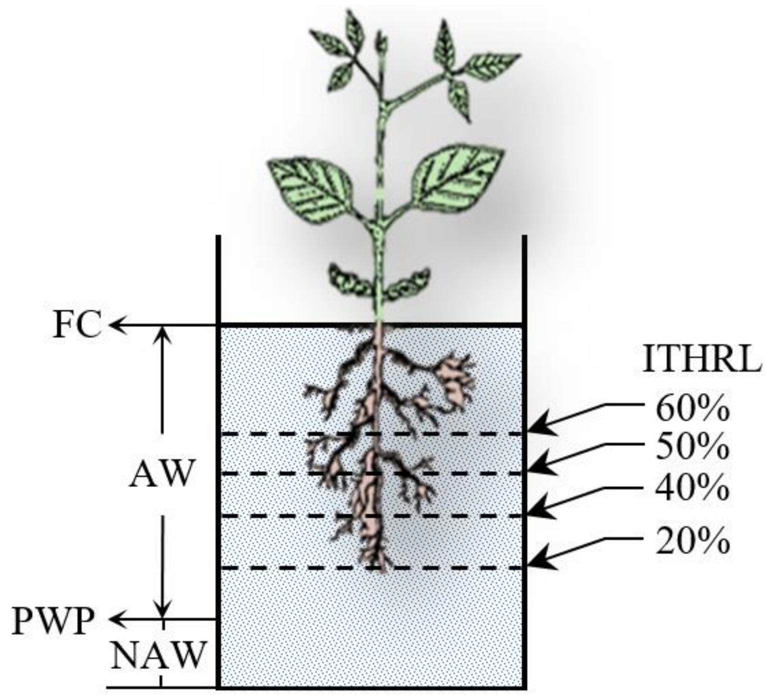

Figure 5.2 Description of the available soil moisture and irrigation thresholds used for this study.

FC = field capacity, PWP = permanent wilting point, AW= water available, NAW = non-available soil water, ITHRL= irrigation threshold

Study Area

The Mississippi Delta region (MDR) is located in the northwest part of Mississippi, and it is characterized for its high production of crops such as cotton, maize, and soybean. Most of the water for irrigation comes from the Mississippi River Valley Alluvial (MRVA) Aquifer, which is one of the most highly used aquifers in the US (Maupin and Barber, 2005). This aquifer represents the most important source of water in the MDR because of its easy accessibility for irrigation.

The CROPGRO-Soybean model was implemented on two different farms located within the MDR, namely the DBES in Stoneville, MS, and the Dulaney Farm in Clarksdale, MS (Burgess et al., 2010). These two farms were selected because they are located in counties where the production of soybean is extensive (NASS, 2003). In addition, a groundwater monitoring well (USGS well code M0038) located in the central part of the MDR in Sunflower County, Mississippi was selected for the SVR prediction (Figure 5.3). The well is used to monitor groundwater levels in the MRVA (USGS, 2014). This county has one of the most critical zones of groundwater depletion in the region. Finally, weather information was collected from the MSU Delta Research and Extension Center (DREC) Agricultural Weather website (DREC-MSU, 2015) and the Global Historical Climatology Network-Daily (GHCN-Daily), Version 3.2 Moorhead weather station (Menne et al., 2012).

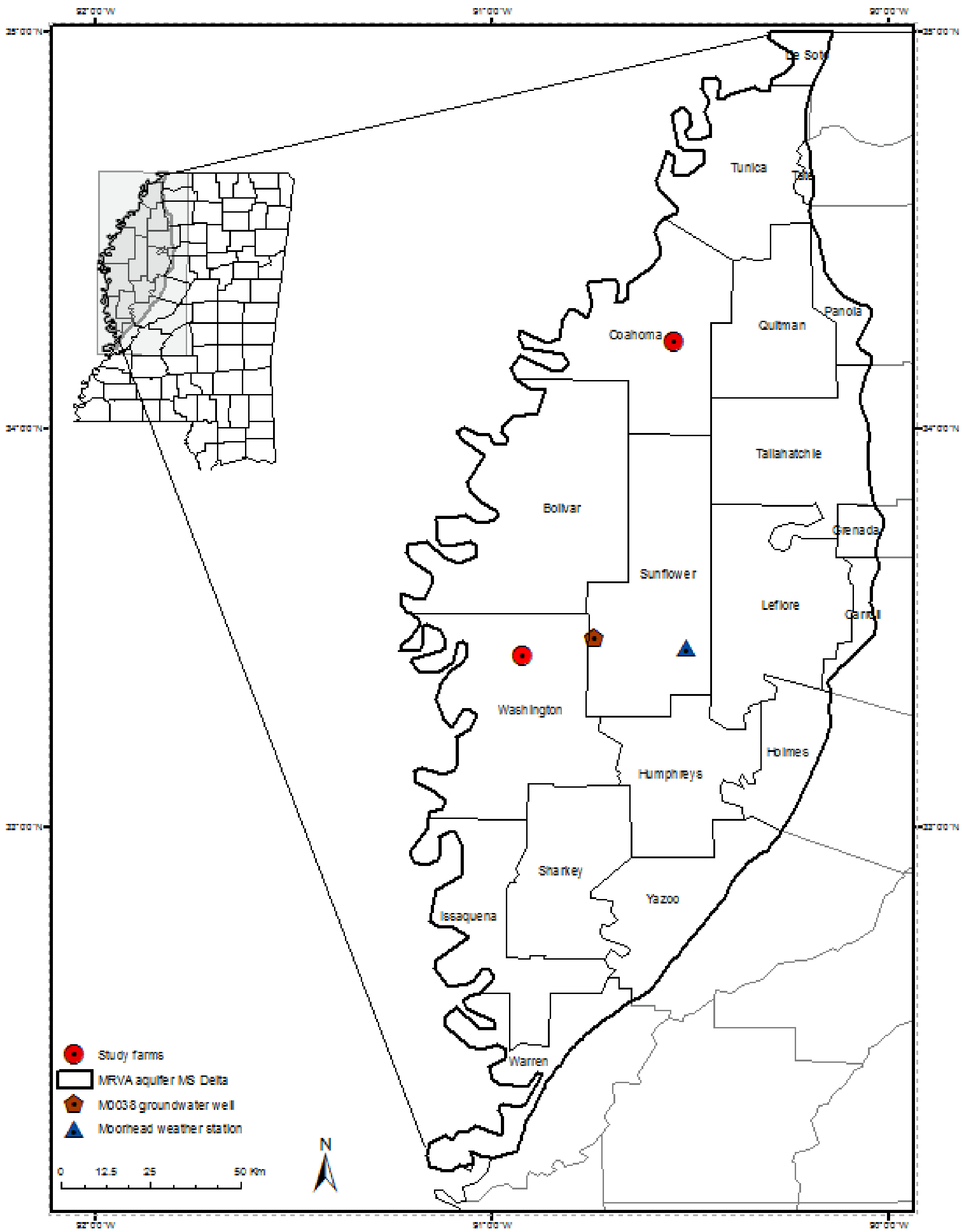

Figure 5.3 Map showing the MDR and locations selected for this study.

Support Vector Regression

The support vector machines (SVM) were introduced by Cortes and Vapnik (1995) to solve classification and regression problems for pattern recognition. The subdivision of SVM that solves regression problems is known as support vector regression (SVR), and is used mostly to characterize time series data. For SVR, the main objective is to find a hyperplane function able to recognize the pattern that generates a specific time series. SVR provides a linear function plus a coefficient called the "kernel" which contains the hyperplane generated and also a confidence interval, called the epsilon intensive band (ε), which is the limit in which the prediction can deviate from the observation (Equation 5. 3). The values that are overlapping the ε boundaries are called the support vectors, and the values that are outside of ε are considered as errors ξ, ξ_i^*. A general description of a non-linear SVR is shown in Figure 5.4.

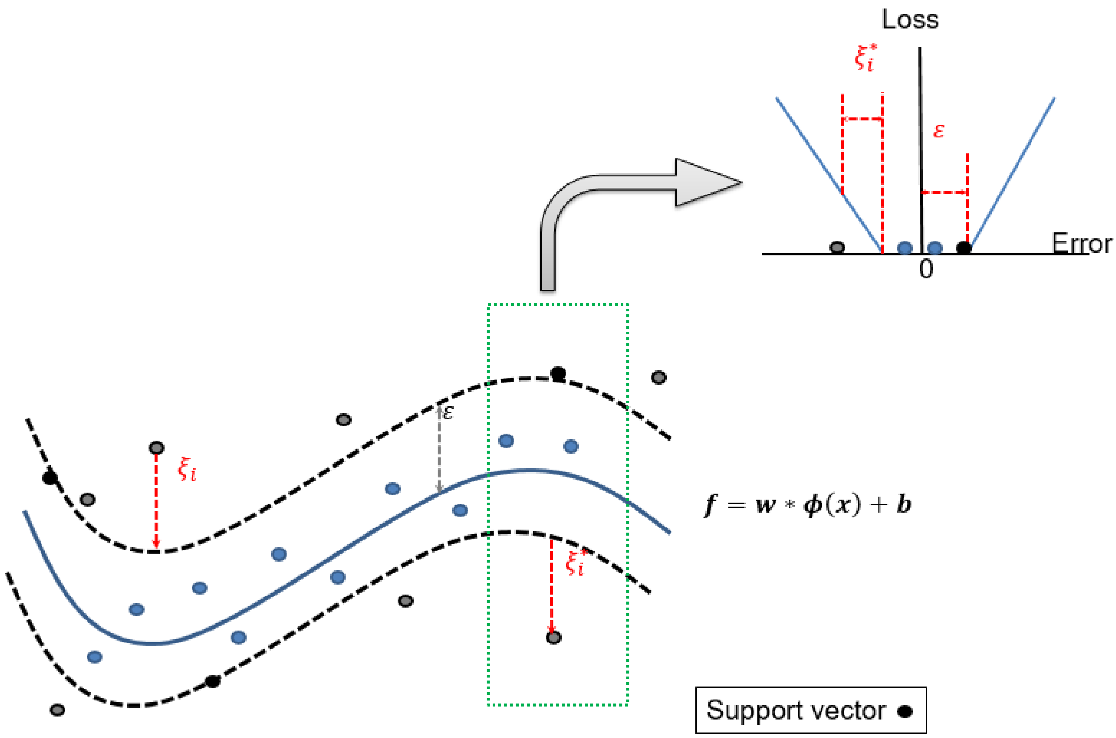

Figure 5.4 Overview of a non-linear SVR.

SVR uses structural risk minimization, which minimizes both empirical risk and the model complexity simultaneously, instead of the empirical risk used in methods such as artificial neural networks. When the structural risk is minimized, the model is able to optimize its data generalization capabilities. A model that generalizes well is able to correctly predict if additional new examples are introduced. The general SVR is described as:

$$f(x) = \langle w, \psi(x) \rangle + b \tag{5.2}$$

where:

w = vector of weights associated to the inputs

x = vector of input values

$\psi(x)$ = kernel function

b = the function bias

For this problem, it is possible to minimize the norm $\|w\|^2 = \langle w.w \rangle$, and to include the ξ, ξ_i^* error terms to optimize the dual function:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{5.3}$$

$$subject\ to \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \qquad\qquad\quad \geq 0 \end{cases}$$

Using the Lagrangian multipliers to optimize the quadratic problem in equation 5.3, the SVR function is transformed into:

$$f(x) = \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)k\langle x_i, x \rangle + b \tag{5.4}$$

where:

α_i, α_i^* = Lagrangian multipliers

$k\langle x_i, x \rangle$ = kernel function.

CROPGRO-Soybean and SVR Model Setup for the Study Area

The linked CROPGRO-Soybean and SVR model was used to simulate the potential impacts of four irrigation management scenarios on crop yield and on groundwater availability. The simulated scenarios evaluated, in particular, how the agro-hydrological system of the MDR may respond if different irrigation thresholds were established and used in the region (e.g. to mimic dry and wet soil water conditions or conservation practices). The daily cumulative soybean water requirements (IRRIC) per scenario were generated from the soil – water module (SoilWat.OUT) of the CROPGRO-Soybean simulation output, with the general calibrated initial conditions shown in Table

5.2. The IRRIC CROPGRO outputs along with the measured daily groundwater levels (Gw) and precipitation (Pr) from 1985 to 1994 were used as inputs to the SVR model for the prediction of daily groundwater levels. A general description of the linked models used for this study is shown in Figure 5.5.

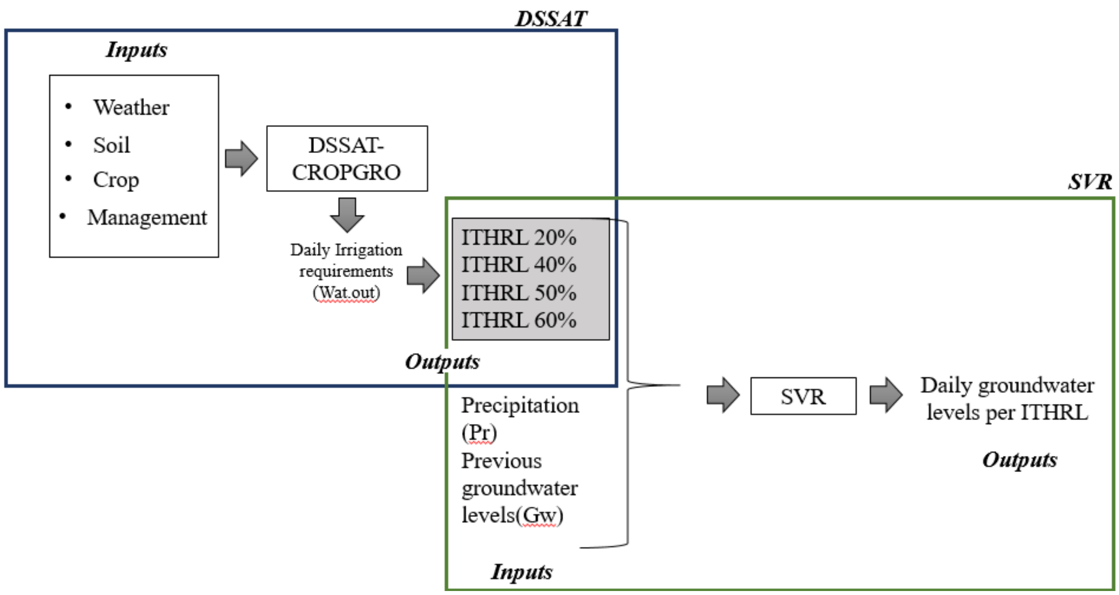

Figure 5.5 Diagram of the inputs and outputs per model for the DSSAT-SVR linked model applied in this study.

Table 5.2 Initial conditions used in CROPGRO-Soybean to generate the daily water requirements for SVR.

Soil Type	Sharkey Clay
Weather Station	Moorhead MS
Planting Date	28-Apr
Plant population (plants/m^2)	23-Jan
Row Spacing (cm)	16-Mar
Planting Depth (cm)	46
Previous crop	Soybean
Cultivar	Late MG IV

In Chapter 4, a sensitivity analysis of the SVR model was performed to determine the optimal set of input data variables needed to construct a predictive model of groundwater levels. We evaluated seven combinations of input variable arrangements that included groundwater, precipitation, and evapotranspiration (ET) to obtain the best SVR architecture for the prediction of daily groundwater levels. The best arrangement of input variables was obtained by combining Gw+Pr as inputs. Thus, in this study we used this combination plus each ITHRL scenario to generate the four input arrangements shown in Table 5.3.

Table 5.3 Input arrangements used for the SVR model.

Model	Arrangement*
Setup 1	Gw+Pr+ITHRL20%
Setup 2	Gw+Pr+ITHRL40%
Setup 3	Gw+Pr+ITHRL50%
Setup 4	Gw+Pr+ITHRL60%

*Where: Gw = Daily groundwater levels, Pr = Precipitation, and ITHRL = 20%, 40%, 50%, and 60% irrigation threshold from the soil water available.

SVR Training and Performance

The SVR structure was set up in the R statistical package (CoreTeam, 2012). The selection of kernels and kernel parameters was initially defined by using literature values from previous studies that used an SVR model to predict hydrological processes, and later refined by trial and error (Asefa et al., 2006; Bray and Han, 2004; Raghavendra and Deka, 2014). The radial basis function (RBF) kernel was selected from four possible kernel functions (linear, polynomial, RBF, and sigmoid), and the RBF parameters were evaluated until the optimal SVR training function was found. More details on the SVR training process and kernel - parameter selection can be found in Guzman et al. (2015). The final SVR configuration for this study is described as follows:

- Support Vectors = 392
- Kernel Function = Radial Basis Function

$$exp(-gamma * |u - v|^2)$$

- $gamma$ = 0.01, u = 0.1, and v = 100

The SVR training was evaluated based on the MSE, coefficient of determination (R^2) and the general trend of prediction. The function with the lowest MSE provides the most efficient SVR performance, while the R^2 reflects the degree of linear correlation between two different datasets or time series. The prediction trend represents the SVR's ability to capture the seasonal changes in the groundwater time series and how groundwater is affected by the inputs. The statistical equations of model performance are defined by:

$$MSE = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N} \tag{5.5}$$

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{5.6}$$

where y_i and $\hat{y}_i$ are the observed and predicted values, $\bar{y}$ is the mean of the observed data, and N is the number of values.

Results and Discussion

CROPGRO-Soybean Calibration Results

The model was able to predict plant growth, development, yield, and water requirements for soybeans grown at two different locations (Figure 5.6). Variables such as plant height (m) and grain weight (kg ha^{-1}) were used to evaluate the model prediction.

To reduce the error between observed and predicted variables, crop model parameters related to soil and root conditions were adjusted for all the years and farms in this study. We compared the observed yields against the predicted ITHRL50% scenario, which is similar to the conventional water management practices in Mississippi.

The model accurately predicted grain weight for both soybean varieties in Stoneville for 2010 and 2011 (Figure 5.6a and 5.6b). The errors in yield prediction were relatively small, indicating that the model performed well based on the conditions set for the model calibration process and the study site management practices. In contrast, the predicted yields varied more compared with the observed values for 2012 and 2013. The measured yields in 2012 were comparatively lower than in the previous years, which was mainly due to the initial dry conditions that delayed plant emergence and, consequently, reduced soybean yield (Burgess et al., 2012). In contrast, for 2013, the adequate initial soil moisture conditions at planting and germination accelerated the soybean emergence date, provided timely variations between reproductive growth stages, and generated a comparatively high yield with measured values around 5,790 kg ha^{-1} (Burgess et al., 2013).

For the Clarksdale site, the average yield prediction was more uniform for all four years. The model accurately predicted soybean yields for AG4831 for all years with differences between measured and predicted that ranged from 268 to 400 kg ha^{-1} from 2011 to 2013. Similarly, the model showed excellent yield predictions for AG4730 with an average error of 134 kg ha^{-1}. However, the model for Stoneville overpredicted the observed values for both AG4730 and AG4831 in 2012 by 1,191 kg ha^{-1}, and this is likely due to the variability in the initial soil moisture conditions for the planting and germination periods, which was not a controlled variable for this study. The crop model was calibrated based on the usual management practices applied in the region and provided a general estimation of yield for multiple years. The model was able to describe the impacts of different levels of irrigation on soybean yields grown under the general management practices in the MDR.

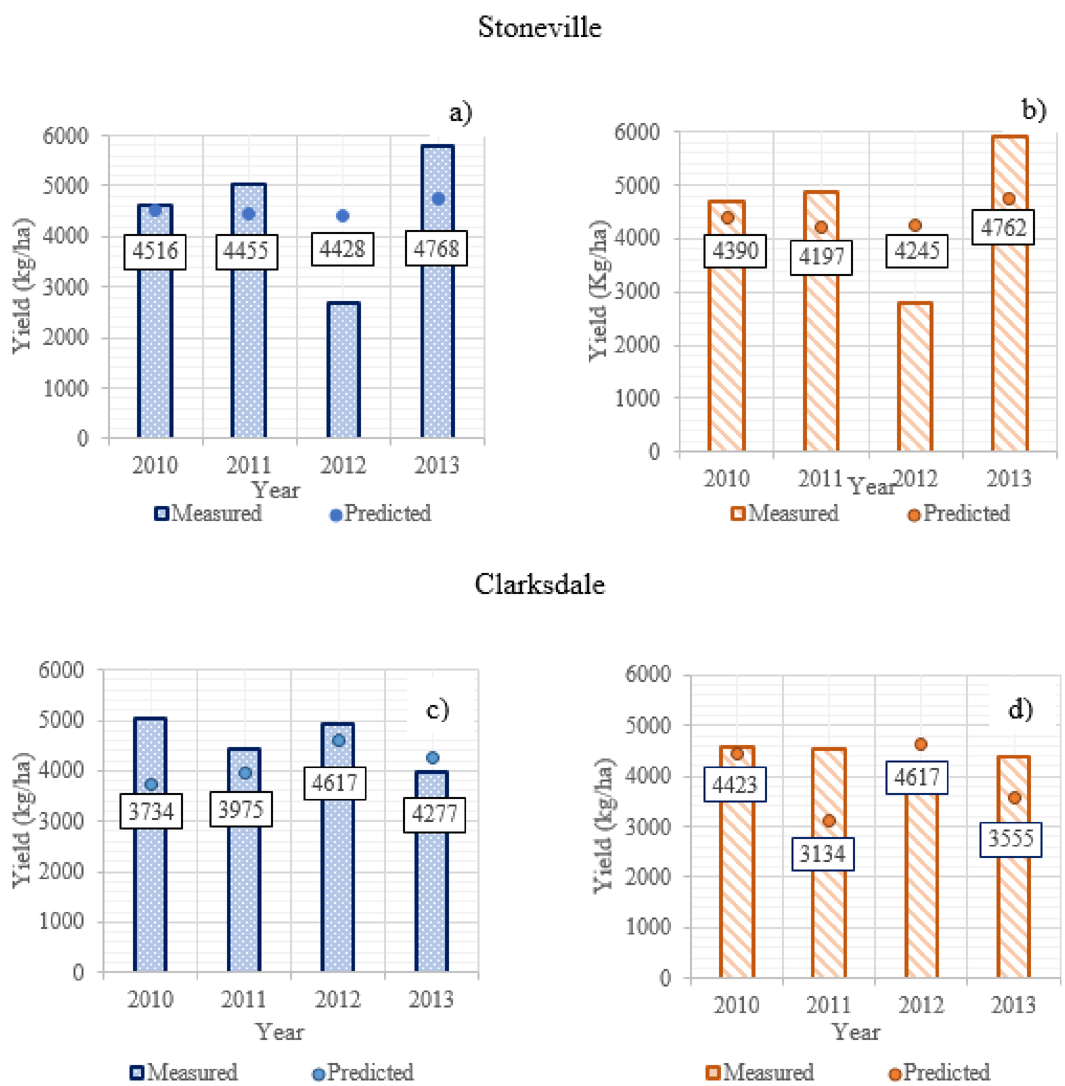

Figure 5.6 Comparison of measured vs predicted yield of two soybean varieties AG4831 and AG4730 planted in Stoneville and Clarksdale.

For Stoneville: a) AG4831, b) AG4730 and for Clarksdale: c) AG4831, d) AG4730

The CROPGRO-Soybean model accurately predicted the plant height of AG4831 grown in Stoneville in 2011 and 2013 (Figure 5.7a). The difference between measured and predicted heights was around 0.2 m for the three calibration years. Plant height data were not collected in 2010. Results for 2013 shows that the variations in plant height are correspondent to the changes in yield. For example, the variety AG4831 in Clarksdale overpredicts both height and yield crop variables. The crop model overpredicted the plant height of AG4730 with an average of 0.4 m for all the years and for both study sites. However, the model accurately predicted yield for variety AG4730 in 2010 at the Clarksdale site, despite ovepredicting the canopy height (Figure 5.7d). In addition, for the variety AG 4831 this correspondence between height and yield is more evident in comparison with the variety AG4730.

For the Clarksdale site, the predicted plant heights were fairly similar. The variety AG4831 showed an average difference of 0.2 m between measured and predicted plant height values, and the variety AG4730 overpredicted the soybean height for the four calibration years. Although the soybean height prediction showed slight errors, the model was able to predict the crop yields efficiently, especially for the variety AG4831, compared with similar soybean studies using CROPGRO (Jagtap and Jones, 2002; Paz et al., 1998).

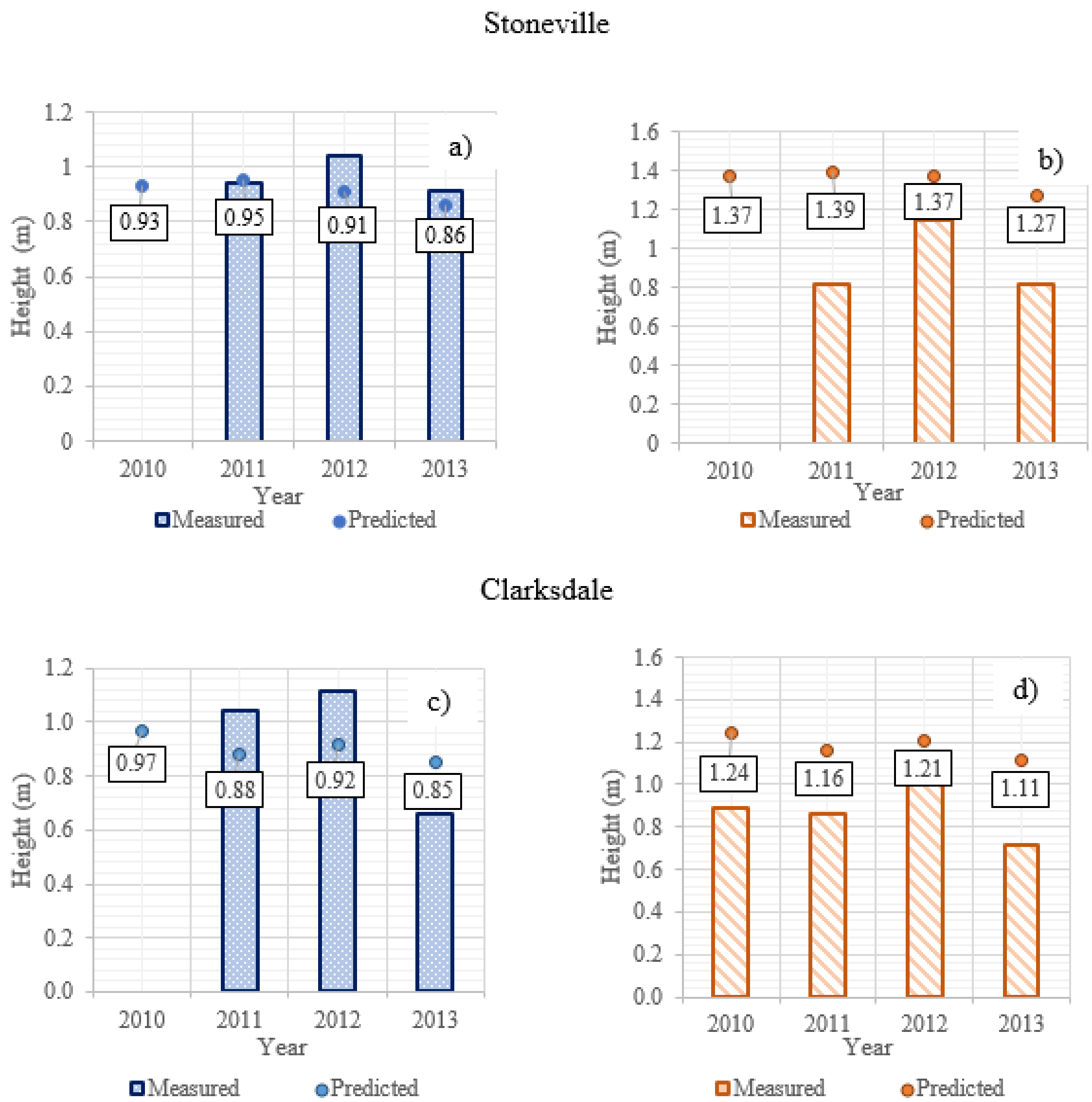

Figure 5.7 Comparison of predicted versus measured height (m) of two soybean varieties AG4831 (left) and AG4730 (right) planted in Stoneville and Clarksdale.

For Stoneville: a) AG4831, b) AG4730 and for Clarksdale: c) AG4831, d) AG4730

Irrigation Requirements

The calibrated CROPGRO-Soybean model was used to determine the yield responses and daily irrigation requirements of soybean under different ITHRL scenarios.

Table 5.4 shows the total irrigation requirements for soybean from 1988 to 1994 under the four irrigation threshold scenarios.

Table 5.4 IRRIC (mm) by threshold for the period of study.

YEAR	ITHRL20%	ITHRL40%	ITHRL50%	ITHRL60%*	PREC (mm)**
1984	79.0	94.0	136.9	160.0	1961
1985	161.0	163.1	191.0	206.0	1449
1986	205.0	255.0	276.1	284.0	1870
1987	119.9	160.0	166.1	202.9	2161
1988	80.0	93.0	109.0	135.9	1504
1989	80.0	96.0	112.0	136.9	3106
1990	150.1	209.0	209.0	214.9	1299
1991	74.9	119.1	128.0	133.1	3165
1992	37.1	59.9	101.5	150.1	1945
1993	119.9	158.0	192.0	184.9	1756
1994	80.0	126.0	165.1	162.1	2110

*The crop is irrigated when the soil available water is below the threshold
** Total precipitation per crop season

In general, selecting lower irrigation thresholds (20%, 40%) resulted in lower cumulative soybean water requirements (IRRIC) than that of ITHRL60%. The ITHRL50% scenario required lower IRRIC than the ITHRL60% scenario for all years except from 1992 to 1994. This means that it is possible to generate an irrigation management schedule that will lower groundwater pumping in the MRVA. Higher IRRIC based on ITHRL50% in 1992 to 1994 can be attributed to the variable irrigation frequency resulting from the number of precipitation events per crop season. As the model takes into account the amount of water that enters the soil system through precipitation and automatically triggers an irrigation event when the soil moisture goes below a specific threshold, the frequency of irrigation days is modified based on these precipitation amounts. For 1992, the ITHRL50% scenario initializes the first irrigation event 27 days after planting, with a total of seven irrigation days during the crop season. Meanwhile, for the ITHRL60% threshold, irrigation started at day 24 and had a total of eight irrigation days. This shows that although the volume of water applied is higher for the 50% threshold, the frequency of applications is lower than with the ITHRL60% scenario. The results highlight the importance of evaluating not only the amount of water required but also the number of irrigation events that are needed.

The water outputs provided by the crop model are the most reliable results for daily irrigation demands based on specific soil and weather conditions. IRRIC is inversely related with the volume of precipitation, so when the precipitation is lower, IRRIC is higher. One example is shown in 1991, where the amount of precipitation for the season fulfilled a considerable percentage of the crop water requirements which reduced the IRRIC volumes. Conversely, the year 1990 was drier and required higher amounts of irrigation. In addition, the highest IRRIC simulated by the model was for the year 1986 with values between 205 mm and 299.7 mm and a yearly precipitation of 1870 mm. The MDR experienced dry conditions and low precipitation in 1985, with an annual volume of 1449 mm. These conditions continued in 1986, which contributed to an unusually high irrigation demand for the nine predicted years. In general, the lowest IRRIC was for the year 1992 with volumes between 37.1 and 59.9 mm and an annual precipitation of 1945 mm. The low volumes can be attributed to the wet conditions for 1991 and normal precipitation conditions in the subsequent year of 1992.

While there were large differences in irrigation volumes between the ITHRL scenarios, the yield differences were fairly small. For example, the comparison of yield responses from ITHRL20% to ITHRL60% in 1984 shows that the application of 81 mm of additional water (from 79.0 mm of irrigation required for the 20% scenario to 160 mm for the 60% threshold as shown in Table 5.4) did not produce a large increase in soybean yield (Table 5.5). In fact, the yield increase for the entire crop season was only minimal (33.6 kg ha^{-1}). In most cases when the irrigation volumes were higher (ITHRL60% scenario), the increase in crop yield was negligible.

For all years, the highest yield was produced under ITHRL50% followed by ITHRL60% and ITHRL40%. For this study, it was assumed that the normal irrigation scenario in the Mississippi Delta has a threshold of 50%. The simulation showed that this scenario provided good crop yields based on the total amount of water applied. Nevertheless, when less irrigation is applied from the ITHRL50% to ITHRL40% scenario, the average yields will decrease around 134 kg ha^{-1} and the average water volume will be reduced by 30.48 mm.

These findings provide evidence that conservative irrigation water management by selecting a lower ITHRL (e.g. ITHRL40%) may result in good crop yields comparable to what is produced by a high water use management practice. Furthermore, if we apply this strategy in the MDR, lower IRRIC as a result of lower ITHRL can have a big impact on reducing the volume of groundwater withdrawal. Also, the adoption of lower ITHRL could help to increase crop production by reducing the costs associated with pumping additional water from groundwater wells. With the scenarios evaluated in this study, it is possible to analyze how yields are affected under future dry or wet climatic conditions and to determine the appropriate irrigation management strategies.

Table 5.5 Simulated yield (kg ha^{-1}) by threshold for the period of study.

YEAR	ITHRL20%	ITHRL40%	ITHRL50%	ITHRL60%
1984	4161.7	4189.9	4281.2	4195.3
1985	4026.2	4073.8	4163.1	4079.9
1986	3951.0	4073.2	4316.8	4077.9
1987	4036.9	4038.9	4202.7	4040.3
1988	3853.0	3863.8	4014.1	3889.9
1989	4120.8	4120.8	4299.3	4116.1
1990	4069.8	4055.7	4169.1	4057.7
1991	4016.8	4114.1	4320.8	4114.8
1992	4283.2	4283.9	4369.8	4279.9
1993	3949.0	3950.3	4130.2	3953.7
1994	4527.8	4528.2	4734.2	4538.9

Groundwater Level Prediction Based on the Crop Water Requirements

Results from the testing performance of each Gw+Pr+ITHRL scenario are presented in Figure 5.8. The figure shows how the adoption of different volumes of water for irrigation have an effect on the variability of the SVR daily groundwater prediction. The results suggest that an efficient irrigation management for soybeans will, in a short period of time, improve groundwater levels. Merrell (2009) also reported the short time response of the groundwater system to small changes in the crop water use methods.

The SVR prediction shows that all scenarios are able to simulate the groundwater withdrawal and recharge sinusoidal trends, and also are able to effectively forecast the groundwater levels for periods up to three months. The ITHRL60% scenario shows a delay in learning the general trend of the groundwater levels for the first withdrawal epoch between 1990 and 1991 (Figure 5.8a), but has an increasing predictive capacity after it reaches the first peak of recharge in 1991. The ITHRL20% scenario has an initial overprediction during the first withdrawal period until it reaches the recharge peak, compared to the ITHRL50% scenario, which has a fast learning rate and provides the best approximation for the first withdrawal. For all input scenarios, the model was able to capture the peaks and lows for the testing period, although the model generally underpredicted the measured daily groundwater levels.

The details of the SVR predictions for the second withdrawal period (Figure 5.8b) show the effect of each ITHRL scenario on the groundwater level trend and prediction. The predicted groundwater levels show a set of small peaks that reflect the periods when water was extracted for irrigation (area circled in green). The total number of small peaks during the predicted withdrawal period matches the average number of irrigation days per season, which reflects the model's sensitivity to groundwater withdrawals for irrigation. The ITHRL20% and ITHRL40% show higher peaks in comparison with the resulted peaks for the ITHRL50% and ITHRL60%. This can be caused by the higher pumping rates to reach the crop field capacity per irrigation application. Although there are differences between the observed and predicted groundwater levels, these differences are smaller than 1 m, which is acceptable for groundwater studies. When comparing the SVR performance of Gw+ Pr as inputs (presented in Chapter 4), with the performance of Gw + Pr + ITHRL scenarios in this study, it was evident that the addition of ITHRL scenarios decreased the SVR predictive performance. However, the evaluation of different irrigation thresholds contributes to the improved understanding of how the demand on groundwater for irrigation has an impact on the groundwater system and how fast this impact will be reflected in the aquifer.

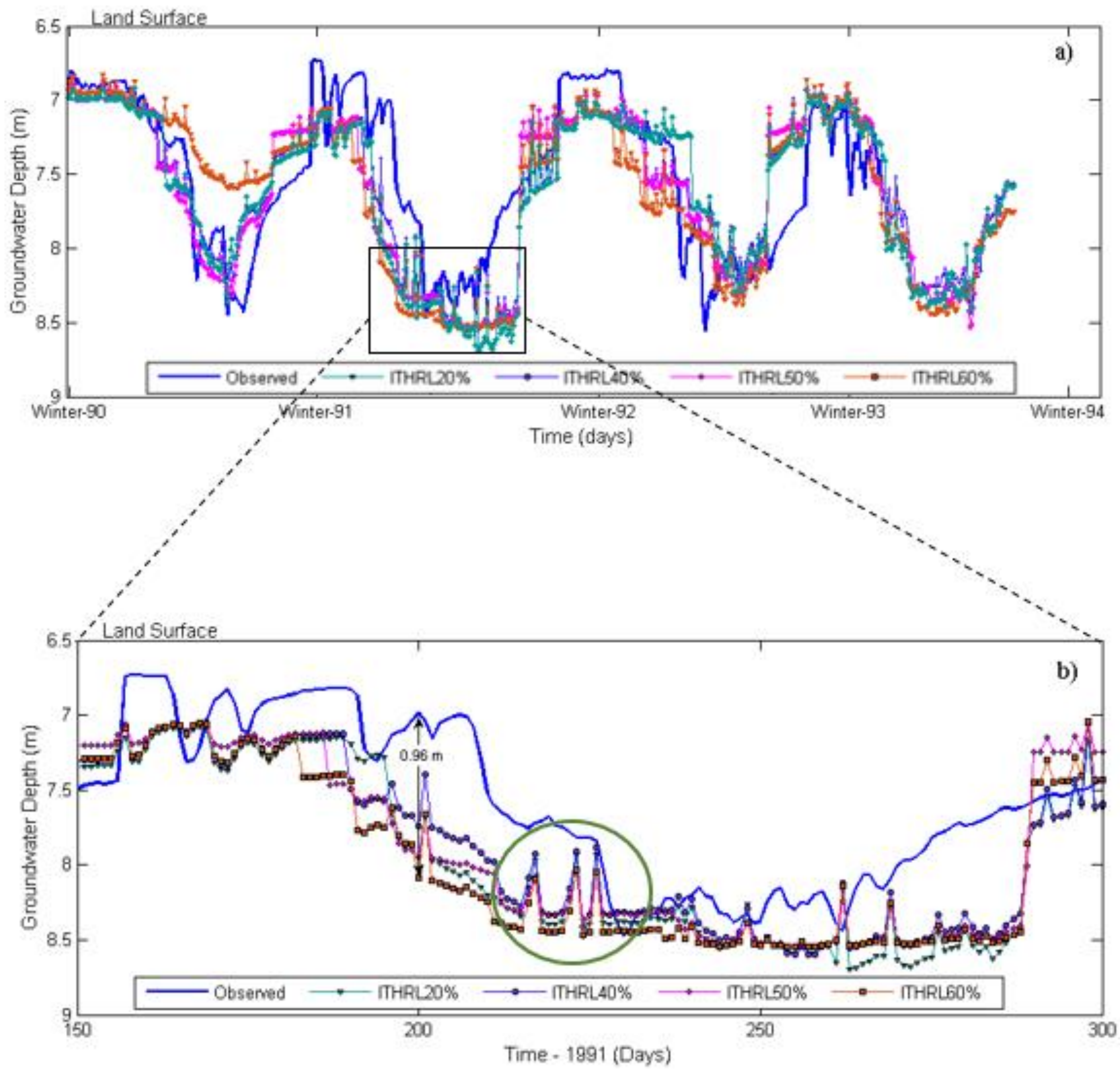

Figure 5.8 Comparison of the measured and predicted daily groundwater levels showing a) testing results for all ITHRL scenarios, b) details of the second withdrawal period.

The summary of observed and predicted groundwater level predictions for each soybean growing season are presented in Table 5.6. In general, predicted groundwater levels ranged from 6.7 m to 8.45 m below the soil surface, which is close to the measured levels which ranged from 7.27 m to 8.55 m. The results showed that the ITHRL60% had the highest prediction error. The SVR model is efficient in predicting the seasonal oscillations of groundwater levels for the observed well and is able to represent how the well responds to the proposed scenarios of full water availability (ITHRL60%) and water scarcity (ITHRL20%) evaluated in this study. For small scale studies, the estimation of fluctuations in daily groundwater levels is an advantage because it provides tools for the decision maker to evaluate how a possible drought scenario could be managed to maintain the crop's physiological development. In addition, for a seasonal estimation of water availability, the model shows the maximum and minimum simulated volume of water that the well can provide, which serves as a basis for developing a water management plan for the following crop season.

Table 5.6 Summary of groundwater levels for each soybean growing season.

Scenario	1991			1992			1993			1994		
	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
Observed4	8.45	6.80	7.46	8.45	6.74	7.60	8.45	6.73	7.56	8.45	6.73	7.16
20	8.36	6.92	7.43	8.36	6.92	8.01	8.36	6.92	7.44	8.36	6.92	7.71
40	8.21	6.90	7.42	8.21	6.90	7.95	8.21	6.90	7.47	8.21	6.90	7.67
50	8.32	6.89	7.44	8.32	6.89	7.99	8.32	6.89	7.50	8.32	6.89	7.72
60	7.59	6.83	7.23	7.59	6.83	8.05	7.59	6.83	7.59	7.59	6.83	7.76

The best SVR model prediction for groundwater levels was obtained with the Gw+Pr+ITHRL50% scenario with an MSE of 0.138 m (Table 5.7). All scenarios except ITHRL60% had coefficients of determination (R^2) greater than 0.56 (Figure 5.9). ITHRL60% had the lowest R^2, which is possibly due to the difficulty the model had predicting the first withdrawal epoch, and subsequently led to an increase in total model error. However, based on the MSE and general predictive trend, the Gw+Pr+ITHRL50% arrangement provided the best SVR performance. The results of this study will support a decision maker by providing information about the effect of different amounts of irrigation on the well response and the effects of water management practices that emphasize groundwater conservation and irrigation efficiency.

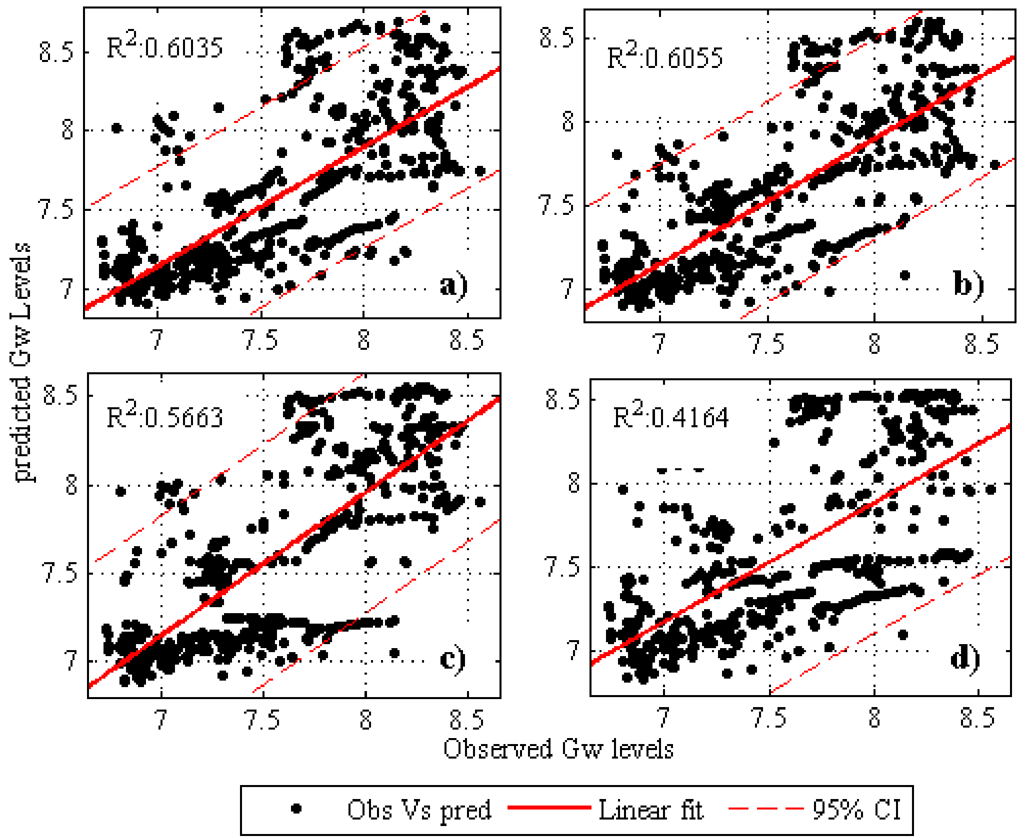

Figure 5.9 Comparison of observed versus predicted daily groundwater levels under different ITHRL scenarios

a) ITHRL20%, b) ITHRL40%, c) ITHRL50%, and d) ITHRL60%.

Table 5.7 MSE coefficient of performance for each irrigation threshold scenario.

Scenario	ITHRL20%	ITHRL40%	ITHRL 50%	ITHRL 60%
MSE	0.168067	0.159318	0.138529	0.152846

Summary and Conclusions

A modeling approach was implemented linking the CROPGRO-Soybean and the SVR model, to examine the potential impact of different irrigation volumes on the variability of daily groundwater levels in the MRVA. Variables such as soybean yield and canopy height were calibrated using the CROPGRO-Soybean simulation to obtain the daily IRRIC outputs, generated from four different ITHRL scenarios. These outputs were included as input variables along with Gw and Pr in the SVR model to predict the variability of daily groundwater levels.

The CROPGRO-Soybean simulation was able to predict the yield and irrigation requirements for soybeans in the MDR. The difference between measured and predicted yield was around 134 kg ha^{-1} for soybean variety AG4730 and 213 kg ha^{-1} for AG4831. In addition, the average height differences were 0.2 m for all years included in this study. An examination of the impacts of other field conditions including soil moisture content prior to planting and emergence may improve the crop model's predictive ability, and this could be a topic for future studies.

This study demonstrated that the implementation of conservative irrigation management by selecting a lower ITHRL can provide good crop yields compared to the yields from higher ITHRL. Thus, lower IRRIC would reduce the volume of groundwater withdrawal and the associated energy costs. The results provide a modeling tool for the decision maker to evaluate the most effective alternative for irrigation management and for groundwater withdrawal.

For the prediction of daily groundwater levels, the results showed the influence of different ITHRL in the SVR predictive response. The best scenario was provided by

ITHRL50% with an MSE of 0.138529 followed by the ITHRL60% and ITHRL40% scenarios. The results of this linked crop – SVR model could provide the baseline information for the evaluation of irrigation methods in the region and for the development of groundwater conservation plans. The modeling approach presented in this study can be applied to characterize and understand the interaction between different scenarios of irrigation management and groundwater variability. Understanding and quantifying the amount of water required from the aquifer to irrigate crops is essential to determine how different amounts of irrigation affect yields and the sustainability of groundwater sources.

References

Asefa, T., Kemblowski, M., McKee, M., Khalil, A., 2006. Multi-time scale stream flow predictions: The support vector machines approach. Journal of Hydrology, 318(1–4): 7-16. DOI:http://dx.doi.org/10.1016/j.jhydrol.2005.06.001

Batchelor, W.D., Basso, B., Paz, J.O., 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. European Journal of Agronomy, 18(1): 141-158.

Bray, M., Han, D., 2004. Identification of support vector machines for runoff modelling. Journal of Hydroinformatics, 6(4): 265-280.

Brouwer, C., Goffeau, A., Heibloem, M., 1985. Irrigation Water Management: Training Manual No. 1-Introduction to Irrigation. FAO, Rome, Italy.

Bruns, A.H., Meredith, W.R., Abbas, H.K., 2003. Effects of Furrow Irrigation on Corn in the Humid Sub-tropical Mississippi Delta. Crop Management, 2(1). DOI:10.1094/cm-2003-1222-02-rs

Burgess, B. et al., 2010. Mississippi soybean variety trials, 2010. Information Bull.

Burgess, B. et al., 2011. Mississippi soybean variety trials, 2011. Information Bull, 465.

Burgess, B. et al., 2012. Mississippi soybean variety trials, 2012. Information Bull.

Burgess, B. et al., 2013. Mississippi soybean variety trials, 2013. Information Bull.

Carruth, G.W., Paz, J.O., Tagert, M.L.M., Guzman, S.M., Oldham, J.L., 2014. Reusing irrigation water from tailwater recovery systems: Implications on field and stream-level nutrient status.Paper number 141913747, 2014 Montreal, Quebec Canada July 13–July 16, 2014. American Society of Agricultural and Biological Engineers (ASABE), pp. 1. DOI:10.13031/aim.20141913747

Chevalier, R., Hoogenboom, G., McClendon, R., Paz, J., 2011. Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks. Neural Comput & Applic, 20(1): 151-159. DOI:10.1007/s00521-010-0363-y

CoreTeam, R., 2012. R: A Language and Environment for Statistical Computing.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach Learn, 20(3): 273-297. DOI:10.1007/BF00994018

DREC-MSU, 2015. Delta Agricultural Weather Center. Delta Research and Extension Center–Mississippi State University.

Garcia y Garcia, A.G., Persson, T., Guerra, L., Hoogenboom, G., 2010. Response of soybean genotypes to different irrigation regimes in a humid region of the southeastern USA. Agricultural water management, 97(7): 981-987.

Graterol, Y.E., Eisenhauer, D.E., Elmore, R.W., 1993. Alternate-furrow irrigation for soybean production. Agricultural Water Management, 24(2): 133-145. DOI:http://dx.doi.org/10.1016/0378-3774(93)90004-T

Guzman, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A., 2015. Artificial Neural Networks and Support Vector Machines: Contrast Study for Groundwater Level Prediction Paper 152181983, 2015 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, New Orleans, Louisiana.

Heatherly, L.G., Wesley, R.A., Elmore, C.D., 1990. Corn, sorghum, and soybean response to irrigation in the Mississippi River alluvial plain. Crop science, 30(3): 665-672.

Holzworth, D.P. et al., 2015. Agricultural production systems modelling and software: current status and future prospects. Environmental Modelling & Software, 72: 276-286.

Hoogenboom, G. et al., 2015. Decision support system for agrotechnology transfer (DSSAT) Version 4.6 In: Foundation, D. (Ed.), Prosser, Washington.

Hook, J.E., 1994. Using crop models to plan water withdrawals for irrigation in drought years. Agricultural Systems, 45(3): 271-289.

Hook, J.E. et al., 2009. Assessing agricultural groundwater needs for the future: Identifying irrigated area and sources, Proceedings of the Georgia Water Resources Conference, Athens, GA.

Jagtap, S.S., Jones, J.W., 2002. Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. Agriculture, Ecosystems & Environment, 93(1–3): 73-85. DOI:http://dx.doi.org/10.1016/S0167-8809(01)00358-9

Jones, J.W. et al., 2003. The DSSAT cropping system model. European journal of agronomy, 18(3): 235-265.

Jones, J.W. et al., 1998. Decision support system for agrotechnology transfer: DSSAT v3. In: Tsuji, G., Hoogenboom, G., Thornton, P. (Eds.), Understanding Options for Agricultural Production. Systems Approaches for Sustainable Agricultural Development. Springer Netherlands, pp. 157-177. DOI:10.1007/978-94-017-3624-4_8

Karam, F., Masaad, R., Sfeir, T., Mounzer, O., Rouphael, Y., 2005. Evapotranspiration and seed yield of field grown soybean under deficit irrigation conditions. Agricultural Water Management, 75(3): 226-244. DOI:http://dx.doi.org/10.1016/j.agwat.2004.12.015

Kebede, H., Fisher, D.K., Sui, R., Reddy, K.N., 2014. Irrigation methods and scheduling in the delta region of Mississippi: current status and strategies to improve irrigation efficiency. American Journal of Plant Sciences, 5(20): 2917.

Liu, H. et al., 2011. Simulating water content, crop yield and nitrate-N loss under free and controlled tile drainage with subsurface irrigation using the DSSAT model. Agricultural Water Management, 98(6): 1105-1111.

MacRobert, J., Savage, M., 1998. The use of a crop simulation model for planning wheat irrigation in Zimbabwe, Understanding options for agricultural production. Springer, pp. 205-220.

Maupin, M.A., Barber, N.L., 2005. Estimated withdrawals from principal aquifers in the United States, 2000 [electronic resource] / by Molly A. Maupin and Nancy L. Barber. Circular: 1279. Reston, Va. : U.S. Dept. of the Interior, U.S. Geological Survey ; Denver, CO.

Menne, M. et al., 2012. Global Historical Climatology Network-Daily (GHCN-Daily), Version 3. 2. NOAA National Climatic Data Center.

Mercer, A.E., Richman, M.B., Bluestein, H.B., Brown, J.M., 2008. Statistical modeling of downslope windstorms in Boulder, Colorado. Weather and Forecasting, 23(6): 1176-1194.

Merrell, T.L., 2009. Development of an interactive model predicting climatological and cultural influences on annual groundwater volume in the Mississippi Delta shallow alluvial aquifer. Master Thesis Thesis, Mississippi State University, 87 pp.

NASS, U., 2003. United States Department of Agriculture-National Agricultural Statistics Service. Cropland Data Layer.

Navarro-Hellín, H., Martínez-del-Rincon, J., Domingo-Miguel, R., Soto-Valles, F., Torres-Sánchez, R., 2016. A decision support system for managing irrigation in agriculture. Computers and Electronics in Agriculture, 124: 121-131. DOI:http://dx.doi.org/10.1016/j.compag.2016.04.003

Ojha, R., Ramadas, M., Govindaraju, R., 2013. Current and Future Challenges in Groundwater. I: Modeling and Management of Resources. Journal of Hydrologic Engineering, 20(1): A4014007. DOI:10.1061/(ASCE)HE.1943-5584.0000928

Paz, J. et al., 1998. Analysis of Water Stress Effects Causing Spatial Yield Variability in Soybeans. Transactions of the ASAE, 41(5): 1527.

Raghavendra, N.j., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Applied Soft Computing, 19(0): 372-386. DOI:http://dx.doi.org/10.1016/j.asoc.2014.02.002

Salazar, M. et al., 2012. Estimating irrigation water use for maize in the Southeastern USA: a modeling approach. Agricultural Water Management, 107: 104-111.

Sassenrath, G.F. et al., 2013. Development of the Mississippi Irrigation Scheduling Tool-MIST, 2013 Kansas City, Missouri, July 21-July 24, 2013. American Society of Agricultural and Biological Engineers, pp. 1.

Thorp, K.R., DeJonge, K.C., Kaleita, A.L., Batchelor, W.D., Paz, J.O., 2008. Methodology for the use of DSSAT models for precision agriculture decision support. Computers and Electronics in Agriculture, 64(2): 276-285.

USDA-NRCS, U.D.o.A.N.R.C.S., 2015. Web Soil Survey.

USGS, U.S.G.S., 2014. National Water Information System: Web Interface.

Woli, P., Paz, J.O., 2012. Evaluation of Various Methods for Estimating Global Solar Radiation in the Southeastern United States. Journal of Applied Meteorology and Climatology, 51(5): 972-985. DOI:10.1175/JAMC-D-11-0141.1

Yang, J., Yang, J., Liu, S., Hoogenboom, G., 2014. An evaluation of the statistical methods for testing the performance of crop models with observed data. Agricultural Systems, 127: 81-89.

Yang, Y., Watanabe, M., Zhang, X., Hao, X., Zhang, J., 2006. Estimation of groundwater use by crop production simulated by DSSAT-wheat and DSSAT-maize models in the piedmont region of the North China Plain. Hydrological processes, 20(13): 2787-2802.

CHAPTER VI

GENERAL CONCLUSIONS

The following are the major conclusions from this study:

1. The nonlinear autoregressive with exogenous inputs (NARX) recurrent neural networks were implemented to forecast daily groundwater levels at various lead times. The Bayesian Regularization (BR) and the Levenberg-Marquardt (LM) training algorithms were evaluated to determine the most efficient training architecture for the forecasting of daily groundwater levels of a USGS monitoring well located in the Mississippi Delta Region (MDR). The BR with 100 time delays and 2 hidden layers was the most accurate architecture based on the measurements of statistical performance and the training and testing results. Although the LM training algorithm was 12% faster in reaching model convergence, the BR algorithm showed higher predictive performance, and it was more robust in the prediction of daily groundwater levels. The NARX-BR model was able to forecast groundwater levels up to three months ahead with a Mean Squared Error (MSE) of less than 0.00119 m.

2. The NARX neural network and the SVR models were compared to identify a machine learning technique that can efficiently describe the complexities of groundwater level variability at a seasonal scale. The models were trained and tested using nine years of summer and winter daily groundwater levels from 1985 to 1994. Both models were able to predict the variations in withdrawal for the summer season. However, the SVR performed better than NARX in predicting changes in groundwater levels during the winter season. An SVR with a radial basis function (RBF) kernel function, $\gamma = 0.01$, $C=100$, and $\varepsilon = 0.1$ provided the best architecture for both the withdrawal and recharge periods. Although the process of finding the predictive function parameters to generate an adequate SVR training architecture was time consuming, SVR was the most effective machine learning technique to predict the seasonal groundwater trends. More importantly, this study demonstrated that seasonally divided input time series helped the SVR model reduce its computational requirements and generate faster iterations.

3. Seven arrangements of input variables that included observed groundwater levels (Gw), precipitation (Pr), and evapotranspiration (ET) were evaluated to determine the best combination for an SVR model to predict daily groundwater levels. This study showed that the Gw+Pr arrangement provided the best input option for the groundwater level prediction using SVR and produced the lowest MSE (0.00123 m). If Pr is not available as an input for the model, a solution can still be generated using only Gw as input, and acceptable predictions can be obtained with an MSE of 0.21058 m. The SVR had a reduced performance when ET was included in the input arrangement (Gr+Pr+ET), and the degraded performance was particularly evident during the recharge periods. The results from this study can be used as a baseline to construct improved data driven models to forecast groundwater levels in individual wells around the MDR. However, it is important to mention that SVR should be calibrated individually for further studies in wells around the region, as the model is based on data and does not account for regional aquifer dynamics.

4. A modeling approach was implemented that linked a calibrated CROPGRO-Soybean model and an SVR model to assess the impacts of varying levels of irrigation thresholds (ITHRL) on daily cumulated irrigation requirements (IRRIC) and the corresponding changes in daily groundwater levels in the MDR. The IRRIC outputs of the crop model, generated from four ITHRL scenarios, were included as input variables along with Gw and Pr in the SVR model to predict daily groundwater levels.

The SVR model was efficient in predicting the seasonal oscillations of the groundwater levels and was able to represent how the groundwater system responded to the proposed scenarios of full water availability (ITHRL60%) and water scarcity (ITHRL20%). These findings provide evidence that a conservative irrigation water management by selecting a lower ITHRL (e.g. ITHRL40%) may still result in good crop yields comparable to what is produced by a high water use management practice. Furthermore, if we apply this strategy in the MDR, lower IRRIC as a result of lower ITHRL can have a big impact on reducing the volume of groundwater withdrawal.