2011

Assimilating in situ soil moisture measurements into the DSSAT-CSM using a Kalman filter

Derek Gene Groenendyk
Iowa State University

Assimilating in situ soil moisture measurements into the DSSAT-CSM using a Kalman filter

by

Derek Gene Groenendyk

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Agricultural Engineering

Program of Study Committee:

Amy Kaleita, Major Professor

Brian Hornbuckle

Rob Malone

Iowa State University

Ames, Iowa

2011

Dedication

To all the support from my loving friends and family, as well as my beautiful wife who were all able to support me during this time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to recognize the hard work of Candace M. Batts whose study and research was invaluable to completion of this project.

Abstract

With the ability to monitor soil moisture in time comes the opportunity to develop ways to incorporate these measurements into predictive models, without compromising or overriding the model physics. The importance of soil moisture to the growth of crops is well understood and because of this it is recognized as one of the more important parts of crop modeling programs. This research focused on improvements to the Decision Support System for Agrotechnology Transfer - Cropping System Model (DSSAT-CSM) as determined by the accuracy of soil moisture estimates. To accomplish this, data assimilation techniques were implemented to process the uncertainty of the model estimates and *in situ* measurements of soil moisture. Consideration of soil parameter uncertainty, which influences model estimates of soil moisture and model output, was taken into account using a Monte Carlo approach. A Kalman filter was used to combine the model estimates of soil moisture with *in situ* soil moisture measurements, while varying several important soil parameters in the model using a Monte Carlo approach. Covariances for the Kalman filter were calculated for the model and measurements based on the model's standard deviation from the Monte Carlo soil moisture estimates and the standard deviation of the *in situ* soil moisture measurements. Data for this study was obtained from a research study conducted on irrigated wheat during the winters of 2003-04 and 2004-05 in Maricopa, Arizona, in which thorough field and crop data were collected. Results of the simulations were compared against biomass and yield measurements to determine the effectiveness of the data assimilation scheme. The Monte Carlo approach with assimilation done in the top layer of the soil profile was only able to moderately address uncertainty present in the soil parameters. Improvement resulted for data assimilation of soil moisture through the reduction of the error between the measured and simulated grain yield and canopy weight for 47% and 37% of the simulations for the 2003-2004 and for 25% and 32% of the simulations for the 2004-2005 season, respectively. Assimilation was more effective for improving the model output of

grain yield for the 2004-2005 than the 2003-2004 season and canopy weight for the 2003-2004 season than the 2004-2205 season. The results of model estimated daily NO_3 levels in the soil layers from data assimilation simulations indicates that assimilation of soil moisture can influence its levels. The data assimilation combined with a Monte Carlo approach showed the use of remotely sensed soil moisture could lead to improvements of frequently studied model outputs, such as grain yield and canopy weight. Further study is needed to fully understand the most desirable conditions for soil moisture assimilation and what other influencing effects data assimilation of soil moisture presents.

CHAPTER 1. Overview

1.1 Introduction

In agriculture and hydrology, computer simulation models are used to carry out experiments and research for environments, conditions and situations that are hard or costly to replicate, reach or find. The models allow researchers and scientists to complete varying experiments without having to physically recreate the experiments or observations which are often costly and time-consuming. Models are used in this manner to simulate different weather, soil, land use, and vegetation conditions.

Food security and production sustainability benefit from the production data and crop yield information provided by crop models. The models benefit scientists, policy-makers, and the general public for use in precision agriculture, crop development, economic forecasting, famine predictions, and global crop forecasting as well[Chen et al. (2008)].

Crop growth models have been used for many years to predict yield, monitor crop growth, and manage farms. The Decision Support System for Agrotechnology Transfer - Cropping System Model (DSSAT-CSM) has been used by scientists and researchers for years to predict crop yields for a family of 25 crops. DSSAT-CSM implements its crop models into a common simulation platform that share key non-plant modules such as soil and weather. Moreover, the DSSAT-CSM is still widely used and have are continually being updated to meet ever present modeling demands, making use of improved cropping models and non-plant modules [Jones et al. (2003)].

In recent years agricultural production has adopted new standards and procedures to meet the worlds demands for food supply while focusing on maintaining the efficiency of production. Agricultural production has been able to take advantage of the rapidly growing technologically

advanced, affordable scientific equipment, such GPS, satellites, and sensors, designed for applications in information management and collection. This implementation has grown into what is considered today to be precision agriculture. Since the evolution of precision agriculture models, more specifically crop models, are now being used more regularly and at higher spatial resolutions [Batchelor (2002)]. The models require inputs of weather, soil property, management practices, and cultivar characteristic information in order to simulate biological processes like crop growth. Now that data and measurements for these parameters or variables are being used more frequently, at finer resolutions and on larger scales, they offer the opportunity to improve model accuracy. However, relatively few studies have been done with the focus being on using measured environmental conditions along with agricultural crop models, particularly in the case of soil moisture with DSSAT-CSM.

Soil water content is considered a significant factor in crop models, because the importance of and influence of soil moisture on crop growth is well known. Soil water content is a key factor specifically for the DSSAT-CSM model because its estimated value is used frequently in the soil-plant-atmosphere, plant, and management modules within the DSSAT-CSM model [Jones et al. (2003)]. These modules in the crop model take advantage of the relationship between soil water content and transpiration rates. The amount of water content most importantly directly relates to water availability for crop growth [Bert et al. (2007)]. Amongst other factors, such as weather, genetics, and plant population, water stress is known to limit yield [Batchelor (2002); Novak et al. (2005)]. Thus soil water content is important for accurately simulating crop growth.

In recent years there have been efforts at improving the water balance module and methods in DSSAT-CSM. These studies recognize and are based on the idea that the DSSAT-CSM water balance methods could benefit from improved ability to estimate soil moisture. Models such as Root Zone Water Quality Management, RZWQM, have been coupled with the crop modules of the DSSAT-CSM to help improve the water profile estimates [Ma et al. (2005, 2006)]. There has also been an increase in the use of *in situ* measurements, observations and remotely sensed data combined with crop models to improve their predictions [Chen et al. (2008)].

Maas (1988) presented the usefulness of several techniques specifically focused on using

remotely-sensed data in improving models, these were again revisited by Moulin et al. (1998).

(a) direct use of a driving variable in the model

(b) updating of a state variable in the model

(c) re-initialization of the model

(d) re-calibration of the model

The four procedures that they covered form the basis of what is called data assimilation.

Data assimilation based on "updating" consists of taking a measurement and inserting into the model as a replacement for or in combination with the estimated state variable in manner to achieve an updated model estimated state variable. A state variable represents the value or state of a condition within the model that changes in time. The goal of incorporating the measurement is to obtain an optimal or true state variable in the model. The state variable is considered optimal or most representative of the value in reality as far as modeling is concerned is when the state variable has the lowest possible uncertainty but sometimes it is considered as having minimum variance.

Because both model estimates and *in situ* measurements contain errors, using a combination of the two while considering their corresponding errors should yield a superior estimate. To combine the model estimate and the measurement, an assimilation algorithm, such as a Kalman filter, is required. The Kalman filter is an algorithm that weights the uncertainty of the measurement against the uncertainty of the model estimate to obtain a gain value, the Kalman gain. The Kalman gain is applied to the residual or difference between the measurement and model estimate which is then added to the previous model estimate to achieve the optimal state variable or new model estimate [Reichle (2008)].

Many successful studies have been conducted in which a Kalman filter has been implemented to assimilate observational data into models. The Kalman filter has the ability to efficiently and effectively take *in situ* measurements and combine them with model estimates to improve DSSAT-CSM predictions, specifically considering soil moisture values. Data assimilation of soil

moisture *in situ* measurements should improve, specifically, the yield prediction of DSSAT-CSM version 4.

This study will be focusing on the "updating" data assimilation scheme that Maas (1988) covered, based on *in situ* measurements. This research is driven by the hypothesis that assimilation of soil moisture observations using a Kalman filter will improve the model output, specifically grain yield and canopy weight, of the DSSAT-CSM.

CHAPTER 2. Review of Literature

2.1 Soil Moisture

With the recent development of precision farming the research community can foreseeably provide the agricultural community with useful data but also important insights on how the data is useful within their area. As precision farming continues to grow the agricultural community would greatly benefit from new technologies such as soil moisture networking systems. The importance of soil moisture and its influence on crop growth and yield is well known. Amongst other factors, such as weather, genetics, and plant population, water stress is known to limit yield [Batchelor (2002)].

Soil moisture is important in agriculture. It is of specific interest for crop irrigation scheduling because of water's limited availability and managing it effectively often requires reliable information about evapotranspiration or ET [Hunsaker et al. (2007a)]. It has also been a consideration in water management, meteorology and land surface climate models [Paniconi et al. (2003)]. The study, understanding, and management of surface geophysical processes is dependent on the temporal and spatial variation soil moisture, since it is the bridge connecting the hydrologic cycle with the energy budget of land surfaces [Houser et al. (1998)]. Therefore, when considering agricultural production and plant science, water, specifically soil moisture, is considered one of the most important factors.

Water is a crucial medium for nutrient transport and exchange, cooling and other processes necessary for plant growth. Furthermore, the drainage of water through the soil profile influences the level of nutrient availability. The necessary soil water content for processes such as nutrient and oxygen movement can only be properly simulated if the components of the water budget such as infiltration, runoff, drainage, evaporation, and root water uptake rates

are reasonably simulated.

This is particularly relevant when modeling crop growth and requires that we are able to reasonably simulate soil moisture [Tsuji et al. (1998a)], specifically in the DSSAT-CSM because the estimated value of soil water content is used in the Soil-Plant-Atmosphere, Plant, and Management modules [Jones et al. (2003)]. Transpiration rate, which is limited by soil water content, is one of the many processes that can affect yield [Novak et al. (2005)]. The water content can indirectly affect yield predictions but most importantly its value directly relates to water availability for crop growth [Bert et al. (2007)].

2.2 DSSAT-CSM

The Decision Support System for Agrotechnology Transfer - Cropping Systems Model (DSSAT-CSM) has been created and utilized to model crop growth for many years. It is a physical model, versus a statistical model, that computationally simulates the biological processes that occur during crop growth. In 2010, the International Consortium for Agricultural Systmes Applications (ICASA) released a new version of the DSSAT called the DSSAT-CSM 4.5. DSSAT-CSM version 4 and beyond have had the underlying code restructured into modular components. The physical layout and organization of the program was modularized so that its components were more accessible and easily maintained. The new organization allows each plant or cultivar module to use the same soil, weather, and management modules [Jones et al. (2003)].

The DSSAT-CSM has been used widely in the academic and research community [Jones et al. (2003)] and provided valuable insight into understanding the processes of diverse agricultural cropping systems. The DSSAT-CSM has been evaluated for various cropping systems creating a good foundation for the testing of new routines and changes to current routines in the model.

2.2.1 DSSAT-CSM Structure and Operation

The DSSAT-CSM's main purpose is to manage information for and control the simulation of 25 cultivar models. To do so it takes into consideration weather, soil, and plant dynamics as

well as irrigation and management practices, and plant stresses.

The Main Program of the DSSAT-CSM, consisting of six operational steps, is the controlling structure for the time loops and modules. The Land Unit Module handles the communication and control between the Main Program and six primary modules; Weather, Management, Soil, Soil-Plant-Atmosphere, CROPGRO Crop Template, and Plant. Each of these primary modules contains sub modules relevant to fulfilling the modules purpose.

The Land Unit Module is responsible for transferring data to the Plant module, and sub-modules of the Plant module such as CERES-Wheat, needed to simulate crop growth. CERES-Wheat has been extensively evaluated and validated for use in many different locations, having unique soil and climate conditions, and varieties. Its successful performance has been well-documented making it a reliable and trusted crop model [Jones et al. (2003)].

2.2.1.1 Weather Module

The DSSAT-CSM will accept user provided weather data or it can simulate site relevant weather data based on historical data using SIMMETEO or WGEN. The required daily inputs are minimum and maximum temperature, dew point temperature, solar radiation, wind, and precipitation.

2.2.1.2 Soil-Plant-Atmosphere Module

Within the Soil-Plant-Atmosphere module the daily plant transpiration and soil evaporation is calculated. Potential evapotranspiration (ET) is scaled up from a Penman-Monteith reference ET using a crop coefficient based on leaf area index. Actual ET is calculated from soil moisture conditions. To calculate the actual evapotranspiration and the potential ET this module requires the daily weather inputs as well as soil properties, soil water content and leaf area index. Soil water content and leaf area index are computed daily but the initial soil water contents and soil properties are provided to the DSSAT-CSM in a specific file. To determine the ET values, root uptake through each soil layer is computed, which also requires the calculation of root length density for each soil layer.

2.2.1.3 Management

Planting, Harvest, Irrigation, Fertilization, and Residue are all the sub-modules included within the Management module. These sub-modules are used to specify and control, through user inputs, when their corresponding management practices take place and at what level. All of these practices can be scheduled in the DSSAT-CSM experiment file or automated by the DSSAT-CSM. Automation is typically based on the number of days from planting or when certain conditions are reached. Automatic planting occurs when the soil water content in the top 30cm and the soil temperature have reached preset limits. Likewise, automatic harvesting can occur when the crop has reached maturity or the soil water content permits machine field operations. If irrigation is desired and a user-specified schedule is not provided, it is triggered when the available plant water drops below a fraction of water holding capacity for a given management depth. Inorganic fertilizer application takes place on the provided days or when, for automatic management, the plants require it, based on the plant nitrogen stress level. Applied organic fertilizer and crop residue are can be accounted for at the beginning of the simulation, after harvest or based on user specified days.

2.2.1.4 CERES-Wheat Sub Module

Each of the Plant sub-modules or crop models represents an different individual crop, each having different phenological parameters. Each model is capable of simulating respective growth stages, plant nitrogen and carbon demands. One specific crop model available is CERES-Wheat. The CERES-Wheat model has been specifically designed to exhibit the behaviors, specifically growth and yield, of individual wheat species.

CERES-Wheat simulates seven stages of the wheat plant life: germination, emergence, terminal spikelet, end ear growth, beginning grain fill, maturity and harvest. Growing degree days, calculated using the maximum and minimum daily temperatures, determine the the rate of development. Progress from one stage of growth to another can be based on specific user defined inputs or it be computed within the module using other user inputs.

Consideration of dry matter or dry biomass accumulation is part of the model's physics as

well. Daily intercepted light, based on leaf area index (LAI), plant population and spacing, is converted to dry matter using a radiation use efficiency parameter. Dry matter totals for each day are also influenced by soil water, soil nitrogen, air temperature, atmospheric CO_2 concentration, depending on the most limiting [Jones et al. (2003)].

2.2.1.5 Soil Water sub module

The Soil module includes a sub-module that utilizes water balance dynamics for calculation of soil moisture. Other sub-modules that are included use the dynamics for nitrogen, carbon and temperature to determine their respective state.

The soil water balance model, originally developed for CERES-Wheat, was created as a one-dimensional model, using irrigation, infiltration, vertical drainage, unsaturated flow, soil evaporation and plant root uptake processes to compute the daily water content changes experienced by each layer in the soil profile.

Soil water movement downward through the soil profile is modeled using a "tipping bucket" method, when the water content is above the drained upper limit, or level at which no further drainage can take places. The soil parameter provided for diffusivity and differences between the adjacent layers' soil water content are used to calculate the upward saturated flow.

Infiltration is calculated as the difference between the precipitation and runoff calculated using the Soil Conservation Service - Curve Number method [Soil Conservation Service (SCS) (1972)]. The DSSAT-CSM however includes a modification to the SCS-CN method, by Williams et al. (1984), that compensates for soil layers and also for initial soil water content at the time of precipitation. Irrigation is assumed an additive component of total precipitation. Water accumulates above a soil layer only if the drainage, downward soil water movement, for the layer is greater then saturated hydraulic conductivity considered over a 24 hr period, which results in the actual drainage being equal to the saturated hydraulic conductivity over a 24 hr perio. Otherwise the actual drainage through the each layer is assumed to be the calculated vertical drainage for the layer. Drainage through each layer is considered only after a total drainage for the soil profile has been calculated, which is determined by a global soil drainage parameter. Also, as mentioned, the Soil-Plant-Atmosphere module calculates the soil evaporation and plant

root uptake. These fluxes are all calculated as equivalent depths and added or subtracted to the soil water content for each layer on each day. The DSSAT-CSM uses a water balance method that at times may oversimplify what is actually happening in the local soil profile. The DSSAT-CSM version 4 uses the same SCS CN, method from version 3.5 and earlier.

The curve number method should not be assumed to provide accurate runoff and infiltration values for specific storms [Tsuji et al. (1998a)]. Sadler et al. (2000) also found that the model overestimates infiltration when using the SCS curve number method. So notably, this will affect the accuracy of the DSSAT-CSM's soil moisture estimation. However, it has been shown that the SCS-CN within the DSSAT-CSM, can in fact perform well [Liu et al. (2011)].

2.3 Data Assimilation

Soil moisture is influenced by vegetation, ground cover and soil conditions. This dependency is one that leads to spacial and temporal variability. To address variability or uncertainty of soil moisture studies focus on improving the soil moisture estimates of the models.

In recent years there have efforts at improving the water balance module and methods in DSSAT. Models such as Root Zone Water Quality Management, RZWQM, have been coupling with the crop growth modules of the DSSAT-CSM to help improve the water profile data [Ma et al. (2005, 2006)]. There has also been an increase in the use of *in situ* measurements, observations and remotely sensed data combined with crop models to improve their predictions [Chen et al. (2008)].

The DSSAT-CSM has been created independently of the research that has been done developed systems or methods with the ability to utilize *in situ* measurements to benefit knowledge of crop production and crop growth. They have the ability to complement each other well and could improve the understanding of crop growth and forecasting if used in combination with each other. Each has their own but different advantages but each has drawbacks. With the marriage of the two we can hope to eliminate or at the very least weaken the drawbacks effect on the prediction of crop production.

We acknowledge that models are by nature imperfect however there may be ways to improve the models without completely rewriting or making them more complex. One way to do this is

through the consideration of the observed or measured data. We also know that measurements can be imperfect and have error but we also know that they are constrained and most likely have a lower error value. We can take advantage of this lower error value and the constraints that the observed data has by applying it to the model's state variable estimates. The use of state variables here could also be changed to focus on the optimizations of state parameters, mainly through iterative model simulations.

Moulin et al. (1998) and Maas (1988) discussed the four main procedures for utilizing data to improve models. These four procedures make up an area that is called data assimilation.

(a) direct use of a driving variable in the model

(b) updating of a state variable in the model

(c) re-initialization of the model

(d) re-calibration of the model

The meteorology field has used data assimilation for quite some time. The hydrological community is starting to use them more regularly and also find success with their implementation [Paniconi et al. (2003); Alavi et al. (2009)].

As we consider data assimilation options or configurations, one of the biggest decisions is choosing an appropriate system or algorithm for melding observed data and the model state variables. The decision becomes important because there are several factors or assumptions associated with the algorithm that can influence results or the efficiency of the overall system.

Before an algorithm or procedure is chosen the appropriateness or the method requires consideration. Many aspects of advantages or disadvantages can play a role in this consideration. Methods for data assimilation are designed for use under certain conditions or assumptions. If these assumptions are not met or considered appropriately when implementing an algorithm, it may lead to inaccurate results.

To choose an appropriate method, a good understanding of what type of data is used within the algorithm and how the model will interact with the algorithm are aspects that are taken into account. The scheme may not need a complex assimilation method to efficiently or effectively

meld the data. An overly complex algorithm or method may be able to produce good estimates but a simpler algorithm or method may be computationally faster and produce results that are sufficiently comparable.

The method or algorithm chosen should not be too computationally costly, meaning that it should not cause the scheme or model to run so slowly the results gained by its computations are not worth the time lost to process them. This is often important because the number of observations and associated properties used can be large due to spatial size or the number of dimensions and in turn cause the need for many computations. Choosing a model can be a simple process if the right representation for the analysis the modeling errors is chosen [Bouttier and Courtier (2002)].

2.3.1 Direct Insertion

Direction insertion is a method that simply replaces the models estimated values for state variables with the values of the measurements. So rather then allowing the model to compute estimates of the state variables the measurements are used instead. This method is very quick computationally however all influence or knowledge coming from the model is lost at the timesteps where the state variables are modified. When measurements are required on a very frequent interval an interpolation method is sometimes used for the timesteps between actual measurements because measurements on this time scale can be difficult to obtain. It does however allow insight into how the model is behaving and how it is likely to behave when important model variables are changed or modified. Walker and Houser (2001) and Walker et al. (2001) were able to produce promising results using direct insertion methods from *in situ* soil moisture measurements.

The direct insertion scheme results in a very simple and unique forcing algorithm. However the errors that are present in the measurements are often propagated all the way through the assimilation scheme. This may weigh heavily on the estimates since all the assimilation weight is placed onto the measurements. Alavi et al. (2009) reports on several experiments that have used direct insertion for many years specifically within meteorology and oceanography and that it has also been used more recently in hydrologic fields.

2.3.2 Re-calibration and Re-initialization

The re-calibration procedure looks at modifying the model parameters for a specific site. These parameters are first tested for model sensitivity to determine which parameters that will have a greater impact on the model. Sets of simulations are created by varying the values for each influential parameter. The model is considered to be calibrated when the sensitive parameters have values that align the model state variable estimates closest to the values of the measurements. The optimized parameters are typically site specific, therefore, calibration must be done for every site because each site characteristic can often vary from the one used during calibration. Consequently, this methodology can require significant amounts of data and computation time.

Re-initialization is conducted in much the same manner as re-calibration. Instead of adjusting the model's state parameters the models' initial values for the state variables are changed until again the state estimates have values that are at a minimum difference from the measurements.

2.3.3 Updating

Data assimilation "updating" schemes consist of taking a observation or measurement and combining it with a model estimated state variable in manner to achieve an updated estimate of the state variable. The goal of incorporating the observation is to obtain an optimal or "true" state variable model estimate. To combine the model estimate and the measurement, an assimilation algorithm or system, such as a Kalman filter, is required.

Updating techniques differ from direct insertion because they do not always completely replace the model estimates but use measurements to correct or update the state variable estimate only when measurements exist. Updating schemes also employ methods the take into consideration the uncertainties that exist with the model state variables estimates and the measurements to weight their values and produce a more optimal state variable estimate [Reichle (2008)]. Updating schemes have shown to be a better alternative to direct insertion [Heathman et al. (2003)].

The goal, as stated, of the updating data assimilation is to determine an optimal state for a given system variable. The reality is that there can be different assumptions or conditions under which "optimal" states are achieved, here the definition of "optimal" is assumed to be any improvement or minimization of errors under any circumstances given.

Updating data assimilation schemes or algorithms are classified into several categories based on the timing of the observations and also the observation processing method that are used to find the optimal state estimate. They are first divided into either sequential, also referred to as real-time or filtering, assimilation or variational, also referred to as retrospective or smoothing, assimilation [Bouttier and Courtier (2002); Reichle (2008)]. Sequential or filtering assimilations only considers the observations that have been made, in real-time, up to the time of assimilation. Whereas smoothing or retrospective assimilation includes observations from times in the future. The assimilation systems can be further defined as either intermittent or continuous at this point [Bouttier and Courtier (2002); Ide et al. (1997); Stauffer and Seaman (1990)]. This categorization was developed because intermittent assimilation systems process the observations in short periods of time or batches and continuous assimilation systems does so over a longer periods in time. Intermittent methods are usually easier to manage computationally but continuous methods produce corrections that are realistically smooth in time [Bouttier and Courtier (2002)]. There are of course many combinations of these methods creating variations on these categorizations, also making differences between systems sometimes hard to distinguish. Most of the assimilation algorithms discussed here will be considered intermittent, and sequential.

The model state estimate or *a priori* state estimate is assumed to be given by a stochastic process or model that has the following form,

$$x_{b,t} = Mx_{b,t-1} + G\mu_t + w_{t-1} \tag{2.1}$$

where M represents the model operator that advances the model forward, G the control input operator, which is optional and assumed to be zero here, μ the control input, and w the model noise. Subscripts with t indicate the location in time of the corresponding value. The

measurement is related to the basic linear model by,

$$z_t = H \cdot x_{b,t} + v \tag{2.2}$$

where v represents the measurement noise.

Given the model for a basic linear model, presented as Equation 2.1, it is assumed that for a given background or predicted state of the model, x_b, the observation operator, H, for any x near x_b, is linear, such that $H(x) - H(x_b) = H(x - x_b)$. It is also assumed that both the observation and model covariances or magnitude of error are non-trivial and positive. Unbiased errors are also assumed, meaning that the expectation of both the observations and model errors is zero, $\mathbf{E}[x_b - x] = \mathbf{E}[z_t - H \cdot x] = 0$, where x represents the true model state and z_t an observation, and that the errors are also uncorrelated, $[x_b - x][z_t - H \cdot x]^T = 0$.

In order to determine the optimal true state estimate, a linear analysis is used to obtain model corrections that are defined by linearly related observation deviations. These corrections are also defined as optimal in the sense that the estimate is at a minimum variance.

The Gauss-Markov Theorem states that when given a linear model with errors that have zero expectation and the errors are also uncorrelated and have equal variances the best linear unbiased estimator, BLUE, is its least squares estimator. The variances are considered equal, which is assumed here, if they come from the same distribution, a normal or Gaussian distribution in this case. The least squares estimator is assumed to have a minimum variance and it is also considered to be a maximum likelihood estimator.

The algorithms considered here, similar to linear regression approaches, are based on this least squares estimator, and are defined by the optimum or smallest sum of the residuals,

$$S = \sum_{i=1}^{n} (x_{b,i} - x_i)^2 \tag{2.3}$$

Weighted least squares is a special form of the least squares method in which the residuals are weighted and forms the BLUE if the weights of those residuals is equal to the inverse of the variance. The weighted residual is commonly referred to as the Mahalanobis distance. For this linear model the residula can be shown as the following cost or objective function [Reichle (2008)],

$$J(x) = \frac{(x_b - x)^2}{\sigma_b^2} + \frac{(z - x)^2}{\sigma_o^2} \tag{2.4}$$

where σ_b^2 and σ_o^2 are the model and observation variance respectively. The least squares in general, attempts to minimize the sum of squares of the errors by setting the gradient of the cost or objective function equal to zero. Weighted least squares follows this form by finding the minimum variance from the minimization of the above cost function, which occurs at the steepest slopes defined using the gradient. The *a posteriori* state estimate is assumed to be,

$$x_{a,t} = x_{b,t} + K(z_t - x_{b,t}) \tag{2.5}$$

Also, if the *a posteriori* error covariance estimate or the updated background error covariance for the model is defined as,

$$\mathbf{A} = \mathbf{E}[(x - x_b)(x - x_b)^T] \tag{2.6}$$

It can be further be represented in terms of the analysis, model, and measurement errors,

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \tag{2.7}$$

The background model error covariance, $\mathbf{B}$, or *a priori* error covariance estimate is considered to be the covariance of the linear state model estimation,

$$\mathbf{B}_t = \mathbf{MB}_{t-1}\mathbf{M}^T + \mathbf{Q} \tag{2.8}$$

where the model error covariance is $\mathbf{Q}$.

Given the updated background model error estimate as $x - x_b$, the aim is to minimize the mean square estimator represented as $E[(x - x_b)^2]$. The equivalent of this is setting the trace of the updated background model error covariance estimate to zero and differentiating it with respect to the Kalman gain, $\mathbf{K}$, $\frac{\partial Tr(\mathbf{A})}{\partial \mathbf{K}}$. When it is then solved for $\mathbf{K}$, the results is,

$$\mathbf{K} = \frac{\mathbf{BH}}{\mathbf{HBH}^T + \mathbf{R}} \tag{2.9}$$

If the Kalman gain is assumed to be optimal then the updated background model error covariance estimate simplifies to,

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B} \tag{2.10}$$

The Kalman gain, K, is assumed to be optimal resulting in the minimum of variances, thus giving the value for the cost function at which it is a minimum.

The advantage of a Kalman filter is in the use of K, the Kalman gain, which allows the measurements and model estimates to be weighted based on their given error levels or covariances. The weighting or gain is then applied to obtain the optimal *a posteriori* estimate, $x_{a,t}$.

This method, as shown, uses *a priori* estimates along with *a posteriori* estimates and a given set of observations to determine the values of particular states. This is a formulation of Bayesian theorem, based on the Bayes rule which states that the likelihood or conditional probability of B, given A is equal to the probability of the A, given B, *a posteriori*, times the probability of A, *a priori*, divided by the probability of B, marginal probability. Since Kalman filter gain matrix results in a minimum variance and a minimum mean square error, which for this Bayesian setting produces what is called a minimum-variance unbiased estimator.

The cost function can be minimized without calculating the Kalman gain, through direct application of Bayes' Theorem, if the assumption that the all the errors are represented by Gaussian probability density functions, *pdf*, is used. In this manner the analysis *pdf* is given as the the product of the observation *pdf*, and the background model *pdf*. The minimization of the cost function is then provided as the inverse log of the analysis *pdf* which by definition of all the assumptions yields the maximum likelihood.

This larger approach provides us with one method for determining the value for which the cost function is minimized. To do this values are found, through iteration, that directly minimize the cost function, by setting its gradient to zero. The process utilizes knowledge of its gradient by applying a minimization algorithm [Holm (2003); Bouttier and Courtier (2002)].

If the gradient of the cost function is set to equal to zero,

$$\nabla J(x) = 0 = 2\mathbf{B}^{-1}(x_a - x_b) - 2\mathbf{H}^T\mathbf{R}^{-1}(z - \mathbf{H}\cdot x_a) \tag{2.11}$$

it can be found that the solution is equal to, equation 2.5, which in matrix form is,

$$x_a = x_b + \frac{\mathbf{BH}}{\mathbf{HBH}^T + \mathbf{R}}(z - \mathbf{H}\cdot x_a) \tag{2.12}$$

and is commonly rewritten as,

$$x_a = (1-K)x_b + Kz, \text{ where } K = \sigma_b^2/(\sigma_b^2 + \sigma_o^2) \tag{2.13}$$

All variational data assimilation schemes are based on this objective function or cost function, J, which relates the disparity between the true value, and the model estimate and observation. The goal of the schemes are to minimize the cost function so that a least squares estimate of the true value can be found. There are several data assimilation studies using this method that have been implemented for soil moisture assimilation [Reichle (2008); Bouttier and Courtier (2002)].

The following assimilation schemes also follow all of the above assumptions, which include, both that the state variable estimates and observations are linear, the estimator itself is unbiased, and the best linear unbiased estimator has a total minimum variance, with errors that are Gaussian, a distribution defined by a mean and standard deviation, and white, given that the distribution has a mean of zero. It should be noted that the following two methods under certain assumptions yield the same result, even though they are conceptually different.

2.3.4 Kalman filter

The most common sequential methods are variations of the basic Kalman filter (KF). The operation of a Kalman consists of tracking the conditional mean of statistically optimal estimation of a state variable [Entekhabi et al. (1994); Walker and Houser (2001)]. This is accomplished through a series of cycles between a measurement correction and update steps [Welch and Bishop (2001); Bouttier and Courtier (2002); Holm (2003)].

The Kalman filter is constructed around the BLUE or least squares concepts [Drecourt and Madsen (2002)], as mentioned above. The Kalman filter is able to pass on error information forward in time through explicit calculation of error covariances [Reichle (2008)]. This allows the filter to address the presence of errors on several levels and as the model is progressing in time.

The Kalman filter is an algorithm that takes the uncertainty of the measurement and the uncertainty of the model estimate and weights their values accordingly. The weighted value is added to the previous model estimate to achieve the optimal state variable or new model estimate [Reichle (2008)].

The equations which form the Kalman filter are divided into two groups: *time update* and

measurement update equations. The first set of equations provide *a priori* estimates of the current state and error covariance while the second set of equations provide the *a posteriori* estimate of the state estimate by including a new observation or measurement. After the measurement update the new measurement has essentially been fused with the *a priori* estimate using the *a priori* error covariance. This process cycles every time there is a new observation, simulating a prediction and correction procedure. The *a posteriori* estimate is used as the new time step *a priori* estimate. [Welch and Bishop (2001)]

There are several assumptions made by the Kalman filter. The first being that the model error consists of Gaussian white noise having a mean of zero and has a covariance value of Q. The second is that the observation error also consists of Gaussian white noise having a mean of zero but has a covariance value of R. Lastly that the state variable is Gaussian and has a mean (X) and covariance (σ) [Walker and Houser (2001); Bouttier and Courtier (2002)].

2.3.4.1 Extended Kalman filter & Ensemble Kalman filter

The Extended Kalman filter (EKF) mainly differs from the Kalman filter in the structure of its model and observation operators. The KF is a applied to linear systems whereas the EKF is its nonlinear equivalent [Welch and Bishop (2001)]. The operators must be linearized in the EKF to account for observations that are typically related nonlinearly to the state variables of interest [Eyre (1997)]. The Extended Kalman filter produces estimations based on the first-order linearized approximations of the non-linear system, typically a Taylor series approximation [Welch and Bishop (2001); Entekhabi et al. (1994)].

The ensemble Kalman filter (EnKF) was developed because higher order statistics of the error covariances are ignored in the EKF resulting in an unbounded error covariance. The EKF requires that error covariances be defined, every time an observation exists, for both the model and observation matrices, making the algorithm computationally expensive to compute for even reasonably sized conditions. To correct this issue a Monte Carlo approach is used in the EnKF to approximate the error covariances. The approximation significantly increases the algorithm efficiency and alleviates the unbounded error covariance problem [Evensen (1994)].

Reichle (2008) listed several formal differences present between the EKF and the EnKF. One

difference being in the manner in which linearizations of the model and measurement operators are determined. The EnKF is capable of taking into account more model errors then the EKF and that the number of EnKF members must be optimized to compare computationally to the EKF. Also present within the EnKF is the ability to consider model and measurement errors that are spatially correlated. While the EKF, during the forecast step, integrates the uncertainty or error covariance of one state estimate, the EnKF is computing in parallel a collection or ensemble of state variable estimates in order to gather a state error covariance. This makes the EKF, for highly nonlinear cases, harder to implement because the matrices are very large. The reduced error covariance present in the EKF is represented as the state estimate itself where as the EnKF uncertainty reduction is provided as the mean of the ensemble state variables or the reduced spread between all of the ensemble members [Reichle (2008)].

Alavi et al. (2009) reports that EnKF can be used more efficiently for highly non-linear systems than the EKF or variational analysis methods. These highly non-linear systems tend to exist and form when trying to assimilate data on large spatial scales in multiple dimensions using remotely sensed data from satellites. The models used for these systems often times use non-linear approaches to relate the measurements or observations to the model estimates.

2.3.5 Optimal Interpolation

Optimal Interpolation, OI, can be best described as a simplified version of the EKF, where the *a priori* error covariance estimate is replaced by an approximation. OI has been widely used for weather prediction and has been useful for oceanographic data assimilation [Ide et al. (1997)]. OI takes advantage of the BLUE by simplifying equations for Kalman gain and prediction by only considering a few of the observations that are geographically close, when considering assimilations that have multiple observations at a given location. OI produces the forecast error covariances, $\mathbf{B}$, using an analytical model.

Two weaknesses arise with OI, one being the formation of a noisy analysis from heavily local data and the second being the computation of the Kalman gain, $\mathbf{K}$, specifically for observations that require use of a linearized observation operator [Bouttier and Courtier (2002); Courtier (1997)]. The required *a priori* observation and first guess errors are assumed to be known.

Other assumptions for OI are that the initial guesses required are unbiased and that the first guess is independent of the observations [Barth et al. (2008)]. Routing OI into four-dimensions, for interpolation of time and space, is relatively easy. Research has been done to computationally improve OI at less cost than using an EKF, specifically for larger data assimilation schemes [Ide et al. (1997)].

2.3.6 Variational Analysis

The general objective of the variational analysis method is to find the optimal linear case for a model by weighting the statistical quality of a minimized deterministic function, *i.e.* cost function, against its minimization [Courtier (1997)]. The method adjusts the initial guess for the *a posteriori* state estimate for the desired time of analysis. Useful information about the guess is usually gathered by first guessing x_a to be equal to x_b.

Variational assimilation schemes often employ advanced numerical methods to minimize the cost function, as opposed to finding where the derivative of *J(x)* is equal to 0, because when trying to minimize *J(x)* often times there are large matrices or vectors that exist, as well as non-linearities within the solution, making analytical solutions impossible [Reichle (2008)]. Often times in variational data assimilations approximations must be made in the analysis to account for the fact that many models within the schemes aren't differentiable and a non-linear operator relating the measurement to the state estimate, known as the adjoint, cannot be found [Liang and Qin (2008)].

The use of an adjoint solution or adjoint operator is computationally convenient because the model and observation operators have been linearized, making solutions possible for the often nonlinear and more complex models that are encountered in earth sciences [Alavi et al. (2009); Bouttier and Courtier (2002); Liang and Qin (2008)].

To account for the non-linearity that can often be present for data assimilation schemes the cost function is minimized using several guesses. However, in the linear case only one minimum exists, making determination of the optimal state estimate relatively plain [Holm (2003)].

Two of the most common methods for minimization of the cost function use simulated annealing or an adjoint solution of the cost function. The non-linearity and discontinuity that

occurs when determining the global minimum of the cost function can often be avoided using the simulated annealing function [Pathmathevan et al. (2003)].

Variational data assimilation schemes range from 1DVAR (one-dimensional) to 4DVAR (four dimensional) based on their inclusion of spatial and time dimensions. Most variational methods assume the model to be 'perfect', denoted as being *strongly-constrained*, and so model errors aren't considered in the cost function. However another term can be included in the cost function to allow for model error and time-changes of boundary conditions, this form is denoted as *weak-constrained* [Reichle (2008)].

Most common 4DVAR data assimilation schemes, which include all space dimensions and the time dimension, assume that the model estimations are perfect [Bouttier and Courtier (2002)]. Liang and Qin (2008) states that although weak-constrained data assimilation schemes can produce poor results if state parameters are incorrect, they can still provide valuable insights into the model which is lost with strong-constrained data assimilations, and can help determine inaccurate state parameters.

The computational differences between the 3DVAR and the 4DVAR are seen because the 4DVAR is smoothing scheme and the assimilation period considers both present and future observations. This produces a state estimation curve that is continuous and smooth in time [Holm (2003)].

In general variational data assimilation schemes typically use a longer period of time than Kalman filter or OI schemes in which to produce an optimal fit considering all observations occurring during this period. This allows information to be propagated forward and backward in time [Bouttier and Courtier (2002); Rihan and Collier (2003)]. During the period of assimilation all the errors are assumed to be independent and additive [Alavi et al. (2009)].

Variational data assimilation schemes have a small cost computationally and are straightforward and require less inputs or analysis to apply [Bouttier and Courtier (2002)]. Beginning a data assimilation system with a 1DVAR allows for a good foundation for developing a 4DVAR scheme [Pathmathevan et al. (2003)].

Analyses done by Ide et al. (1997) reveals that a three-dimensional variational assimilation method has high similarity with sequential methods. 1DVAR data assimilation schemes have

been applied with success for remotely-sensed data [Pathmathevan et al. (2003); Sabater et al. (2007)]. Typically the 3DVAR or 4DVAR are used for soil moisture data assimilation schemes. Less complex 1DVAR and also the more complex 4DVAR have been used successfully to model soil moisture, [Entekhabi et al. (1994); Reichle (2008); Huang et al. (2008)]. 3DVAR data assimilation systems essentially eliminate the need for local data selection of only the near-by observations [Bouttier and Courtier (2002)].

The main drawbacks of the variational schemes are that the model is considered perfect and there is no direct access to the analysis covariances [Courtier (1997)]. Also, the initial guess for error covariance in the Kalman filter is less crucial than the initial guess for the variational methods [Bouttier and Courtier (2002)].

2.3.7 Summary

The Kalman filter is the best conceivable algorithm if we are assuming a linear model and that the errors are white, meaning they are uncorrelated and do not vary in time, and are Gaussian, being defined by the mean or first statistical moment and the standard deviation or second statistical moment [Bierman (1979)].

Reichle (2008) states that the EKF and EnKF are the best considerations for data assimilation schemes because often times adjoint models, usually necessary for variational data assimilation, are often not available or difficult to construct.

Even though Evensen (1994) stated that the linear approximation of the model operator, M, and observation operator, H, in the Kalman filter may produce instabilities or possibility divergence, the Kalman filter has still been used with success on a number of occasions [Walker et al. (2001), Galantowicz et al. (1999), Reichle (2008); Reichle et al. (2008), Entekhabi et al. (1994)]. The Kalman filter has been known to operate well even under circumstances that are necessary for optimality are not present [Welch and Bishop (2001)].

It is known that for large systems where large error covariance matrices exist, the propagation of these matrices slows down the Kalman filter algorithms significantly, except in the case of the EnKF [Drecourt and Madsen (2002); Reichle (2008)]. This occurs typically when large spatial scales are used and also when the horizontal and vertical error covariances.

The Kalman filter, as studied by Walker et al. (2001), was used to retrieve estimates of the entire soil moisture profile. Variations of the EKF have also been frequently chosen to model the one-dimensional vertical soil moisture profiles [Reichle et al. (2002)]. Entekhabi et al. (1994) and Galantowicz et al. (1999) found success with the Kalman filter successful for updating soil moisture profiles.

In some situations where the linear Kalman filter fulfills several assumptions, including that the model and observations errors are assumed uncorrelated in time, it can produce results that are identical to variational methods. For example, the 1DVAR weak-constrained and Kalman Filter schemes have shown experimentally to produce the same result at the end of the simulation period [Reichle (2008)]. The 4DVAR, as well, produces the same state estimations as the Kalman filter at the end of the assimilation window assuming the linear case with no model errors [Holm (2003)].

2.4 Conclusion

One concern about using data assimilation within a model is that changing state variables could theoretically create divergence within the model and possibly halt the simulation prematurely if the assimilated values are significantly different than the model estimates. The most severe case of this would occur when using direct insertion or having model errors that are extremely high. It is the assumption that most algorithms will provide some constraint to the model estimates to prevent this type of issue.

With this knowledge, the next in the process of data assimilation is to implement the Kalman filter again using soil moisture observations to determine what benefit, if any, there is for predictions of yield and biological crop growth, such as canopy weight, especially if poor soil parameters exist.

Hunsaker et al. (2007a) and Hunsaker et al. (2007b) obtained a dataset from two wheat experiments conducted during the winters of 2003-2004 and 2004-2005, which has soil moisture measurements at several depths within the entire soil profile, as well as soil texture information, grain yield and canopy weight measurements. Thorp et al. (2010a) and Thorp et al. (2010b) created a calibrated DSSAT-CSM v4.5 for this location using this dataset. Having this dataset

available and the calibrated DSSAT-CSM [Thorp et al. (2010b)], both of which have been extensively studied and reiviewed, for this research site makes for cost and time effective research and allows for an in-depth analysis of the results and implications against well-documented and published work.

The Kalman filter algorithm has been evaluated as the best option for this research scenario, because the direct measurements of soil moisture allow use of a linear relation to model estimates. It is also assumed that the errors present in the model and measurements are Gaussian and have a mean of zero. This filter has flexibility in implementation, specifically meaning that the model estimates can be updated in real time, as well as the ability to address easily several different sources of error. The algorithm has shown effectiveness in the area of soil moisture already and successful implementation with several models.

CHAPTER 3. Assimilating in situ soil moisture measurements into the DSSAT-CSM using a Kalman filter

a paper to be submitted to Transactions of the ASABE

3.1 Abstract

With the ability to monitor soil moisture in time comes the opportunity to develop ways to incorporate these measurements into predictive models, without compromising or overriding the model physics. The importance of soil moisture to the growth of crops is well understood and because of this it is recognized as one of the more important parts of crop modeling programs. This research focused on improvements to the Decision Support System for Agrotechnology Transfer - Cropping System Model (DSSAT-CSM) as determined by the accuracy of soil moisture estimates. To accomplish this, data assimilation techniques were implemented to process the uncertainty of the model estimates and *in situ* measurements of soil moisture. Consideration of soil parameter uncertainty, which influences model estimates of soil moisture and model output, was taken into account using a Monte Carlo approach. A Kalman filter was used to combine the model estimates of soil moisture with *in situ* soil moisture measurements, while varying several important soil parameters in the model using a Monte Carlo approach. Covariances for the Kalman filter were calculated for the model and measurements based on the model's standard deviation from the Monte Carlo soil moisture estimates and the standard deviation of the *in situ* soil moisture measurements. Data for this study was obtained from a research study conducted on irrigated wheat during the winters of 2003-04 and 2004-05 in Maricopa, Arizona, in which thorough field and crop data were collected. Results of the simulations were compared against biomass and yield measurements to determine the effectiveness

of the data assimilation scheme. The Monte Carlo approach with assimilation done in the top layer of the soil profile was only able to moderately address uncertainty present in the soil parameters. Improvement resulted for data assimilation of soil moisture through the reduction of the error between the measured and simulated grain yield and canopy weight for 47% and 37% of the simulations for the 2003-2004 and for 25% and 32% of the simulations for the 2004-2005 season, respectively. Assimilation was more effective for improving the model output of grain yield for the 2004-2005 than the 2003-2004 season and canopy weight for the 2003-2004 season than the 2004-2205 season. The results of model estimated daily NO_3 levels in the soil layers from data assimilation simulations indicates that assimilation of soil moisture can influence its levels. The data assimilation combined with a Monte Carlo approach showed the use of remotely sensed soil moisture could lead to improvements of frequently studied model outputs, such as grain yield and canopy weight. Further study is needed to fully understand the most desirable conditions for soil moisture assimilation and what other influencing effects data assimilation of soil moisture presents.

3.2 Introduction

The usefulness of crop models is well known, not only for improving economic returns but for gaining knowledge in the research community [Batchelor (2002); Liang and Qin (2008); Ma et al. (2009)]. In precision farming applications crop models have been specifically employed for yield forecasting [Thorp et al. (2010a); Chen et al. (2008)]. Crop models allow researchers and agricultural production managers to make well-informed research and crop management decisions [Tsuji et al. (1998b); Jones et al. (2003); Tsuji et al. (1998a); Bert et al. (2007); Heinzel et al. (2007)]. These decisions can be based upon the model estimates of biomass or grain yield predictions, as well as nitrogen or water balance estimates produced by the model. Inputs related to weather data, soil characteristics, management practices, and the cultivars are the information used by crop models in applications to inform production goals and management decisions like yield forecasting, in season biomass production, irrigation scheduling, and fertilizer application [Tsuji et al. (1998b); Jones et al. (2003); Boote et al. (1998)].

The DSSAT-CSM is a crop modeling system that has been particularly and extensively

utilized for research and agricultural production [Jones et al. (2003)]. The individual crop models within the DSSAT-CSM have successfully simulated crop growth in a wide range of locations and conditions with the results being used for various applications [Sarkar and Kar (2008); Rezzoug et al. (2008); Paz et al. (1998, 1999); Ma et al. (2005, 2006); Casanova et al. (2005)].

Although the DSSAT-CSM has shown to be useful, in some situations the simplified model physics of the crop model aren't able to perform well enough to maintain its usefulness. Because of this researchers have a desire for continued improvements for the DSSAT-CSM and crop models in general [Batchelor (2002); Liang and Qin (2008); Sau et al. (2004); Ma et al. (2007); Stastná and Zalud (1999)]. One area for continued improvement is within model soil moisture estimates [Ma et al. (2007)]. The soil parameters of hydraulic conductivity, saturation, drained upper limit, and lower limit influence improvements of soil moisture estimates because of their influence and importance to soil water behavior [Balland et al. (2008); Stastná and Zalud (1999)].

Soil moisture is of specific importance because it is crucial to many different aspects of crop growth and consequently crop modeling predictions as well [Tsuji et al. (1998a); Jones et al. (2003)]. Correctly simulating soil moisture is important because water is a crucial medium for nutrient transport and exchange, cooling and other processes necessary for plant growth. It has been stated that it is the most important factor for energy balance and flux consideration is stored water [Casanova et al. (2005); Houser et al. (1998)]. Additionally, the drainage of water through the profile influences the on level of nutrient availability. The necessary soil water content for processes such as nutrient and oxygen movement can only be properly simulated if the components of the water budget such as infiltration, runoff, drainage, evaporation, and root water uptake rates are accurate [Tsuji et al. (1998a); Houser et al. (1998); Entekhabi et al. (1994)]. Because soil moisture is particularly relevant to modeling crop growth it requires that we are able to adequately model soil moisture [Tsuji et al. (1998a)].

Soil water estimates are used in the DSSAT-CSM specifically within the Soil-Plant-Atmosphere, Plant, and Management modules making use of the relationship between soil water content and transpiration rates for determination of the water balance [Jones et al. (2003)]. This relation-

ship is one of the many processes that can effect grain yield [Novak et al. (2005)]. The soil water content can indirectly effect yield predictions but more importantly its value directly relates to water availability for in season crop growth and biomass production [Bert et al. (2007)]. Yield forecasting consequently relies on the accuracy of model in season biomass predictions, and more specifically the accuracy model soil moisture estimates.

The CERES-Wheat model is one of several plant sub-modules that are specifically designed to exhibit the behaviors, specifically growth and yield, of an individual plant species or a group of plant species. Each model is capable of simulating crop development and crop growth processes while considering the effects of crop water and nitrogen deficits. Information exchange for conditions such as weather and environment, is controlled by the DSSAT-CSM main program, called the Land Unit Module. The Land Unit Module is responsible for transferring data to the Plant module, and sub-modules such as CERES-Wheat, needed to simulate crop growth. CERES-Wheat has been extensively validated in many different locations including a wide array of soil and climate conditions, and varieties. Its successful performance has been well-documented [Jones et al. (2003)]. The CERES-Wheat model has been used successfully with assimilation studies using several types of remotely sensed data other than soil moisture [Heinzel et al. (2007)].

Site specific management (SSM) opportunities, as a result of the growth in precision agriculture techniques and technologies, rely on having knowledge about the level of spatial variability in yield forecasts to improve agriculture and economic returns [Braga and Jones (2004)]. This makes accurate estimates of the state parameters that describe field conditions, specifically soil parameters, critical to properly implement crop models for use in SSM and situations were detailed management practices are used.

There are cases where important soil parameters, such as hydraulic conductivity, are not known or are highly spatially variable, which leads to high uncertainty in model outputs. Accurate yield predictions and soil moisture estimates can result when soil parameters are obtained using soil moisture measurements [Braga and Jones (2004)]. This suggests that soil moisture measurements could be used in data assimilation schemes to address poor model inputs. To conduct a thorough study of the soil moisture estimates provided by a crop model,

it would be best to utilize *in situ* soil moisture measurements throughout the soil profile. This would provide an opportunity for the study of the soil moisture estimates without having to rely only on model physics when soil parameter uncertainty is present.

To incorporate measurements of state variables such as soil moisture, within modeling, several data assimilation techniques have been developed [Maas (1988); Moulin et al. (1998)]. State variables are model conditions or estimates representing biological and environmental quantities that change in time. The main purpose of data assimilation is to provide more accurate model estimates which ultimately lead to better model predictions. Data assimilation aims at improving model predictions by reducing the model output variation or producing more accurate model output. Variation of model output can be high due to poor state parameter estimates, initial condition estimates or model physics.

Data assimilation techniques are divided in four general approaches: forcing, updating, re-initialization, and re-parameterization. The measurements are fused with the model in a different manner for each technique. Direct insertion or forcing schemes simply replace the state variable of interest within the model overwriting the model estimates altogether at every time step of the model. Direct insertion has been employed over the years for meteorology as well as hydrological applications [Alavi et al. (2009)] and was able to influence the model performance and output, providing useful insight into model behavior [Walker and Houser (2001); Walker et al. (2001)]. Updating techniques differ from direct insertion because they do not always completely replace the model estimates but use measurements to correct or update the state variable estimate only when measurements exist. Updating schemes also employ methods that take into consideration the uncertainties that exist with the model's estimates of state variables estimates as well as the measurements in order to produce a more optimal state variable estimate [Reichle (2008)]. Updating schemes have shown to be a better alternative to direct insertion [Heathman et al. (2003)]. Updating schemes have been used for the assimilation of soil moisture because of their flexibility to handle various levels of model and measurement uncertainty due to varying spatial scales. Re-parameteritization and re-initialization are similar in the fact that they both apply optimization algorithms to optimize the initial conditions (model start values for the state variables) and state parameters (environmental conditions that generally

don't change in time). State parameters are different from state variables in the fact that state parameters are model conditions or estimates representing biological and environmental quantities that are assumed to remain constant. Through iteration these algorithms adjust the initial conditions or state parameters until a minimum difference between the measured and model estimated state has been obtained.

The Kalman filter is an updating data assimilation schemes that takes into account uncertainties that are present in the model estimates as well as measurements. Variational analysis methods such as three dimensional variational analysis (3DVAR) and four dimensional variational analysis (4DVAR) typically don't consider model uncertainty and also process a collection measurements over a longer period of time calculating more accurate model estimates in batches. More advanced updating assimilation schemes that use variations of the Kalman filter take into account uncertainties defined in multiple dimensions and on a larger spatial scale [Reichle (2008); Anderson (2001)]. The ability to consider these uncertainties make the use of the Kalman filter more appealing than other assimilation schemes such as variational analysis methods [Galantowicz et al. (1999)]. Kalman filters are appealing because it understood the model have uncertainty due to simplifications or poor model inputs and measurements have uncertainty due to natural random noise. Kalman filters operate in time with the model real-time offering real-time updating of model estimates also making their use more desirable over variational analysis methods which typically operate over longer periods processing the data in small batches. Accounting for both model and measurement errors allows for a less uncertain or varied state estimation within the crop model. The Kalman filter is applicable for assimilating either *in situ* or remotely sensed data [Huang et al. (2008)].

All of these data assimilation techniques have been explored with soil moisture measurements with various levels of success. Huang et al. (2008) reported a promising study using the combination of a crop model, other than CERES-Wheat, and an ensemble Kalman filter utilizing observed soil moisture measurements. Similar experiences have been reported by Reichle et al. (2008, 2002), Burgers et al. (1998); Huang (2004); Dewit and Vandiepen (2007); Kumar and Kaleita (2003), and Koo et al. (2007) focusing on soil moisture as well as other significant model state variables. Research has been conducted with success focusing on soil moisture as-

similation in one dimension using a Kalman filter [Galantowicz et al. (1999); Walker and Houser (2001); Walker et al. (2001)]. However, assimilation of soil moisture into the CERES-Wheat model found in the DSSAT-CSM and consequently its ability to effect yield have not yet been studied.

The objectives of this research were:(1) to develop a Kalman filter strategy for assimilating soil moisture into the DSSAT-CSM, (2) to evaluate the improvement towards measured of assimilation predictions over model-only predictions for grain yield and canopy weight using this data assimilation scheme, and (3) to use a Monte Carlo approach to improve model output accuracy by reducing the variation of model output when soil parameter uncertainty is present.

3.3 Methods and Materials

3.3.1 CERES-Wheat

The CERES-Wheat module, in the Plant Module of the DSSAT-CSM, has broken down the wheat plant life cycle into seven development stages: germination, emergence, terminal spikelet, end ear growth, beginning grain fill, maturity and harvest. Growing degree days calculated using the maximum and minimum daily temperatures determine the the rate of development. Growth stage progression is dependent on either user defined days or internal calculations.

Consideration of dry matter is part of the module's physics as well. Daily intercepted light, based on LAI, plant population and spacing, is converted to dry matter using a radiation use efficiency parameter. Dry matter totals for each day are also influenced by water, nitrogen, temperature, CO_2 concentration, depending on the most limiting.

In crop models soil characteristics are viewed as some of the most crucial model parameters [Stastná and Zalud (1999)], but they are often difficult and costly to obtain. They also tend to have high uncertainty due to high spatial variability [Chirico et al. (2007)]. After collecting soil texture data soil parameters are commonly obtained using a pedotransfer function (PTF), such as ROSETTA [Schaap et al. (2001); Wosten et al. (2001)]. Soil parameters produced by PTFs have been validated by using fitted retention curve data [Romano and Santini (1997)]. This

method however can cause soil parameters, like hydraulic conductivity and field saturation to produce values to result in less then accurate energy balance estimates [Ma et al. (2009)].

3.3.2 Kalman filter as a data assimilation algorithm

As described by Reichle et al. (2002) advanced data assimilation techniques, including the ensemble Kalman filter (EnKF) and 4DVAR algorithms, vary in their ability to account for both model or measurement uncertainties. The Kalman filter algorithm specifically accounts for both, and is one reason why it was chosen as the data assimilation technique for this research. Since directly soil moisture measurements are being used resulting in a linear relationship to the model soil moisture estimates, the Kalman filter (KF) is considered to provide the most optimum results, having an estimate with the least amount of variation, if both model and measurement uncertainties are considered. For this situation the EnKF would produce identical estimates to the KF. Variational filtering algorithms, such as 3DVAR and 4DVAR, at the end of their assimilation period will provide the same estimates as the KF, if the model is considered perfect, having zero errors.

When the KF must process large error or uncertainty matrices variational analysis methods cost less computationally because they process all of the data simultaneously in the given assimilation window. However, when the measurements arrive continually, the KF methods allows for real-time data assimilation. The KF methods are also able to provide uncertainty information about the filter's optimal estimates, something the variational analysis methods do not provide [Alavi et al. (2009)].

Since the DSSAT-CSM runs on a one day time step and the soil moisture measurements are infrequent, the Kalman filter for this research will update model soil moisture estimates on a non-continuous basis sequentially, meaning only measurements made up to the model time step will be considered. This will then only update the model soil moisture estimates whenever there is an *in situ* measurement available. The Kalman filter requires uncertainties from both the model and measurements, are their determination is crucial to a properly operating Kalman filter.

The Kalman filter considers the magnitudes of the covariances, which estimate the amount

of noise or errors likely to be present, for the model estimates and measurements to track the mean of the optimal state estimate. These estimates are assumed to be represent a value that could be taken from a distribution described by the associated covariances. The linear stochastic difference equation, on which the Kalman filter is based, represents a process, such as soil moisture. The soil moisture value in time, x_t, is represented by,

$$x_t = Ax_{t-1} + B\mu_k + w_{t-1} \tag{3.1}$$

it is also assumed that the model has a linear relationship to a measurement, z_t, that is represented by,

$$z_t = Hx_t + v_t \tag{3.2}$$

H, the observational operator relates the value of the measurement to the state variable. The measurement and model noise or error are represented by v_t and w_t, respectively. In Equation 3.1 A is the model operator and propagates the model estimates forward in time. The input control variables μ_t, the optional control input, and B, the input control operator, represent outside influence on the model. The model time-steps are represented using the subscripts t for the current time-step and $t-1$ for the previous time-step.

If we were to represent the present research conditions using the general model given by Equation 3.1 the right-hand side would represent the DSSAT-CSM model processes that propagate the state estimates of soil moisture forward in time. The measurement operator, H, is equal to 1, since we have direct measurements of soil moisture.

In the case of non-linearly related model estimates and measurements, H generally consists of a Taylor series approximation relating the measurements to the model estimates [Entekhabi et al. (1994)]. The Kalman filter in this circumstance is referred to as the Extended Kalman filter or EKF.

The model and measurement noise are assumed to be uncorrelated, or independent of each other. They are also assumed to be white (having a mean of zero) and Gaussian or a distribution fully described by a mean, μ, and standard deviation, σ. The Kalman filter estimate, under these assumptions, is assumed to be the optimal least squares estimator or best linear unbiased

estimator, and providing an estimate having a variation that is considered to be at its minimum value.

Covariances representing the model and measurement errors, w and v, are given by Q and R in Equations 3.3 and 3.4, respectively.

$$pdf(w) \sim N(0, Q) \tag{3.3}$$

$$pdf(v) \sim N(0, R) \tag{3.4}$$

In order to account for the noise or errors, w and v, that are present in Equations 3.1 and 3.2, the Kalman filter process consists of *time update* and *measurement update* equations. The forecasting or time update equations producing *a priori* estimates, for the Kalman filter are,

$$\hat{x}_t^- = A\hat{x}_{t-1} + B\mu_t \tag{3.5}$$

$$P_t^- = AP_{t-1}A^T + Q \tag{3.6}$$

followed by the measurement update equations which generate *a posteriori* estimates,

$$K_t = P_t^- H^T (HP_t^- H^T + R)^{-1} \tag{3.7}$$

$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - H\hat{x}_t^-) \tag{3.8}$$

$$P_t = (I - K_t H)P_t^- \tag{3.9}$$

where hats ($\hat{}$) represent state variable estimates and the superscript minus ($^-$) represents an estimate that is made prior to a measurement update. The model soil moisture estimate is given as $\hat{x}_t^-$ and the optimal soil moisture as $\hat{x}_t$. The overall Kalman filter error covariance is represented by P, and the Kalman gain by K.

The time update equations, Equations 3.5 and 3.6, can also be viewed as prediction equations for the model only estimates and likewise the measurement update equations, Equations 3.7, 3.8, and 3.9 , can be viewed as correction equations to those model estimates.

The advantage of a Kalman filter is in the use of K, the Kalman gain; defined by Equation 3.7, which allows the measurements and model estimates to be weighted based on their given error levels or covariances. The weighting or gain is then applied to obtain the optimal

estimate $\hat{x}_t$, the *a posteriori* state estimate. When the Kalman gain is calculated, the measurement is weighted higher if R, the measurement error, is small making the Kalman gain larger. Conversely, the model estimate is weighted higher when Q, the model error, is low or the measurement error, R, is high. Lower Kalman gain results from a lower *a priori* covariance estimate, P_t^-.

An initial value of the *a posteriori* error covariance estimate, P_0, is required for the Kalman filter. The initial *a posteriori* error estimate was assumed to be equal to Q.

3.3.3 DSSAT Soil Water Balance

The soil water balance is part of the DSSAT-CSM Soil module. The Soil module's sub-modules not only include water balance principles but also soil nitrogen balance principles, the dynamics for soil temperature and carbon balance principles.

The soil water balance for DSSAT-CSM was adapted from CERES-Wheat and was originally developed by Ritchie and Otter (1985) [Jones et al. (2003)]. The soil water balance model for CERES-Wheat was developed as a one-dimensional model, using irrigation, infiltration, vertical drainage, unsaturated flow, soil evaporation and plant root uptake processes to compute the daily water content changes experienced by each layer in the soil profile. Infiltration is simply the difference between precipitation and runoff as determined by the Soil Conservation Service Curve Number [Soil Conservation Service (SCS) (1972)]. Irrigation is considered an additive component of total precipitation.

The DSSAT-CSM uses a water balance method that at times may oversimplify what is actually happening in the local soil profile. The DSSAT-CSM version 4 uses the same SCS CN, method from version 3.5 and earlier. The SCS CN method was designed to estimate runoff from a watershed, but has been modified by Williams et al. (1984) to compensate for soil layers and also for initial soil water content at the time of precipitation.

The curve number method should not be assumed to provide accurate runoff and infiltration values for specific storms [Tsuji et al. (1998a)]. Sadler et al. (2000) also found that the model overestimates infiltration when using the SCS curve number method. So notably, this will affect the accuracy of the DSSAT-CSM's soil moisture estimation and ultimately grain yield

and canopy weight. However, it has been shown that the SCS-CN within the DSSAT-CSM can simulate near-surface (0-30 cm) soil moisture that is in agreement with measured values [Liu et al. (2011)].

To simulate soil moisture profiles the DSSAT-CSM models soil water drainage using a "tipping bucket" method, when the water content is above the upper drained limit. The soil parameter for diffusivity and differences between the adjacent layers' soil water content are applied to calculate the upward saturated flow.

Water accumulates above a soil layer only if the drainage, downward soil water movement, for the layer is greater then saturated hydraulic conductivity for the day, which results in the actual drainage being equal to the saturated hydraulic conductivity for the day. Otherwise the actual drainage through the each layer is assumed to be the calculated vertical drainage for the layer. Drainage through each layer is considered only after a total drainage for the soil profile has been calculated, which is determined by a global soil drainage parameter. Also, as mentioned, the Soil-Plant-Atmosphere module calculates the soil evaporation and plant root uptake. These fluxes are all calculated as equivalent depths and added or subtracted to the soil water content for each layer on each day.

For this given research it was assumed that there was no runoff encountered because all irrigation took place inside dikes, thus the runoff curve number was set intentionally low to allow all of the applied water to infiltrate.

3.3.4 Field Experiments

Hunsaker et al. (2007a) and Hunsaker et al. (2007b) obtained a dataset from two wheat experiments conducted during the winters of 2003-2004 and 2004-2005, containing soil moisture measurements at several depths within the entire soil profile, as well as soil texture information, grain yield and canopy weight measurements. Thorp et al. (2010b) calibrated the CERES-Wheat model in the DSSAT-CSM v4.5 for this location against yield data from this same dataset. The calibrated DSSAT-CSM included adjustments to cultivar parameters and an ET correction. The dataset and the calibrated DSSAT-CSM were used for this study because both have been comprehensively studied and peer-reviewed.

The two wheat irrigation research experiments conducted during the winters of 2003-2004 and 2004-2005 by Hunsaker et al. (2007a) resulted in a corresponding dataset containing soil moisture, soil texture, biomass, and final yield measurements. The wheat fields consisted of 32 equal sized plots with each representing one of 12 different treatments. The soil was mapped as a Casa Grande sandy loam classified to be fine-loamy, mixed, superactive, hyperthermic, Typic Natrargid. The treatments were arranged in a complete random design with incomplete blocking. The main objective of the experiment was to determine the effectiveness of irrigation scheduling based on two different methods, FAO-56 (F) or NDVI (N), for determining the basal crop coefficient, K_{cb}. The basal crop coefficient represents the influence of a specific crop on evapotranspiration (ET). To develop the treatments each of the irrigation schedules included high and low nitrogen applications and three levels of planting densities. The seasonal nitrogen applications were 80 kg N ha^{-1} (L) and 215 kg N ha^{-1} (H). The high nitrogen level is the locally recommended amount for the given sandy-loam soil type. After emergence, which was in early February for both seasons, the nitrogen was injected during irrigation in the form of soluble urea ammonium nitrate (32% N). Planting densities were divided as sparse (S; 75 plant m^{-2}), typical (T; 150 plant m^{-2}), and dense (D; 300 plant m^{-2}) [Hunsaker et al. (2007a)].

Table 3.1

	Experimental Variables			
Subtreatment Abbreviation	K_{cb} Method	Plant Density	Nitrogen Level	No. of Replicates
FSH	FAO (F)	Sparse (S)	High (H)	2
FSL	FAO (F)	Sparse (S)	Low (L)	2
FTH	FAO (F)	Typical (T)	High (H)	4
FTL	FAO (F)	Typical (T)	Low (L)	4
FDH	FAO (F)	Dense (D)	High (H)	2
FDL	FAO (F)	Dense (D)	Low (L)	2
NSH	NDVI (N)	Sparse (S)	High (H)	2
NSL	NDVI (N)	Sparse (S)	Low (L)	2
NTH	NDVI (N)	Typical (T)	High (H)	4
NTL	NDVI (N)	Typical (T)	Low (L)	4
NDH	NDVI (N)	Dense (D)	High (H)	2
NDL	NDVI (N)	Dense (D)	Low (L)	2

Planting of the hard red spring wheat (*Triticum aestivum L.*, cv. Yecora Rojo) took place

on 10-12 December 2003 and 22 December 2004 with row the rows on 0.20 m spacing and a dry soil surface. Irrigation dikes were constructed around all four sides of each plot along with boardwalks across the center, supported by concrete blocks. The boardwalks allowed non-destructive access to the plots as well as to the neutron access tubes located 1.0m away from the center and 3.0m. Time-domain reflectometry (TDR) probes, 0.3m in length, were installed 0.5 m away from the neutron access tubes [Hunsaker et al. (2007b)].

Each of the NDVI plots had irrigations that were scheduled individually, because of the variability of the K_{cb} coefficient, however all 16 of the FAO-56 had irrigations scheduled on the same day. Irrigation was scheduled for the plots the day after the daily soil water depletion of the effective root zone was greater then 45% of the total available water. To account for irrigation inefficiencies, 110% of the estimated depth of soil water depletion was provided. This irrigation procedure was expected to minimize water stress [Hunsaker et al. (2007b)]. This study focused on the FAO-56 treatments just as did the assimilation study performed by Thorp et al. (2010a).

3.3.5 Field Measurements

The crop stages, soil moisture, and soil texture measurements were collected for each of the 32 individual treatment plots in the field, and were collected at different times. Soil moisture measurements from all of these plots were considered for calculation of the soil moisture values used in the assimilation procedure.

To gather biological crop data, destructive measurements were made using a sampling of six plants in different areas with pre-assigned spots based on sampling date located in the northern half of each plot. The biological measurements were taken to monitor wheat growth, development, and ultimately yield. The plant density for each level was able to be verified using these measurements. Measurements were categorized into various plant characteristics by weight, including canopy and grain weight. This biomass data was collected every two weeks and phenology data, every week until the end of the season. The Zadok's number, indicating the stage of plant growth, showed that for the 2003-2004 season maturity occurred near DOY 119 and just before DOY 123 for the 2004-2005 season. The canopy weight measurements

around these dates took place on DOY 111, 125 and DOY 109, 123 for the 2003-2004 and 2004-2005 seasons respectively. For the 2003-2004 season the average canopy weight between DOY 111 and 125 was used for comparison with the simulations and for the 2004-2005 season the measurement from DOY 123 was used for comparison. The average maturity date in 2003-2004 as given by the open-loop or model only simulations and all assimilation simulations occurred on DOY 177 and DOY 123 for the 2004-2005 season.

Soil moisture measurements were, on most occasions, collected weekly but also two to four days after, the day or morning before each irrigation. Measurements were taken starting the day before the first post-planting irrigation was scheduled. Measurements were taken for the top 30 cm using time domain reflectometry (TDR, Trase1, Soil Moisture Equipment Corp., Santa Barbara, Cal.) [Hunsaker et al. (2007a)] and below 30 cm at 20 cm intervals down to 290 cm using neutron probes(model 503, Campbell Pacific Nuclear, Martinez, Cal.). Both probes were calibrated using gravimetric soil samples and achieved volumetric soil water content accuracies of 0.02 m^3 m^{-3} .

The soil moisture measurements were processed to determine the average across the whole field. The average soil moisture across the field was only calculated and used for assimilation if there were at least half of the 32 plots that reported a soil moisture measurement on any day. The average soil moisture for the field was calculated on these days for each individual soil layer. Out of the 51 days that had soil moisture measurements for at least one plot, only 32 days had at least half of the plots with a soil moisture measurement for the 2003-2004 season. Hence, the data assimilation for the 2003-2004 season used 32 soil moisture measurements. For the 2004-2005 season a total of 48 days had any plots with a soil moisture measurement, and 33 out of those 48 were used for data assimilation for the 2004-2005 season.

Soil texture data was also collected for each plot. This texture information was used in the ROSETTA [Schaap et al. (2001)] program to determine characteristic soil parameters, specifically hydraulic conductivity, saturation, drained upper limit and lower limit.

Weather data throughout the experiment was provided by a University of Arizona AZMET weather station approximately 200 m away from the field site. An AZMET technician regularly inspected the station to ensure it was operating properly.

Complete senescence for each season occurred on DOY 135, and DOY 138, respectively. The wheat was harvested and grain yields were collected on DOY 147 for both seasons. The samples were collected from the south half of each plot having areas measuring 24 m^2.

3.3.6 Kalman filter evaluation

Thorp et al. (2010a) used calibrated cultivar parameters for CERES-Wheat from Thorp et al. (2010b) for a data assimilation strategy using leaf area index (LAI) was used to account for uncertainties present in the model inputs. Despite having comprehensive field data the calibrated model output from the assimilation of LAI still resulted in some high degree of uncertainty. An updating data assimilation scheme based on a Kalman filter was selected so uncertainties of both the model and measurements could be considered, and help to address errors relating to large variation of model output and also model output that differed from measured.

The simulations consisted of two unique configurations:

1. open-loop (calibrated model only)

2. calibrated model with the Kalman filter

Simulations for each of these configurations were completed for the 2003-2004 and 2004-2005 season using each of the FA0-56 treatments. The simulations made use of Monte Carlo methods which included the selection of sets from soil parameter distributions of hydraulic conductivity, saturation, drained upper limit and lower limit. Model output of the wheat grain yield and final canopy weight was collected from these simulations for analysis.

The DSSAT-CSM estimates soil moisture states every day for each layer but was updated for this research to also calculate an *a priori* estimate of error covariance, P_t^-, every day as well. On days with no measurement meaning no assimilation was done, the *a priori* estimate of error covariance was set equal to the equal to the *a posteriori* estimate, allowing the estimate of error covariance to be tracked in time. This procedure is considered to be the time update equations. On days with measurements a separate calculation was included and this calculated

the optimal model estimate and an *a posteriori* based on the calculated Kalman gain given the measurement value. These calculations represent the measurement update equations.

Soil moisture measurements for the entire DSSAT-CSM soil profile were available [Hunsaker et al. (2007b)]. Only the measurement of the top 30 cm was used for the assimilation. Because soil moisture in the soil layers is estimated using water balance methods based on soil parameters, the soil moisture in the top of the soil profile governs the soil moisture in the lower layers of the profile and is important to obtaining an accurate soil moisture profile. If accurate soil parameters are used experiments could focus on using only measurements in the top few layers because of this relationship. The daily water balance across the entire profile prior to assimilation was compared the water balance across the entire profile after assimilation to judge how significantly the soil water content was changing. Nitrogen concentration influences crop growth and stress , ultimately impacting grain yield and canopy weight. The concentration of nitrogen can change with the presence of soil water, so daily NO_3 levels were compared between the open-loop and data assimilation simulations.

The DSSAT-CSM uses a default soil layer structure with three layers in the top 30 cm. Data assimilation schemes were configured for use of two soil layer structures with first being the default soil layer structure and the second being a soil layer structure where instead of three layers in the top 30 cm there were only two. The two-layer structure included layers 0-3 cm and 3-30 cm. This layer structure was used because the soil moisture measurements from 0-30 cm would be best represented by the soil water partitioning given by this soil structure. Warnings are given by the DSSAT-CSM to maintain a top layer that isn't much larger or smaller than 5 cm because of the instabilities that it could cause within the soil water balance functions. Since measurements are given as volumetric water content across the 0-30 cm depth, the measurement assimilated into a layer of any thickness in the top 30 cm is assumed to have the same value as the volumetric water content for each 0-30 cm measurement. This is because the measurements do not provide any information to how the water is partitioned in the top 30 cm.

The improvement of the model output for grain yield and canopy weight from these simulations towards the measured values, over the open-loop simulations' output, combinations of

	Top 30 cm Layer Structure	
# of Layers Assimilated	3 Layers (5 cm, 15 cm, 30 cm)	2 Layers (3 cm, 30 cm)
1	0-5 cm	0-3 cm
	5-15 cm	3-30 cm
	15-30 cm	-
2	0-5,5-15 cm	0-3, 3-30 cm
	5-15, 15-30 cm	-
3	0-5, 5-15, 15-30 cm	-

Table 3.2: Kalman filter simulation combinations

different layers will assimilate the soil moisture measurements into the respective layers.

Table 3.2 shows the combinations of the data assimilation schemes that were considered for the top 30 cm given two soil structures.

3.3.7 Determination of the Covariances

The Kalman filter requires a model error covariance as well as a measurement error covariance for calculating the optimal estimate. The model and measurement errors are assumed to be represented by a Gaussian distribution allowing the Kalman filter to be qualified as optimal for our case. Random processes in nature are modelled well by Gaussian distributions [Welch and Bishop (2001); Bierman (1979)]. This supports the assumptions necessary for applying the Kalman filter are not violated when used for these conditions. The model and measurement error distributions can be described by their respective covariance, Q and R.

A measurement error covariance, R, is required by the Kalman filter to be used for the measurement update equations. Common sources of noise or error that are introduced into measurements include random electrical noise and degradation because of the physical limitations of the device [Welch and Bishop (2001)]. Most often the covariance of measurements is dependent on the physical characteristics of the measuring device [Walker et al. (2001); Galantowicz et al. (1999)]. Even devices such as TDR probes that have been calibrated using gravimetric data in the lab can have responses that differ from similar field measurements [Zhang and Van Geel (2007)].

To quantify the measurement errors to determine an error covariance, R, a standard deviation for the measurements on the same days, for each layer across the field was calculated. Just as with the soil moisture measurement averages, the days in which less then 15 plots reported a soil moisture measurement weren't included in the determination of the standard deviation. The measurement error covariance was calculated from the measurements in 2003-2004 only and assumed to be the same for the 2004-2005 season. The way in which the measurement covariance was determined represents random noise that was produced during the measurement process. The measurement covariance wasn't adjusted any further because it sufficiently captured the random noise expected to be present and arbitrarily increasing the covariance more could add redundant and needless error. We assume that the error covariance doesn't directly relate to spatial variability of the soil moisture measurements. This assumption was based on the idea that we dealt with a relatively small scale and the soil moisture was averaged over the entire field. It does however account for sensor error as well as random noise experienced by the sensor.

The Kalman filter requires a covariance of the model error, Q, for the time update equations and representing errors present in the model. The model error arises from several sources, but can be hard to quantify. Models will inherently have errors due to the linearization or simplifications of the state physics; they also suffer error due to inaccuracies present in input data and state parameters. Estimates of these errors are often hard to determine and are usually determined ad hoc [Walker et al. (2001); Evensen (1994); Alavi et al. (2009)].

The model error covariance, Q, calculation was done in a similar manner to the determination of the measurement error covariance. An open-loop model simulation for each of the 32 treatments plots was produced and the standard deviation across each layer on every day throughout all simulations was found. The standard deviation was again averaged over the whole soil profile. This calculation was considered to estimate the errors associated with poor input parameters and in part estimates the errors of the model in response to these poor input parameters. It was not assumed to be associated with the natural spatial variability of the soil moisture. Each simulation used soil parameters that were derived from the soil texture measurements taken for each of the 32 corresponding plots.

3.3.8 Soil Parameters

The soil parameters produced by ROSETTA resulted in variability across all of the plots that was assumed to influence the soil moisture estimates and ultimately the grain yield and canopy weight. The average and standard deviation of the values across all of the plots for the hydraulic conductivity, saturation, drained upper limit and lower limit were used to create distributions for each parameter. Creation of these distributions was the first step taken to implement a Monte Carlo approach. The Monte Carlo approach was used to address the uncertainty of the soil parameters and to lessen the effect of the uncertainty on soil moisture and ultimately grain yield and canopy weight. This in part due to the natural spatial variability of the fields collected texture measurements but also to limitations of the ROSETTA model [Balland et al. (2008)]. To address the variability of these soil parameters a Monte Carlo sampling method was applied to produce sets of soil parameters that would be used in the simulations. These distributions would result in a spread of model output that would be based on the soil parameter uncertainty.

The soil parameter distributions had sample sizes of 1,000 for each of the soil parameters: the lower limit (SLLL), drained upper limit (SDUL), saturation (SSAT), and hydraulic conductivity (SKSS). These distributions were based on their respective mean and standard deviations resulting from the ROSETTA soil parameter calculations. In Table 3.3, the desired means and standard deviations for the soil parameters are listed. The distributions maintained their physical relationships meaning that for a given simulation SSAT $>$ SDUL $>$ SLLL. The mean field soil parameter values were used by Thorp et al. (2010a) in the field average soil profile to obtain the results of the simulations for each of the FAO-56 treatments: FSL, FSH, FTL, FTH, FDH, FDL.

Soil Parameter	Description	Mean	Standard Deviation
SKSS	Hydraulic conductivity (cm hr^{-1})	1.65	0.53
SSAT	Saturation (mm mm^{-1})	0.407	0.005
SDUL	Drained upper limit (mm mm^{-1})	0.236	0.027
SLLL	lower limit (mm mm^{-1})	0.091	0.019

Table 3.3: Means and standard deviations across the field for the soil parameter distributions given by the ROSETTA calculations

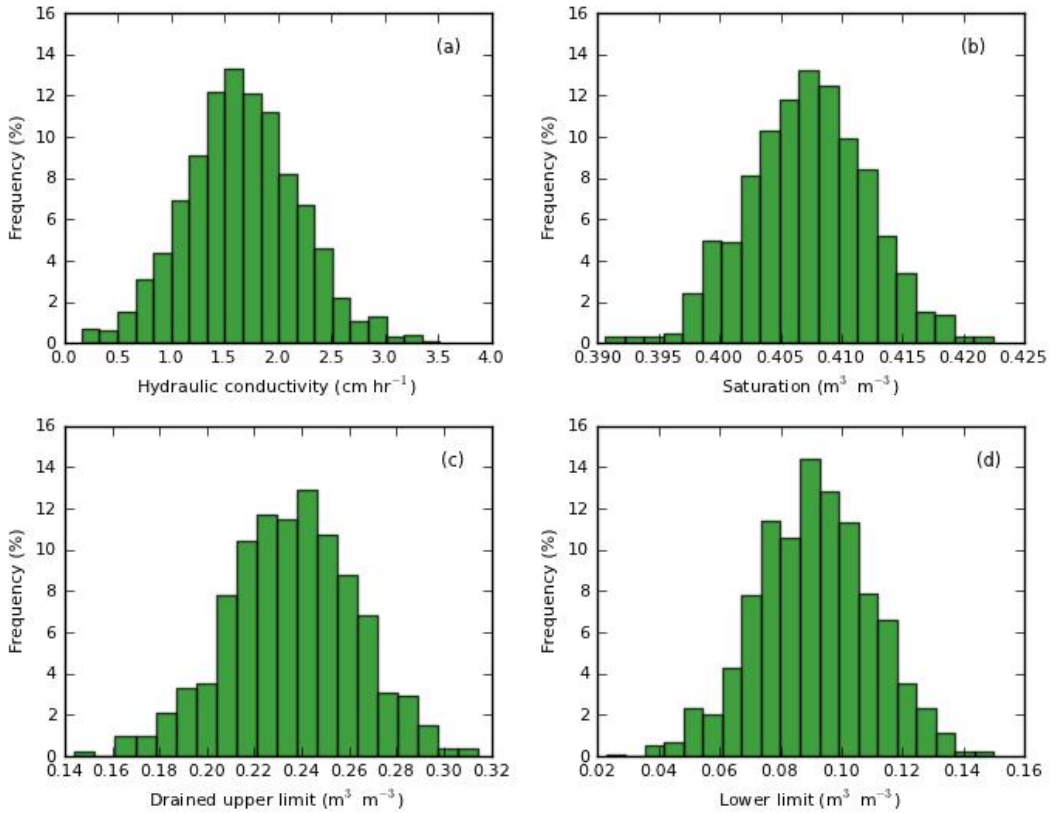

Figure 3.1: Simulated distributions of the four soil parameters used for the Monte Carlo approach

The DSSAT-CSM default soil profile structure consists of 10 layers: a small top layer (5 cm), a 10 cm second layer, followed by two 15 cm layers, and after that layers end every 30 cm, down to 210 cm. The soil parameter distributions were only created and applied for the top 30 cm soil layers and the lower layers used the ROSETTA field average values for their respective soil parameters.

It was confirmed that the 1,000 member distributions created for the each of the soil parameters matched the desired values (Table 3.3). Figure 3.1 shows a histogram for each distribution. Having sufficiently large distributions ensures that the Monte Carlo sampling method would represent the domain of plausible values for each parameter.

Evaluation of the data assimilation schemes was based on results from sets of 1,000 simulations that were completed for each of the FAO-56 treatments for the open-loop and data assimilation configurations. The results for grain yield and canopy weight from the data assimilation Monte Carlo simulations were evaluated using the improvement towards measured values over the open-loop output, and by the reduction of simulation output variability over the open-loop simulations.

When the model output for grain yield or canopy weight results in improvement towards measured values then the model is considered to have more accurately simulated grain yield and canopy weight. The data assimilation shows that it is capable of overcoming uncertainty in soil parameters if the standard deviation of the simulations are lower for the data assimilation than for the open-loop simulations. Percent improvement indicates how many of the simulations, represented by the spread of soil parameters, the data assimilation simulations produced improvement of model output over the open-loop.

3.4 Results

Both seasons resulted in grain yield averages that were closest to measured for the data assimilation scheme consisting of three top 30 cm layers with assimilation occurring into the second layer. Slightly different responses in average difference of grain yield from measured were seen between the two seasons for the data assimilation scheme consisting of three top 30 cm layers with assimilation occurring into the third layer. In the 2004-2005 season it was the lowest but this was not the case for the 2003-2004 season.

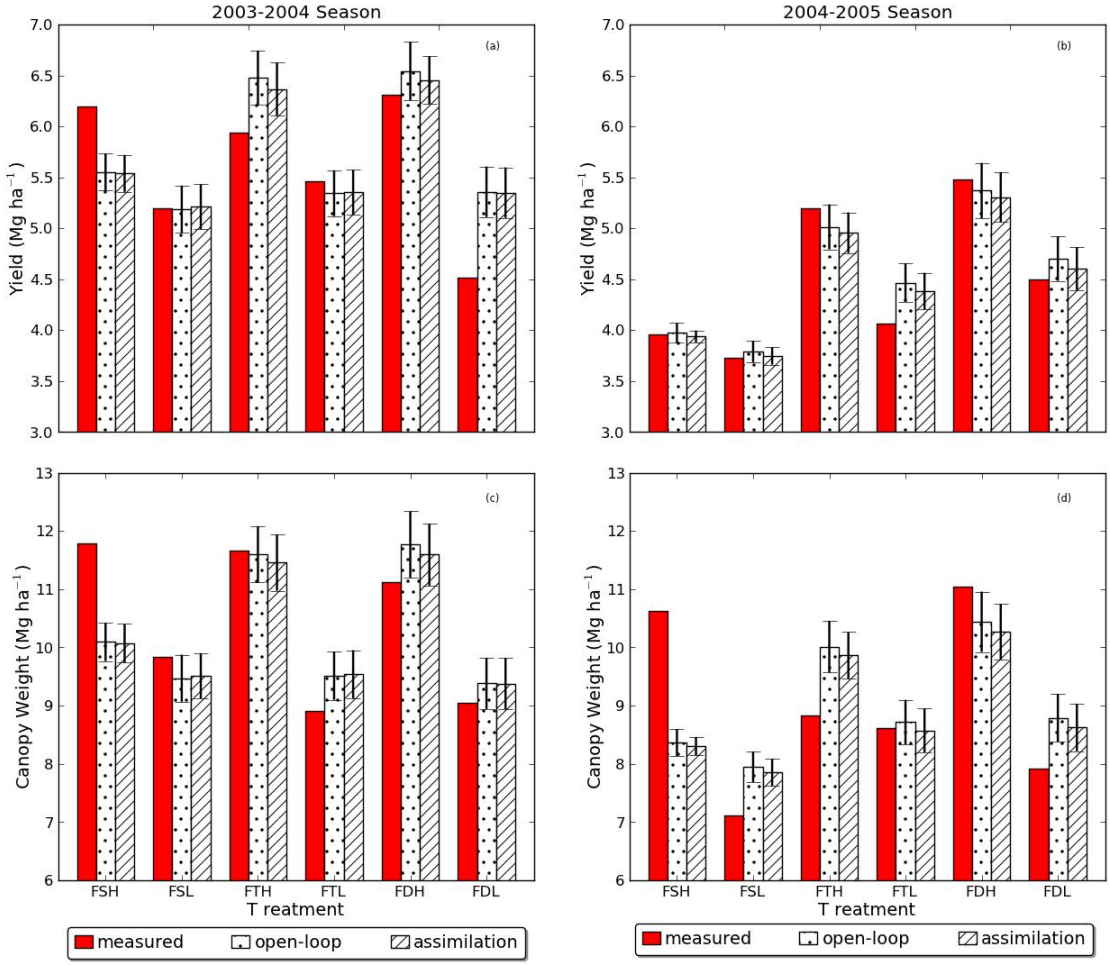

Figure 3.2: The model output for both seasons is from assimilating into the first layer of two top 30 cm layers. The six FAO-56 (F) treatments are divided by planting densities of sparse (S), typical (T), and dense (D) and by high (H) and low (L) applied nitrogen level. The error bars represent the standard deviation across 1,000 simulations

Layers Assimilated	2003-2004 Model Only Mean (kg ha^{-1})	2003-2004 Model Only SD (kg ha^{-1})	2003-2004 Assimilation Mean (kg ha^{-1})	2003-2004 Assimilation SD (kg ha^{-1})	2004-2005 Model Only Mean (kg ha^{-1})	2004-2005 Model Only SD (kg ha^{-1})	2004-2005 Assimilation Mean (kg ha^{-1})	2004-2005 Assimilation SD (kg ha^{-1})
0-5 cm	286	182	244	176	130	122	152	114
5-15 cm			324	192	122	113	136	198
15-30 cm			383	227	129	132	141	250
0-5, 5-15 cm	377	185	362	194	129	132	163	239
5-15, 15-30 cm			408	283	106	121	210	329
0-5, 5-15, 15-30 cm			404	335	129	132	322	377
0-3 cm	259	261	193	239	162	184	146	163
3-30 cm	396	240	407	280	182	199	226	350
0-3, 3-30 cm			423	302	162	184	208	363

Table 3.4: The average difference of simulated grain yield for each respective treatment from measured grain yield along with the average standard deviation across all treatments for each layer combination

Layers Assimilated	2003-2004 Model Only Mean (kg ha^{-1})	2003-2004 Model Only SD (kg ha^{-1})	2003-2004 Assimilation Mean (kg ha^{-1})	2003-2004 Assimilation SD (kg ha^{-1})	2004-2005 Model Only Mean (kg ha^{-1})	2004-2005 Model Only SD (kg ha^{-1})	2004-2005 Assimilation Mean (kg ha^{-1})	2004-2005 Assimilation SD (kg ha^{-1})
0-5 cm			606	309	905	241	897	240
5-15 cm			655	343	942	387	952	227
15-30 cm			720	375	1191	235	1216	485
0-5, 5-15 cm	594	301	686	366	1159	253	1174	473
5-15, 15-30 cm			855	598	1159	236	1184	622
0-5, 5-15, 15-30 cm			943	620	958	253	1019	801
0-3 cm	362	486	337	469	974	375	935	342
3-30 cm	620	441	806	679	1002	397	980	700
0-3, 3-30 cm			869	645	974	375	924	731

Table 3.5: The average difference of simulated canopy weight for each respective treatment from measured canopy weight along with the average standard deviation across all treatments for each layer combination

The simulated data assimilation output of canopy weight for the average difference from measured was closer to measured than the open-loop average difference from measured for 3 of the 4 schemes in which assimilation was done only in the top layer. The average canopy weight and grain yield for both open-loop and data assimilation schemes were both on average 6% away from measured in the 2003-2004 season. For the 2004-2005 season data assimilation and open-loop schemes had average canopy weights that were on average 10% away from measure and average grain yields that were on average 3% away from measured.

The assimilations into the first layer produced interesting results in the 2003-2004 season by having, on average, the lowest difference from measured values for both yield and canopy weight for a data assimilation scheme consisting of two layers in the top 30 cm. This is surprising because it is unlikely that the average measured soil moisture values accurately represent this layer. The measured soil moisture values represent an average over the entire top 30 cm depth, where the water in this soil layer is likely to be lower in the soil right before irrigation events, and also represents a small percentage of water over the entire 30 cm layer even if it were saturated. The model output of grain yield and canopy weight shows improvement when soil moisture was assimilated into layers within 0-15 cm, especially in the top 5 cm. In 2004-2005 the average difference from measured for model output of grain yield and canopy weight across data assimilation schemes had a response more fitting to what would be expected by the soil moisture assimilation. That is, assimilation into the 30 cm layer alone mostly resulted in the grain yield and canopy weight averages that were closest to measured.

Both seasons had only a data assimilation scheme that was able to reduce the standard deviation and an average model output closer to measured than the open-loop with a soil layer configuration of two top 30 cm layers and assimilation done into the layer from 0-3 cm. Figure 3.2 shows the average grain yield and canopy weight by treatment of the open-loop and the assimilation schemes that showed these improvements, alongside the measured values.

Reduction of standard deviation would reflect that using the Monte Carlo approach within the data assimilation was able to address uncertainty in the soil parameters. The standard deviation in Table 3.5 and 3.4 was lower for the simulated grain yield than for the simulated canopy weight when comparing corresponding schemes. The data assimilation schemes with

Layers Assimilated	2003-2004 (%)	2004-2005 (%)
5 cm	52	38
15 cm	54	18
30 cm	46	12
5, 15 cm	52	20
5, 30 cm	44	11
5, 15, 30 cm	46	13
3 cm	37	62
30 cm	47	21
3, 30 cm	44	30
Overall	47	25

Table 3.6: Percentage of simulations where the assimilation grain yield was closer to measured than the open-loop averaged over the treatment simulations

Layers Assimilated	2003-2004 (%)	2004-2005 (%)
5 cm	42	52
15 cm	39	26
30 cm	38	16
5, 15 cm	44	24
5, 30 cm	34	21
5, 15, 30 cm	32	22
3 cm	35	57
30 cm	36	28
3, 30 cm	35	38
Overall	37	32

Table 3.7: Percentage of simulations where the assimilation canopy weight was closer to measured than the open-loop averaged over the treatment simulations

lower standard deviations are the data assimilation schemes that assimilate soil moisture into a layer of either 0-5 cm or 0-3 cm. No data assimilation scheme simulations resulted in a reduction of the standard deviation for simulated canopy weight where the top 30 cm consisted of three layers. Some reduction of standard deviation was seen for data assimilation simulations using two top 30 cm layers.

Standard deviations for 2003-2004 season when considering grain yield for three top 30 cm were not lower when compared to the open-loop for simulated grain yield. Data assimilation into the first of two top 30 cm layers for both seasons showed reduced standard deviations. Data assimilation into the first layer for a three top 30 cm soil layer structure for the 2004-2005 season had reduced standard deviations.

The model output shown in Figure 3.2 was created by the data assimilation scheme that had the highest percentages of improvement. The percentage of improvement was considered to be the number of data assimilation simulations that were closer to the measured grain yield or canopy weight out of the entire set of simulations. The percentage of improvement trended along with the amount of reduction of standard deviation when compared to the open-loop standard deviation.

The same trends didn't seem to exist for both seasons when considering nitrogen application

or plant density for the treatments. The percentage of improvement for grain yield was higher than for canopy weight in the 2003-2004 season. For the 2004-2005 season the canopy weight had a higher percentage of improvement than the grain yield. Typically for treatment simulations in both seasons resulting having a high percentage of improvement for one model output had the other model output with a much lower percentage of improvement. In the 2004-2005 two treatment simulations resulted in percentages of improvements for both canopy weight and grain yield that were around 50%.

Large differences from measured could be seen for the data assimilation scheme and the open-loop simulations, specifically treatment FSH, in Figure 3.2. When the open-loop simulated output for specific treatments for grain yield and canopy weight was closer to measured than the data assimilation simulated output it was only marginally better.

Statistical differences between open-loop and data assimilation simulations was determined if the p-value was lower than 5% for a rank-sum test. All of the treatments for the 2004-2005 season in Figure 3.2 had data assimilation simulations that were statistically different from open-loop simulations for both grain yield and canopy weight. In the 2003-2004 season treatment FSH, FTL, and FDL did not have data assimilation simulations that were statistically different from open-loop simulations for grain yield and canopy weight.

Figure 3.2, Table 3.4 and Table 3.5 show that the data assimilation simulations and the open-loop simulations trend in the same direction across treatments. The model output of several of the treatment simulations in Figure 3.2 indicates that in the model the open-loop and data assimilation schemes are being influenced more strongly by an estimate or parameter other than soil moisture. The treatment simulations where this is most relevant are when the model output is significantly further from measured.

In Figure 3.3 unrealistic soil moisture behavior as a consequence of the data assimilation can be seen, for example, from day 80 to 100 in the 2003-2004 season. Once a measurement occurs the model will adjust itself quickly and then the soil moisture will continue to decrease as it had before. These sharp changes occur specifically when the Kalman gain weights the measurement more strongly than the model and also when the model estimate and the general trend of the estimate is in large disagreement with the measurement.

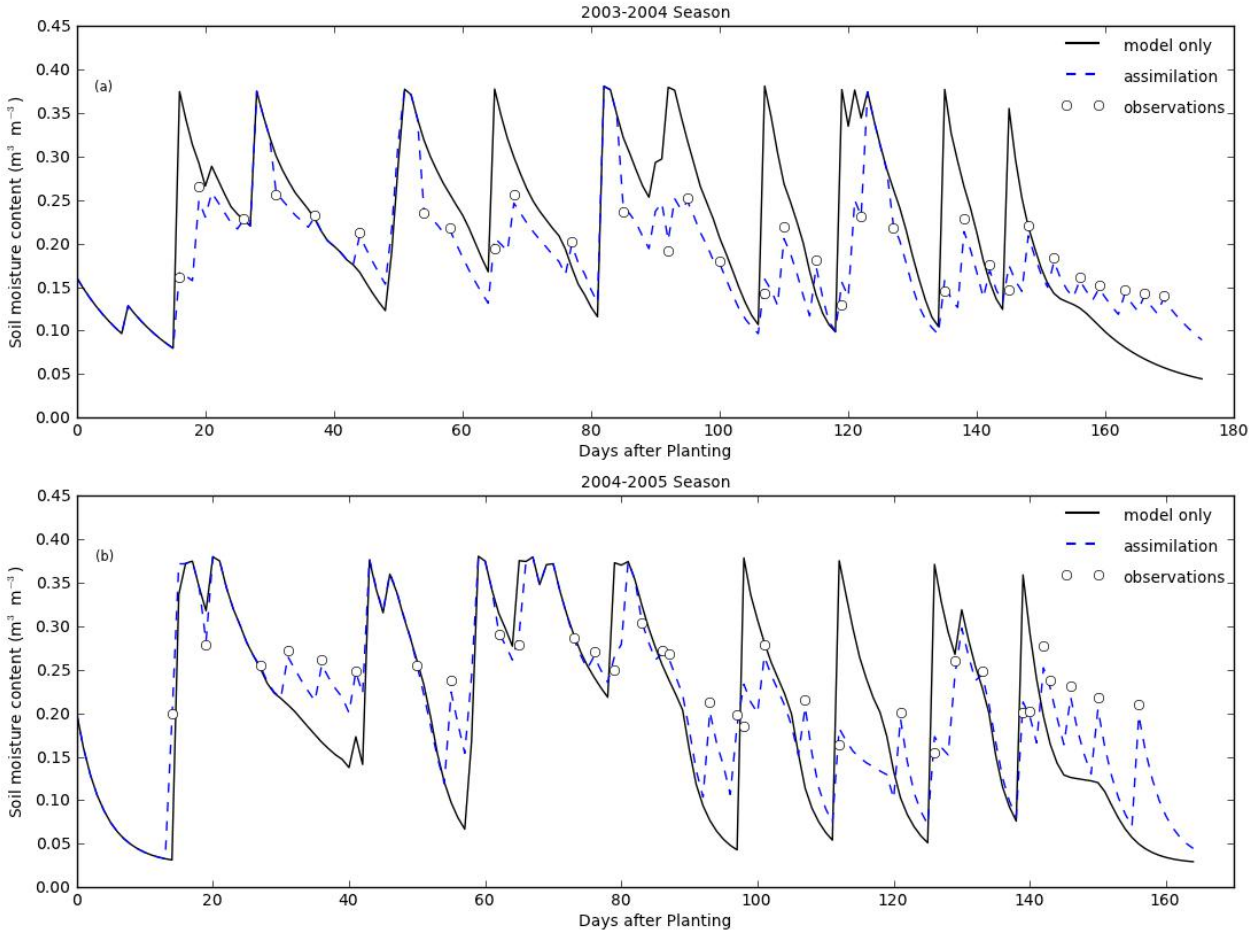

Figure 3.3: Seasonal measurements and soil moisture levels in the 0-3 cm layer where two layers are in the top 30 cm. for the open-loop and data assimilation scheme for each season, from the assimilation of soil moisture into the 0-3 cm layer

The large variations in the soil moisture over time in the top layer of the soil profile layer, seen in Figure 3.3, are due to evapotranspiration, ET, as well as the drainage into lowers layers that occurs, especially for this well-drained soil. Since this layer is the crucial boundary layer between soil processes such as drainage and ET, it affects key modeling processes, such as crop and root growth. The crop growth depends on the energy balances that take place at this surface and the soil moisture in this layer will influence these processes making it a critical estimate.

Figure 3.3 indicates that the soil moisture for a layer of 0-3 cm tends to be overestimated by the open-loop simulations in the 2003-2004 season. The soil moisture of the 0-3 cm layer

is generally underestimated by the open-loop simulations for the 2004-2005, however. The soil moisture measurements were likely smaller then the actual soil moisture for the layer from 15-30 cm because the soil moisture measurement represented 0-30 cm. The measurement was also presumably larger then the actual soil moisture content in the 0-3 cm but yet the model output experienced improvements for both seasons.

Tracking the daily water balance of the entire soil profile on days in which assimilations was done in the top layer for the soil layer structure consisting of three top 30 cm layers showed that the average difference between the depth of water before assimilation and after assimilation wasn't largely different. The actual average differences between daily water balance before assimilation and after for this scheme were 0.127 mm and 0.016 mm for the 2003-2004 and 2004-2005 season respectively. Considering assimilation into a thicker layer such as the second layer with a soil layer configuration of two 30 cm layers for the 2003-2004 the water depth has an average difference of 0.653 mm. For the same season assimilating into the second of three top 30 cm layers results in a difference between daily water depth before assimilation and after assimilation of 0.338 mm.

Soil nitrate (NO_3) levels are important to and vary based on crop growth [Boote et al. (1998)]. If the crop growth is varied due to changes in soil moisture, the levels and need for nitrogen are likely going to be influenced as well. For both seasons the level of mineralized N and denitrified N had little to no change when assimilating into the top layer. When assimilating into the second of three top layers in the 2003-2004 year the mineralization of N was decreased by 3 kg ha^{-1} and denitrified N was decreased by 1 kg ha^{-1}. This was the largest effect that was present throughout any of the assimilation schemes for both seasons. Most schemes and treatments did not experience much change.

Daily levels of soil nitrate, NO_3, throughout the season occur at similar levels across the treatments when the data assimilation is compared to open-loop simulations for the data assimilation schemes that perform well, specifically the top layer assimilation schemes. However the daily nitrogen levels of the data assimilation scheme simulations that assimilated into the second layer and below were higher for every layer in the top 30 cm. These data assimilation schemes were also the simulations that had model output that was the same or worse as the

open-loop simulations. These trends were most notable for 2003-2004 season. Thorp et al. (2010a) reports that high levels of nitrogen stress occurred in the 2004-2005 season. This resulted in the model output for grain yield for this season to be driven by nitrogen sensitive grain parameters. The performance of the data assimilation, based on the average grain yield and canopy weight along with the standard deviations and percentage of improvement, during this season could be attributed to the levels of nitrogen stress.

3.5 Conclusions

- A Kalman filter data assimilation scheme assimilating *in situ* soil moisture measurements into layers of 0-5 cm was more likely to improve the model output of grain yield and canopy weight, by improving the average model output when compared to the model open-loop results.

- Assimilation proved more effective then the model open-loop simulations for a soil configuration of two top 30 cm layers when assimilation was done into the layer from 0-3 cm, by improving the average model output, lowering the standard deviation and having a higher percentage of improvement when compared to the model open-loop results.

- Monte Carlo methods show a moderate ability to address soil parameter uncertainty for assimilation of soil moisture into a small top layer.

- A soil layer structure with layers from 0-3 cm and 3-30 cm resulted in slightly higher standard deviations as well as average model output that was moderately further away from measured, for grain yield and canopy weight for both open-loop and data assimilation configurations.

- Assimilation improvements were more likely to occur when assimilation was done in layers, specifically for layers of 0-5 cm when the soil layer structure consists of three top 30 cm layers and 0-3 cm for a soil layer structure consisting of two top 30 cm layers.

- Assimilation using soil moisture measurements averaged across the top 30 cm are not as likely to overcome improper soil moisture estimates due to poor model inputs of soil

parameters, indicating there is another variable governing model output.

CHAPTER 4. Conclusions

4.1 Conclusion

Data assimilation simulations showed the most improvement of model output towards measured when assimilating soil moisture into layers from 0-5cm and 0-3cm of the soil profile. When data assimilation did not perform better than the open-loop, the model output of the open-loop simulations were generally not much different from those of the data assimilation simulations.

Assimilation into the top two layers of three layers in the top 30cm showed comparable performance to assimilation done into each of these layers individually. Assimilation of soil moisture into the layer from 15-30cm showed slight improvement towards measured as well. Alternatively, assimilating the soil moisture in 0-30cm didn't improve model output likely due to the lack of accurate water distribution because of the average soil moisture value represented 0-30cm.

Overestimation in the top 15cm likely had more of influence on the lower layers, due to the physical relationship between layers, for the 2003-2004 season. In the 2004-2005 season underestimation of soil moisture by the open-loop model was noticed specifically for the layer from 0-3cm. It is likely that the measurements from 0-30cm underestimated the depth of soil water likely to be in the lower 15cm and overestimated the water that was in the top 5cm. Both seasons showed model output improvement towards measured for the 0-3cm even

The results of the simulations using the Kalman filter data assimilation successfully implemented in the DSSAT-CSM, signifies that the DSSAT-CSM for this case seems to overestimate or underestimate soil moisture predominately in the top layer. The influence that this has on the lower layers and also crop growth can be seen by some of the improvements noted from the

assimilation schemes. The results of the soil moisture assimilation also seem to indicate crop growth's strong relationship to soil water partitioning in the top 30cm.

Partitioning of the soil moisture within the top 30cm likely effects crop root growth, and likewise the crop biomass partitioning. The location of water within the soil for the top layers also determines where the nutrients travel and finally how the crop produces biomass. The partitioning of the biomass within the plant itself may be influenced, strongly under some circumstances, by the partitioning of soil water within the soil. The water distribution is dependent on the physics of the water balance model, that is based around fundamental soil parameters. This is a reason why the uncertainty of the input soil parameters was considered in the simulations.

A high uncertainty in soil parameters results with more frequent soil moisture observations being required for a assimilation scheme to provide accurate soil moisture estimates [Walker (2002)]. Because of the importance of water distribution among layers to crop growth, location of soil moisture observations are also a consideration. The uncertainty of the soil parameters could require observations at multiple layers, specifically the layers where soil moisture most significantly influences crop growth and have a higher amount of variation. Bert et al. (2007) reports that priorities can be established for data collection if the influence of the uncertainty that the measured variables have on the model is known. Which also can lead to more accurate model estimates and results analysis as well as informing of model design. Bert et al. (2007) also notes that little literature exists on the effect of input uncertainties on simulation results, in light of their wide use.

Even though the soil moisture observations used with the Kalman filter assimilation schemes had limitations as far as the location and frequency of the measurements were concerned, the dataset nonetheless was still an invaluable source considering the time and effort that went into its collection and calculation. The results from the Kalman filter highlight challenges commonly face when modeling and also coupling observations with models. The Kalman filter performed as was intended and provided beneficial model output allowing for a useful analysis.

The combination of the direct insertion scheme followed by the application of Kalman filter and Monte Carlo methods overall has been useful to the understanding of the model soil

moisture estimation and the influence of soil parameters within the DSSAT-CSM. This by no means validates this research as encompassing but is just another building block to construct a more detailed view of the ideas, concepts and applications presented throughout the research.

The results leave room for further research and a more in depth review or application of the research already done. The determined frequency of observations is strongly linked to the uncertainty of the soil parameter inputs. The improvement of the soil parameters as inputs for the model would lead to the need for less frequent observations from fewer layers, and vice versa. The ability to quantify the error or deviation of results from the simulations based on the uncertainty of these input parameters should be considered, as well.

Data assimilation could still address some of these errors that are still present, drawing on the results presented to provide further insight into the use of the DSSAT-CSM and data assimilation.

4.2 Future Direction

Based on the previous discussions several aspects of the research provide possible suggestions for a more in depth study and examination for specifically addressing soil moisture estimation. Since, the soil moisture observations that were used, were not evenly distributed throughout the season, it may be beneficial to consider what frequency or schedule of observations are most appropriate. The frequency of the soil moisture observations are assimilated has more of an impact on the data assimilation effectiveness especially with uncertain soil parameters. Soil moisture behavior varies based on the water content of the soil so considering at what level of saturation is it important to have moisture observations. This updating frequency as indicated by Walker (2002) is also dependent upon the errors present by the model as well as those expected from the measurements. An analysis to determine at what interval and depths the soil moisture observations are effective based on soil parameter uncertainty.

It would be best to look at other alternative or more intensive soil parameter determination is required if uncertainty of the soil parameters are to be decreased. Research has shown that site specific PTFs are a better choice for determining the van Genuchten soil parameters than ROSETTA [Rubio (2008)]. Looking at methods to develop equations such as these could

address the soil parameter uncertainty.

A thorough sensitivity analysis may be necessary for the soil parameters when using the CERES-Wheat crop model, for soil moisture as well as nitrogen sensitive crop growth parameters. In doing so the soil parameters can be optimized to help adequately partition soil moisture and also nitrogen for future simulations. Since the weather dramatically effects the variability of the soil moisture in the top 5cm, another aspect of consideration would be the difference of the weather conditions for each season for this experiment. From this comparison it may be able to explain situations in which the DSSAT-CSM overestimates soil moisture. Since the overestimation could be linked to evapotranspiration and root uptake processes as well.

The wide implementation of crop models is dependent on their performance success and effectiveness. The DSSAT-CSM model has proven its ability to accurately and perform well under many different circumstances and geophysical conditions. Since the use of the DSSAT-CSM for research and precision agriculture has led to an increased importance in its ability to inform decisions based on the model predictions. These decision are based on economic and educational returns but also on the basic need for food. With this in mind, continued improvement of the DSSAT-CSM should still be a focus to be able to utilize new knowledge and data. As technology improves the amount of information and its detail will increase as well. The information is only useful if has well designed applications and well informed implementation. It is necessary to further understand the relationship that exists when coupling data, specifically in real-time, into the DSSAT-CSM. The effects and benefits of this type of research can lead to improvements of the model design and predictions and in turn help improve all types of benefits.

APPENDIX A. Fortran Code

A.1 WATBAL_ASSIM.FOR

SW_ASSIM being called during the SEASINIT stage in the file for calculating water balance components of the DSSAT-CSM.

```
      IF (Assim_sw) THEN
        CALL SW_ASSIM(CONTROL, ISWITCH, SOILPROP, assim_cont_mod,
     &     assim_data, SW)
        CALL OPASSIM(CONTROL, ISWITCH, assim_data, assim_cont_mod)
      ENDIF
```

SW_ASSIM being called during the INTEGR stage in the file for calculating water balance components of the DSSAT-CSM.

```
!       Perform assimilation of soil water measurements in all layers
        IF (Assim_sw) THEN
            CALL SW_ASSIM(CONTROL, ISWITCH, SOILPROP, assim_cont_mod,
     &          assim_data, SW)
        ENDIF
```

A.2 SW_ASSIM.FOR

The SW_ASSIM file is where the soil moisture estimates are updated if a data assimilation scheme is being used.

```
C=====================================================================
C  SW_ASSIM, Subroutine, Candance M. Batts and Derek G. Groenendyk
```

```
C  Calculates water balance assimilation.
C-----------------------------------------------------------------------
C  REVISION       HISTORY
C  07/14/2008 CMB Written
C  06/14/2010 DGG Updated KFProp Calls and general code
C-----------------------------------------------------------------------
C  Called by: WATBAL_ASSIM module
C  Calls:     YR_DOY
C             KFProp  (File KFProp.FOR)
C=======================================================================
SUBROUTINE SW_ASSIM(CONTROL, ISWITCH, SOILPROP, assim_cont_mod,
     &    assim_data, SW)

        USE ModuleDefs
        USE AssimDefs
        IMPLICIT NONE   ! in Fortran 90, compiler flags more mistakes
        SAVE

        TYPE (ControlType)  CONTROL
        TYPE (AssimContType) assim_cont_mod
        TYPE (AssimDataType) assim_data
        TYPE (SwitchType) ISWITCH
        TYPE (SoilType)    , INTENT(IN) :: SOILPROP

        CHARACTER*6  ERRKEY
        INTEGER model_doy, model_year, SL, Assim_obs_per, DYNAMIC
        INTEGER Start_layer, Assim_type, Assim_layers, bottomLayer
        INTEGER SoilLayer, obs_doy
        INTEGER obs_period, AL, NLAYR !period in between observation assims
```

```
      INTEGER yr, doy, indx, nlen ! phase of kalman filter
      LOGICAL last_assim
      REAL, DIMENSION(NL) :: SW
      REAL swold, sw1delta, stemp1, theta, swobs, swpred

      nlen          = assim_cont_mod % Assim_obs
      Assim_type    = assim_cont_mod % Assim_type
      Assim_layers  = assim_cont_mod % Assim_layers
      Assim_obs_per = assim_cont_mod % Assim_obs_per
      Start_layer   = assim_cont_mod % Start_layer

      NLAYR = SOILPROP % NLAYR
      DYNAMIC = CONTROL % DYNAMIC

      bottomLayer = Assim_layers + Start_layer - 1
      AL = 1

      IF (DYNAMIC .EQ. SEASINIT) THEN
          indx = 1
          last_assim = .FALSE.
!         To initialize the Kalman filter, specifically paposteriori.
          IF (Assim_type .EQ. Assim_kalman) THEN
              assim_data % kf_mode = 0
              DO SoilLayer = Start_layer, bottomLayer
                  AL = SoilLayer - Start_layer + 1
                  CALL KFProp (CONTROL, ISWITCH, assim_cont_mod,
     &                 assim_data, theta, swold, nlen, AL)
              ENDDO
          ENDIF
```

```
        ELSEIF (DYNAMIC .EQ. INTEGR) THEN
!            called during integrate (CONTROL % DYNAMIC)
!            for continuous (daily) data

            CALL YR_DOY(CONTROL % YRDOY,YR,model_doy)

            IF (indx .GT. nlen) THEN
             last_assim = .TRUE.
            ELSE
             obs_doy = assim_data % obs_data(indx,1)
            ENDIF

             IF (obs_doy .EQ. model_doy) THEN
                !update
                ERRKEY = "Layers"
                IF (bottomLayer .GT. 10) THEN
                    print *,bottomLayer
                    CALL ERROR (ERRKEY,1,"FILEX",55)
                ENDIF
                assim_data % kf_mode = 2
                DO SoilLayer = Start_layer, bottomLayer
                  !print *,SoilLayer
                  !pause
                  SL = SoilLayer + 1
                  !also could use Assim_layers
                  IF (assim_cont_mod % Assim_obs_layers .LT.
     &               NLAYR) THEN
                    SL = SL - Start_layer + 1
```

```
                ENDIF

!               simple direct insertion
                IF (Assim_type .EQ. Assim_direct) THEN
                  SW(SoilLayer) = assim_data % obs_data(indx,SL)

!               call kalman subroutine
                ELSEIF (Assim_type .EQ. Assim_kalman) THEN
                  !print *,SL-1
                  !pause
                  theta = assim_data % obs_data(indx,SL)
                  swold = SW(SoilLayer)
                  CALL KFProp(CONTROL, ISWITCH, assim_cont_mod,
     &                 assim_data, theta, swold, nlen, SL-1)
                  SW(SoilLayer) = theta
                ENDIF
              ENDDO
              indx = indx + 1

            ELSE
                IF (Assim_type .EQ. Assim_kalman) THEN
                 assim_data % kf_mode = 1
                 DO SoilLayer = Start_layer, bottomLayer
                     AL = SoilLayer - Start_layer + 1
                     !AL = AssimLayer
                     !print *,AL
                     CALL KFProp (CONTROL, ISWITCH, assim_cont_mod,
     &                    assim_data, theta, swold, nlen, AL)
                  ENDDO
```

```
                ENDIF
            END IF
          ENDIF
        RETURN
        END SUBROUTINE SW_ASSIM
```

A.3 KFPROP.FOR

In KFPROP the Kalman filter calculations are made based on the stage of the DSSAT-CSM and the presence of a measurement.

```
C=====================================================================
C  KFProp, Subroutine, Derek G. Groenendyk
C  Calculates the Kalman filter variables
C---------------------------------------------------------------------
C  Soil Processes subroutine.  Calls the following modules:
C---------------------------------------------------------------------
C  REVISION HISTORY
!  06/14/2010 DGG Written
C=====================================================================
       SUBROUTINE KFProp(CONTROL, ISWITCH, assim_cont_mod, assim_data,
      &    theta, swold, nlen, AL)

         USE ModuleDefs
         USE AssimDefs

         IMPLICIT NONE
         SAVE

         TYPE (AssimContType) assim_cont_mod
```

```
        TYPE (ControlType) CONTROL
        TYPE (SwitchType) ISWITCH
        TYPE (AssimDataType) assim_data

        CHARACTER FILEASSIM*(6)
        INTEGER kf_mode, nlen, AL, Assim_layers, FILENUM
        INTEGER bottomLayer, Start_layer, DYNAMIC, status
!       phase of kalman filter
        REAL swold, theta, swobs, swpred, X_0
        REAL xaposteriori_0, paposteriori_0, A, H, Q, R
        CHARACTER*6, PARAMETER :: ERRKEY = 'ALLCTE'

        REAL, DIMENSION(assim_cont_mod % Assim_layers) :: Z
        REAL, DIMENSION(assim_cont_mod % Assim_layers) :: xapriori
        REAL, DIMENSION(assim_cont_mod % Assim_layers) :: xaposteriori
        REAL, DIMENSION(assim_cont_mod % Assim_layers) :: residual
        REAL, DIMENSION(assim_cont_mod % Assim_layers) :: papriori
        REAL, DIMENSION(assim_cont_mod % Assim_layers) :: paposteriori
        REAL, DIMENSION(assim_cont_mod % Assim_layers) :: K

        DYNAMIC = CONTROL % DYNAMIC
        Assim_layers = assim_cont_mod % Assim_layers
        Start_layer  = assim_cont_mod % Start_layer
        Q = assim_cont_mod % Filter_Q  ! process/model noise covariance
        R = assim_cont_mod % Filter_R  ! measurement noise covariance
        kf_mode = assim_data % kf_mode

        FILEASSIM = "KFProp"
        bottomLayer = Start_layer + Assim_layers - 1
```

```
   swpred = swold ! model's projected day value
   swobs = theta  ! obs value
   !Define the system
   A = 1
   H = 1

   IF (DYNAMIC .EQ. SEASINIT .AND. AL .EQ. 1) THEN
     ALLOCATE (assim_data % Z(Assim_layers), STAT=status)
     IF (status .NE. 0) CALL ERROR (ERRKEY,2,FILEASSIM,FILENUM)
     ALLOCATE (assim_data % xapriori(Assim_layers), STAT=status)
     IF (status .NE. 0) CALL ERROR (ERRKEY,3,FILEASSIM,FILENUM)
     ALLOCATE (assim_data % xaposteriori(Assim_layers),
&     STAT=status)
     IF (status .NE. 0) CALL ERROR (ERRKEY,4,FILEASSIM,FILENUM)
     ALLOCATE (assim_data % residual(Assim_layers), STAT=status)
     IF (status .NE. 0) CALL ERROR (ERRKEY,5,FILEASSIM,FILENUM)
     ALLOCATE (assim_data % papriori(Assim_layers), STAT=status)
     IF (status .NE. 0) CALL ERROR (ERRKEY,6,FILEASSIM,FILENUM)
     ALLOCATE (assim_data % paposteriori(Assim_layers),
&     STAT=status)
     IF (status .NE. 0) CALL ERROR (ERRKEY,7,FILEASSIM,FILENUM)
     ALLOCATE (assim_data % K(Assim_layers), STAT=status)
     IF (status .NE. 0) CALL ERROR (ERRKEY,8,FILEASSIM,FILENUM)
   ENDIF

   IF (DYNAMIC .EQ. INTEGR .OR. DYNAMIC .EQ. SEASINIT) THEN
       If (kf_mode .NE. 0) THEN
           xapriori     = assim_data %  xapriori
           xaposteriori = assim_data %  xaposteriori
```

```
    residual     = assim_data %  residual
    papriori     = assim_data %  papriori
    paposteriori = assim_data %  paposteriori
    K            = assim_data %  K
ENDIF

IF (kf_mode .EQ. 2) THEN !obs exists
    ! Assimilate observation theta1 into SW(1). Using swold,
    ! model predicted reading.
    ! Initial guesses for state and a posteriori covariance.
    xaposteriori_0 = swpred   !  model prediction
    Z(AL) = swobs             !  measurement observation
    paposteriori_0 = paposteriori(AL)

    ! Predictor (Time Update) equations
    xapriori(AL) = A*xaposteriori_0 ! model prediction
    residual(AL) = Z(AL)-H*xapriori(AL) !does this make sense?
    papriori(AL) = A*A*paposteriori_0+Q

    ! Corrector (Measurement Update) equations
    K(AL) = H*papriori(AL)/(H*H*papriori(AL)+ R)!Kalman Gain
    paposteriori(AL) = papriori(AL)*(1-H*K(AL)) ! P_k
    xaposteriori(AL) = xapriori(AL)+K(AL)*residual(AL)!New Pred.

ELSEIF (kf_mode .EQ. 1) THEN !no obs exists
  paposteriori_0 = paposteriori(AL)
  papriori(AL) = A*A*paposteriori_0+Q
  paposteriori(AL) = papriori(AL)
  K(AL) = 0.0
```

```
      ELSEIF (kf_mode .EQ. 0) THEN !initialization run...
        paposteriori_0 = R
        papriori(AL) = A*A*paposteriori_0+Q
        paposteriori(AL) = papriori(AL)
        K(AL) = 0.0
      ENDIF

    theta = xaposteriori(AL)
    assim_data % xapriori     = xapriori
    assim_data % xaposteriori = xaposteriori
    assim_data % residual     = residual
    assim_data % papriori     = papriori
    assim_data % paposteriori = paposteriori
    assim_data % K            = K

ENDIF
RETURN
END SUBROUTINE KFProp
```

APPENDIX B. Extra Figures

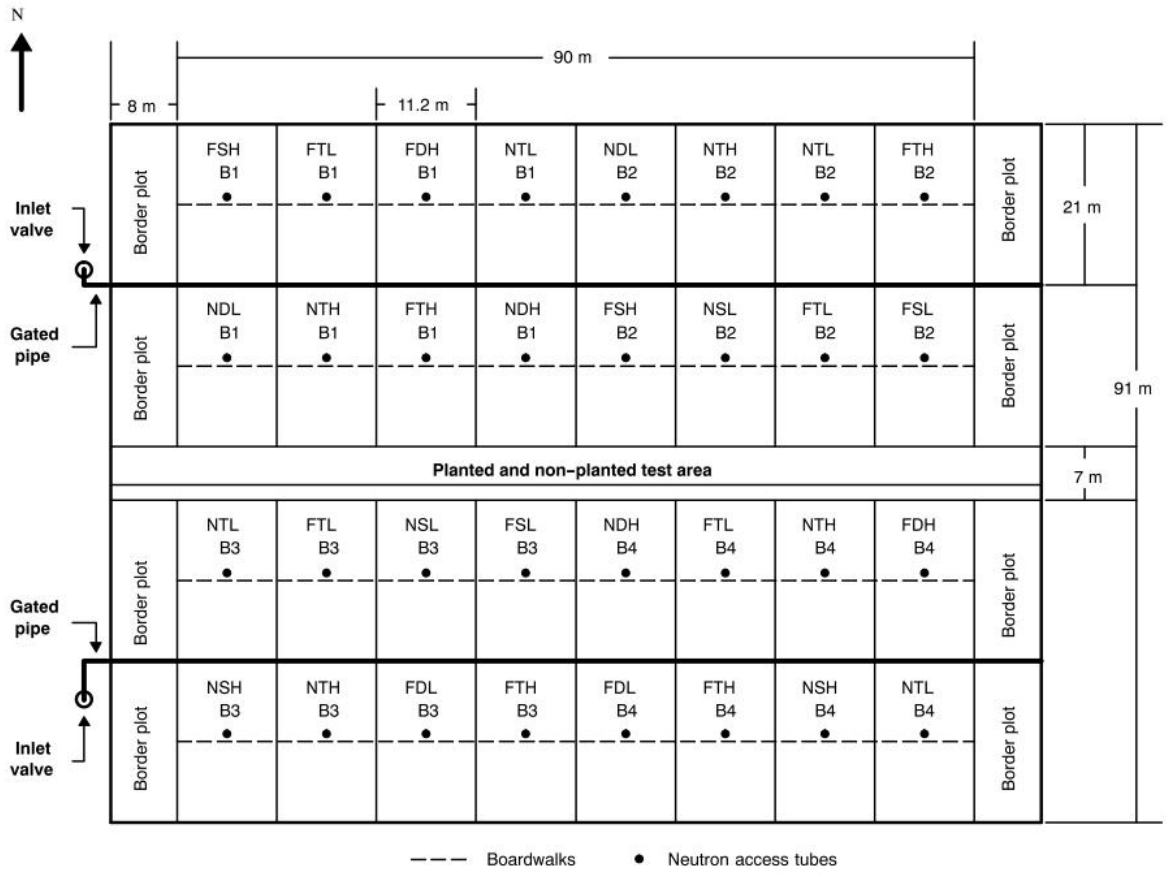

Figure B.1: Hunsaker et al. (2007b) field layout with each plot labeled with its respective treatment. Detailed field distances and locations are also shown. The same layout was used for both seasons.

BIBLIOGRAPHY

Alavi, N., Warland, J. S., and Berg, A. A. (2009). Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications. *Earth*, pages 429–448.

Anderson, J. L. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, 129(12):2884.

Balland, V., Pollacco, J., and Arp, P. (2008). Modeling soil hydraulic properties for a wide range of soil conditions. *Ecological Modelling*, 219(3-4):300–316.

Barth, A., Azc, A. A., Joassin, P., and Troupin, C. (2008). Introduction to Optimal Interpolation and Variational Analysis.

Batchelor, W. (2002). Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy*, 18(1-2):141–158.

Bert, F., Laciana, C., Podesta, G., Satorre, E., and Menendez, a. (2007). Sensitivity of CERES-Maize simulated yields to uncertainty in soil properties and daily solar radiation. *Agricultural Systems*, 94(2):141–150.

Bierman, G. (1979). *Stochastic models, estimation, and control*. Academic Press, New York, volume 1 edition.

Boote, K. J., Jones, J. W., Hoogenboom, G., and Pickering, N. B. (1998). *The CROPGRO model for grain legumes*. Kluwer Academic Publishers, Dordrecht, volume 7 edition.

Bouttier, F. and Courtier, P. (2002). Data assimilation concepts and methods. Technical Report March 1999, ECWMF.

Braga, R. P. and Jones, J. W. (2004). USING OPTIMIZATION TO ESTIMATE SOIL INPUTS OF CROP MODELS FOR USE IN SITE-SPECIFIC MANAGEMENT. *Transactions Of The ASABE*, 47(5):1821–1831.

Burgers, G., Jan Van Leeuwen, P., and Evensen, G. (1998). Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, 126(6):1719.

Casanova, J. J., Judge, J., and Jones, J. W. (2005). Calibration of the CERES-Maize Model for Linkage with a Microwave Remote Sensing Model. *ASAE Mtg. Presentation*, 053027:1–12.

Chen, Z., Li, S., Ren, J., Gong, P., and Zhang, M. (2008). *Monitoring and Management of Agriculture with Remote Sensing*, pages 397–421. Springer Netherlands.

Chirico, G., Medina, H., and Romano, N. (2007). Uncertainty in predicting soil hydraulic properties at the hillslope scale with indirect methods. *Journal of Hydrology*, 334(3-4):405–422.

Courtier, P. (1997). Variational Methods. *Journal of the Meteorological Society of Japan*, 75(1B):211–218.

Dewit, A. J. W. and Vandiepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1-2):38–56.

Drecourt, J. P. and Madsen, H. (2002). Uncertainty estimation in groundwater modelling using Kalman filtering. In *Proceedings of the 4th International Conference on Calibration and Reliability in Groundwater Modelling*, volume 46, pages 306–309, Prague, Czech Republic.

Entekhabi, D., Nakamura, H., and Njoku, E. (1994). Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2):438–448.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143–10162.

Eyre, J. R. (1997). Variontional Assimilation of Remotely-Sensed Observations of Atmosphere. *Journal of the Meteorological Society of Japan1*, 75(1B):331–338.

Galantowicz, J., Entekhabi, D., and Njoku, E. (1999). Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4):1860–1870.

Heathman, G., Starks, P., Ahuja, L., and Jackson, T. (2003). Assimilation of surface soil moisture to estimate profile soil water content. *Journal of Hydrology*, 279(1-4):1–17.

Heinzel, V., Waske, B., Braun, M., and Menz, G. (2007). Remote sensing data assimilation for regional crop growth modelling in the region of Bonn (Germany). *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 3647–3650.

Holm, E. V. (2003). Lecture Notes on Assimilation Algorithms. Technical report, European Center for Medium-Range Weather Forecasts, Reading, UK.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., and Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, 34(12):3405–3420.

Huang, C., Li, X., Lu, L., and Gu, J. (2008). Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. *Remote Sensing of Environment*, 112:888 – 900.

Huang, G. (2004). Modeling soil water regime and corn yields considering climatic uncertainty. *Plant and Soil*, 259(1/2):221–229.

Hunsaker, D. J., Fitzgerald, G. J., French, A. N., Clarke, T. R., and Ottman, M. J. (2007a). Wheat Irrigation Management Using Multispectral Crop Coedffients: I. Crop Evapotranspiraton Prediction. *Transactions Of The ASABE*, 50(1):2017–2033.

Hunsaker, D. J., Fitzgerald, G. J., French, A. N., Clarke, T. R., and Ottman, M. J. (2007b). Wheat Irrigation Management using Multispectral Crop coefficients: II. Irrigation Schedul-

ing Performance, Grain Yield, and Water Use Efficiency. *Transactions Of The ASABE*, 50(6):2035–2050.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C. (1997). Unified Notation for Data Assimilation: Operational, Sequential and Variational.

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., and Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*.

Koo, J., Bostick, W. M., Naab, J. B., Jones, J. W., Graham, W. D., and Gijsman, A. J. (2007). ESTIMATING SOIL CARBON IN AGRICULTURAL SYSTEMS USING ENSEMBLE KALMAN FILTER AND DSSAT-CENTURY. *Transactions Of The ASABE*, 50(5):1851–1865.

Kumar, P. and Kaleita, A. L. (2003). Assimilation of near-surface temperature using extended Kalman filter. *Advances in Water Resources*, 26(1):79–93.

Liang, S. and Qin, J. (2008). *Data Assimilation Methods for Land Surface Variable Estimation*, chapter 12, pages 313–339. Springer Netherlands.

Liu, H., Yang, J., Tan, C., Drury, C., Reynolds, W., Zhang, T., Bai, Y., Jin, J., He, P., and Hoogenboom, G. (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*, 98:1105–1111.

Ma, L., Ahuja, L. R., and Malone, R. W. (2007). Systems Modeling For Soil and Water Research and Management; current status and needs for the 21st Century. *Transactions Of The ASABE*, 50(5):1705–1713.

Ma, L., Hoogenboom, G., Ahuja, L. R., and Nielsen, D. C. (2005). Development and Evaluation of the RZWQM-CROPGRO Hybrid Model for Soybean Production. *Agronomy Journal*, 97:1172–1182.

Ma, L., Hoogenboom, G., Ahuja, L. R., and Saseendran, S. A. (2006). Evaluation of the RZWQM-CERES-Maize hybrid model for maize production. *Agricultural Systems*, 87:274–295.

Ma, L., Hoogenboom, G., Saseendran, S. a., Bartling, P. N. S., Ahuja, L. R., and Green, T. R. (2009). Effects of Estimating Soil Hydraulic Properties and Root Growth Factor on Soil Water Balance and Crop Production. *Agronomy Journal*, 101(3):572–583.

Maas, S. (1988). Use of remotely-sensed information in agricultural crop growth models.

Moulin, A., Bondeau, R., and Delecolle, S. (1998). Combining agricultural crop models and satellite observations: from field to regional scales. *International Journal of Remote Sensing*, 19(6):1021–1036.

Novak, V., Hurtalova, T., and Matejka, F. (2005). Predicting the effects of soil water content and soil water potential on transpiration of maize. *Agricultural Water Management*, 76(3):211–223.

Paniconi, C., Marrocu, M., Putti, M., and Verbunt, M. (2003). Newtonian nudging for a Richards equation-based distributed hydrological model. *Advances in Water Resources*, 26(2):161–178.

Pathmathevan, M., Koike, T., and Li, X. (2003). A New Satellite-Based Data Assimilation Algorithm to Determine Spatial and Temporal Variations of Soil Moisture and Temperature Profiles. *Journal of the Meteorological Society of Japan*, 81(5):1111–1135.

Paz, J. O., Batchelor, W. D., Babcock, B. A., Colvin, T. S., Logsdon, S. D., Kaspar, T. C., and Karlen, D. L. (1999). Model-based technique to determine variable rate nitrogen for corn. *Agricultural Systems*, 61(1):69–75.

Paz, J. O., Batchelor, W. D., Colvin, T. S., Logsdon, S. D., Kaspar, T. C., and Karlen, D. L. (1998). Analysis of Water Stress Effects Causing Spatial Yield Variability in Soybeans. *Transactions Of The ASABE*, 41(5):1527–1534.

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *PoLAR*, 31:1411–1418.

Reichle, R. H., Crow, W. T., and Keppenne, C. L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resour. Res.*, page 101029/.

Reichle, R. H., Walker, J. P., Koster, R. D., and Houser, P. R. (2002). Extended versus Ensemble Kalman Filtering for Land Data Assimilation. *Journal of Hydrometeorology*, 3(6):728.

Rezzoug, W., Gabrielle, B., Suleiman, A., and Benabdeli, K. (2008). Application and evaluation of the DSSAT-wheat in the Tiaret region of Algeria. *Journal of Agricultural Research*, 3(4):284–296.

Rihan, F. A. and Collier, C. G. (2003). Four-Dimensional Data Assimilation and Numerical Weather Prediction. Technical Report July, Centre for Environmental System Research, Unversity of Salford, Manchester, England.

Romano, N. and Santini, A. (1997). Effectiveness of using pedo-transfer functions to quantify the spatial variability of soil water retention characteristics. *Journal of Hydrology*, 202(1-4):137–157.

Rubio, C. M. (2008). Applicability of Site-Specific Pedotransfer Functions and Rosetta Model for the Estimation of Dynamic Soil Hydraulic Properties under Different Vegetation Covers. *Journal of Soil and Sediments*, 8(2):137 – 145.

Sabater, J. M. n., Jarlan, L., Calvet, J.-C., Bouyssel, F., and De Rosnay, P. (2007). From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques. *Journal of Hydrometeorology*, 8(2):194.

Sadler, E., Gerwig, B. K., Evans, D. E., Busscher, W. J., and Bauer, P. J. (2000). Site-specific modeling of corn yield in the SE coastal plain. *Agricultural Systems*, 64(3):189–207.

Sarkar, R. and Kar, S. (2008). Sequence Analysis of DSSAT to Select Optimum Strategy of Crop Residue and Nitrogen for Sustainable Rice-Wheat Rotation. *Agronomy Journal*, pages 87–97.

Sau, F., Boote, K. J., Bostick, W. M., Jones, J. W., and Ine, M. (2004). Testing and Improving Evapotranspiration and Soil Water Balance of the DSSAT Crop Models. *American Soceity of Agronomy*, 1257:1243–1257.

Schaap, M., Leij, F., and Vangenuchten, M. (2001). ROSETTA: a Computer Program for Estimating Soil Hydraulic Parameters With Hierarchical Pedotransfer Functions. *Journal of Hydrology*, 251(3-4):163–176.

Soil Conservation Service (SCS) (1972). National Engineering Handbook, Hydrology Section 4, Chapters 4-10.

Stastná, M. and Zalud, Z. (1999). Sensitivity analysis of soil hydrologic parameters for two crop growth simulation models. *Soil and Tillage Research*, 50(3-4):305–318.

Stauffer, D. R. and Seaman, N. L. (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Monthly Weather Review*, 118(6):1250–77.

Thorp, K. R., Hunsaker, D. J., and French, A. N. (2010a). Assimilating Leaf Area Index Estimates from Remote Sensing into the Simulations of a Cropping Systems Model. *Transactions Of The ASABE*, 53(1):251–262.

Thorp, K. R., Hunsaker, D. J., French, A. N., White, J. W., and Clarke, T. R. (2010b). Evaluation of the CSM-CROPSIM-CERES-Wheat Model as a Tool for Crop Water Management. *Transactions Of The ASABE*, 53(2006):1–17.

Tsuji, G. Y., Hoogenboom, G., and Thornton, P. K., editors (1998a). *Soil water balance and plant water stress*, pages 41–54. Kluwer Academic Publishers, Great Britain, 7 edition.

Tsuji, G. Y., Hoogenboom, G., and Thornton, P. K. (1998b). *Understanding Options for Agricultural Production*. Kluwer Academic Publishers, Dordrecht, volume 7 edition.

Walker, J. P. (2002). Three-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: Simplified Kalman filter covariance forecasting and field application. *Water Resources Research*, 38(12):101029/.

Walker, J. P. and Houser, P. R. (2001). A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *Journal of Geophysical Research*, 106(D11):11761–11774.

Walker, J. P., Willgoose, G. R., and Kalma, J. D. (2001). One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms. *Advances in Water Resources*, 24(6):631–50.

Welch, G. and Bishop, G. (2001). An Introduction to the Kalman Filter.

Williams, J. R., Jones, C. A., and Dyke, P. T. (1984). A Modeling Approach to Determining the Relationship Between Erosion and Soil Productivity. *Transactions Of The ASABE*, 27(1):129–144.

Wosten, J., Pachepsky, Y., and Rawls, W. (2001). Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*, 251(3-4):123–150.

Zhang, X. and Van Geel, P. J. (2007). Development of a vertical TDR probe to evaluate the vertical moisture profile in peat columns to assess biological clogging. *Journal of Environmental Engineering and Science*, 6(6):629–642.