

# Soil Moisture Estimation by Microwave Remote Sensing for Assimilation into WATClass

by

Damian Chi-Ho Kwok

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master's of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2007

©Damian Kwok 2007

## **Author's Declaration**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## **Abstract**

This thesis examines the feasibility of assimilating space borne remotely-sensed microwave data into WATClass using the ensemble Kalman filter. WATClass is a meso-scale gridded hydrological model used to track water and energy budgets of watersheds by way of real-time remotely sensed data. By incorporating remotely-sensed soil moisture estimates into the model, the model's soil moisture estimates can be improved, thus increasing the accuracy of the entire model.

Due to the differences in scale between the remotely sensed data and WATClass, and the need of ground calibration for accurate soil moisture estimation from current satellite-borne active microwave remote sensing platforms, the spatial variability of soil moisture must be determined in order to characterise the dependency between the remotely-sensed estimates and the model data and subsequently to assimilate the remotely-sensed data into the model. Two sets of data – 1996-1997 Grand River watershed data and 2002-2003 Roseau River watershed data – are used to determine the spatial variability. The results of this spatial analysis however are found to contain too much error due to the small sample size. It is therefore recommended that a larger set of data with more samples both spatially and temporally be taken.

The proposed algorithm is tested with simulated data in a simulation of WATClass. Using nominal values for the estimated errors and other model parameters, the assimilation of remotely sensed data is found to reduce the absolute RMS error in soil moisture from 0.095 to approximately 0.071. The sensitivities of the improvement in soil moisture estimates by using the proposed algorithm to several different parameters are examined.

## **Acknowledgements**

I would like to express my appreciation for my supervisors, Prof. David Clausi and Prof. Eric Soulis for their guidance and support through my thesis. I would also like to thank Frank Seglenieks for his expertise with WATClass and the Grand River watershed dataset.

Thank you also to the Dr. Alain Pietroniro from the National Water Research Institute for providing the Roseau River watershed dataset and his guidance.

I would also like to acknowledge my family for their support and encouragement during the completion of my thesis.

# Table of Contents

Author’s Declaration.....	ii
Abstract.....	iii
Acknowledgements.....	iv
Table of Contents.....	v
List of Figures.....	vii
List of Tables.....	viii
Chapter 1 Introduction.....	1
Chapter 2 Background.....	3
2.1 Soil Moisture.....	3
2.2 WATClass.....	5
2.3 Soil moisture detection.....	7
2.3.1 Soil moisture estimates by active microwave.....	8
2.3.2 Soil moisture estimates by passive microwave.....	13
2.3.3 Summary of remote sensing soil moisture estimates.....	14
2.4 Soil moisture assimilation into hydrological models.....	15
2.4.1 The Kalman filter.....	16
2.4.2 The extended Kalman filter.....	18
2.4.3 The ensemble Kalman filter.....	19
Chapter 3 Data.....	22
3.1 Grand River watershed data.....	22
3.2 Roseau River data.....	27
Chapter 4 Methodology.....	29
4.1 Objectives.....	29
4.2 Scale and spatial correlation of soil moisture.....	30
4.3 Soil moisture assimilation.....	35
4.4 Summary of proposed algorithm.....	39
Chapter 5 Results.....	43
5.1 Spatial variation of soil moisture content.....	43

5.1.1 Spatial variation of soil moisture content in WATClass.....	44
5.1.2 Spatial variation of soil moisture content in ground data .....	46
5.1.3 Producing soil moisture measurement estimates .....	52
5.2 Ensemble Kalman filter implementation.....	53
5.2.1 Parameter selection .....	53
5.2.2 Simulation .....	55
5.2.3 Ensemble size.....	57
5.2.4 Observation frequency .....	58
5.2.5 Measurement error .....	59
5.2.6 Model error .....	63
Chapter 6 Conclusions and Recommendations .....	68
6.1 Conclusions.....	68
6.2 Recommendations .....	69
References.....	71
Appendix A Glossary/Acronyms .....	74
Appendix B Spatial Statistics.....	77
B.1 Correlation Coefficient.....	77
B.2 Variogram.....	78
B.3 Average distance .....	78
Appendix C Imaging Radar Beam Modes .....	82

## List of Figures

Figure 1 – WATClass soil column .....	6
Figure 2 – Measured dielectric constant for five soils at 1.4GHz. ....	10
Figure 3 – Kalman filter example.....	18
Figure 4 – Site placement in Grand River watershed.....	24
Figure 5 – Measured vs. calculated volumetric soil moisture values for regression using all data.....	25
Figure 6 – Measured vs. calculated volumetric soil moisture values for regression with field dependent constant .....	26
Figure 7 – Measured vs. calculated volumetric soil moisture values for regression using daily sums.....	26
Figure 8 – Roseau sample site and met-station locations.....	27
Figure 9 – Variogram for WATClass estimated volumetric water content of the Grand River watershed during the summer months of 1996 and 1997.....	45
Figure 10 – Variogram for Grand River watershed ground-sampled soil moisture content .....	47
Figure 11 – Correlation of soil water content between fields as a function of distance in the Grand River dataset.....	48
Figure 12 – Standard error for correlation of soil water content between fields in the Grand River dataset (N = 13) .....	49
Figure 13 – Variogram for Roseau River watershed ground-sample soil water content.....	50
Figure 14 – Correlation of soil water content between fields as a function of distance in Roseau River dataset .....	51
Figure 15 – Standard error for correlation of soil water content between fields in the Roseau River dataset (N = 7).....	51
Figure 16 – Soil moisture value and estimates for nominal values.....	56
Figure 17 – Average distance from a square .....	80

## List of Tables

Table 1 – Orbital SAR systems launched since 1994.....	12
Table 2 – Grand River watershed field data.....	23
Table 3 – Roseau River dataset field data .....	28
Table 4 – Soil moisture estimate standard deviation normalised by mean for Roseau River dataset ..	52
Table 5 – Simulation parameters.....	54
Table 6 – Nominal variable values for initial simulation .....	55
Table 7 – Error in soil moisture value as a function of several ensemble sizes .....	58
Table 8 – RMS error in soil moisture value as a function of time between measurements .....	59
Table 9 – Error in soil moisture value as a function of estimated measurement error level .....	61
Table 10 – Error in soil moisture as a function of error in measurement error level .....	62
Table 11 – Error in soil moisture as a function of bias in the observation (%of true value).....	62
Table 12 – Error in soil moisture as a function of actual saturation level .....	63
Table 13 – Error in soil moisture as a function of decay constant error variance .....	64
Table 14 – Error in soil moisture as a function of error in decay constant error level .....	65
Table 15 – Error in soil moisture as a function of error in rain measurement.....	67
Table 16 – Error in soil moisture as a function of error in estimated rain measurement error.....	67
Table 17 – Average distance to square of side length $L$ .....	81
Table 18 – RADARSAT-1 beam modes .....	82
Table 19 – RADARSAT-1 standard beam positions .....	82
Table 20 – RADARSAT-2 beam modes .....	83
Table 21 - Envisat ASAR beam modes.....	83
Table 22 – Envisat ASAR image and alternating polarisation mode beam positions .....	83

# Chapter 1

## Introduction

WATClass is a hydrological model maintained by the Civil Engineering department at the University of Waterloo (Soulis, 2000). It models the water and energy budgets in a watershed by making use of real-time meteorological data. WATClass tracks the flow of water in a watershed as it passes from precipitation or snowmelt into the streams, lakes and rivers. Soil water content is an important state variable in this model as it dictates the amount of rainfall or snowmelt that the ground can absorb. Currently, WATClass does not have a direct method to verify that the soil moisture estimate is accurate. By using remotely sensed soil moisture data, the WATClass model can reduce the amount of error in the soil moisture estimates, thereby presumably increasing the accuracy of the overall flood prediction model.

Methods for global monitoring of soil moisture values in agriculture fields from satellite borne microwave radars and radiometers are currently being developed. Satellite borne synthetic aperture radar (SAR) systems, such as RADARSAT and Envisat ASAR, can provide soil moisture measurements at a 30×30 m resolution (Deschamps, 2004). However, surface roughness and vegetation reduce the ability of these instruments to accurately determine soil moisture values. The effect of surface roughness and vegetation in soil moisture estimation from passive microwave radiometers is much lower than that of active radar. However, the spatial resolution of the microwave radiometers is also much lower than SAR systems' resolution. Therefore, higher resolution soil moisture maps must be interpolated from microwave radiometers for use in hydrological models.

The ensemble Kalman filter (Evensen, 2003) is the proposed algorithm to assimilate the remotely sensed soil moisture estimates into WATClass. The Kalman filter is an algorithm designed to assimilate noisy observations of a state variable into a noisy model of that state variable. The Kalman filter tracks the state variable and the estimated error of the state variable in the model.

When an observation is made, the estimate is updated using the observation. The degree to which the observation is used as a replacement of the model's estimate is determined by the ratio between the observation's estimated error and the model's estimated error. The ensemble Kalman filter is a form of the Kalman filter used in non-linear models, such as WATClass. It uses a large collection of randomly assigned state variables to represent the error distribution of the state variable.

This thesis examines the feasibility of using remotely sensed data, particularly active microwave, to correct the soil moisture estimates in the WATClass model. The goal of this thesis is to determine to what extent satellite borne remotely sensed data can reduce the error in WATClass soil moisture values. To accomplish this, the spatial variability of soil moisture is examined at different scales associated with WATClass and soil moisture measurements. The spatial variability is used to determine the expected variance between the soil moisture measurements and the model soil moisture estimates. Scale in the measurements and the model play a large part in the variability. Soil moisture is highly variable and the variation profile is dependent on the scale in which it is examined.

WATClass is simulated in MATLAB to determine the proposed algorithm's efficiency. Using the simulation, the sensitivity of the estimated error in soil moisture is examined as a function of the errors in the inputs to the model. Specifically, the roles of ensemble size, observation frequency, observation measurement error and model error (error in saturation level, rain measurement and rate of drying) are examined.

Chapter 2 provides some background into soil moisture's role in the hydrologic cycle, its role in WATClass, how microwave remote sensing detects soil moisture and how soil moisture can be assimilated into hydrological systems. Chapter 3 describes the datasets used in this thesis. Chapter 4 describes the methodology of the algorithm used. Chapter 5 presents the results of the thesis. Chapter 6 presents the conclusions and recommendations.

## Chapter 2

### Background

#### 2.1 Soil Moisture

Soil moisture, or soil wetness, is liquid water occupying the empty spaces between soil particles. In dry soil, these spaces are filled with air. The proportion of empty space in a volume of soil, the porosity, is determined by the texture of the soil, i.e. how much of the soil is made up of sand, silt or clay. The soil's water content, or the soil moisture value, for a particular soil is the ratio of water to soil. This can be defined several different ways:

- gravimetric water content: ratio of the weight of water to the weight of dry soil;
- volumetric water content: ratio of the volume of water to the volume of soil; and
- degree of saturation: the ratio of the amount of water currently in the soil to amount of water that would be in the soil if the soil were completely saturated.

During heavy rain, water begins pooling on the surface of the soil. Gravity and capillary forces pull the water into the empty spaces, completely saturating the soil. This process, by which surface water enters the soil, is called infiltration. The boundary between the saturated soil at the surface and non-saturated soil underneath is called the wetting front. As long as there is still pooled water on the surface, gravity continues to pull the wetting front down, increasing the depth of the saturated surface layer of soil. This water is pulled down toward the water table, which is the layer of saturated soil at the bottom of the soil column. This process is called drainage, or ground water recharge. In the water table, the water is either stored or flows underground into the streams and rivers.

If the rate of precipitation exceeds the rate of infiltration, much of the excess water travels as overland flow directly into streams and rivers. When the rate of precipitation drops below the rate of infiltration, air begins to be pulled into the soil, causing the soil moisture at the surface to decrease

below saturation. While drainage will continue indefinitely, the rate at which water drains from the unsaturated zone into the groundwater will slow down as the capillary forces counteract the force of gravity. The soil moisture level at which the drainage rate can be considered negligible is called the field capacity.

In addition to drainage, other processes act to lower the soil moisture: interflow, exfiltration and transpiration. Interflow is the process by which water in the unsaturated layer of soil flows downslope. Exfiltration is water in the surface soil layer evaporating directly into the atmosphere. Transpiration is the process by which plants absorb water through their roots and release the water into the atmosphere through their leaves. Evapotranspiration is the term that envelops all processes by which water near the ground – surface soil moisture, snow, lakes or rivers – enters the atmosphere.

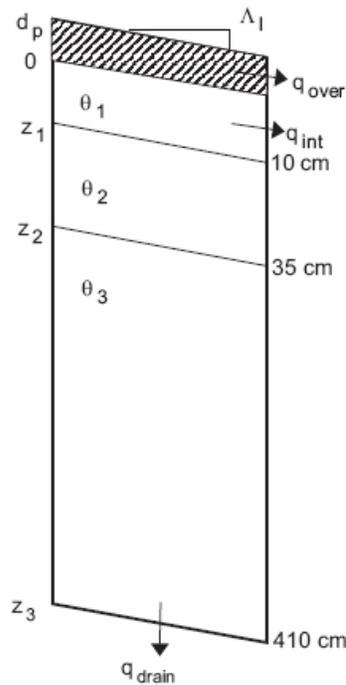
While the upper limit of soil water content is determined by the soil porosity, there are several lower limits other than field capacity that can be considered (Dingman, 2002). The absolute lowest limit is zero water content, which is only achievable by drying the soil in an oven. The next lower limit of water content is hygroscopic water, the state where a thin film of water surrounds each soil particle, held there by electrostatic forces. A soil with water content below this level will begin to draw water directly from the humidity in the air. Finally, permanent wilting point (PWP), the lower limit prescribed to many hydrological systems, is the level of water content below which plants begin to wilt. PWP is a function of soil type, plant species and atmospheric conditions. Thus, when soil water content reaches this value, transpiration ceases and soil moisture values stabilise. Therefore, between rainfall events, the water content of a saturated soil will decrease exponentially due to gravity until it reaches field capacity, at which point it will continue losing water at a much slower rate until it reaches PWP.

## 2.2 WATClass

WATClass is a distributed hydrological model that tracks the energy and water budgets at a meso-scale, or watershed-scale, level (Soulis, 2005). For computational purposes, WATClass divides the watershed into grid squares that range in size from 1×1 km to 25×25 km. In each grid square, the state variables, such as soil temperature and soil moisture, are tracked for each individual land cover type. Examples of land cover types include agriculture, forest, urban, water and wetlands. Remotely sensed LANDSAT imagery and land cover data from land-surface schemes, such as the Canadian Land Surface Scheme (CLASS) (Soulis, 2000) are used to determine the proportion of each land cover that makes up each square. During rainfall events, radar and rain gauges are used to estimate the amount of rainfall in each grid square. The rainfall is then divided amongst the land cover types in each square proportionally to the fraction of the square covered by each land cover type. The runoff from each land cover type is calculated independently from one another and the total runoff from all land types is added to the stream network for that grid square.

The stream network is the network of surface streams and rivers that carry water between grid squares. In WATClass, each square must be directly connected to the stream network. A routing algorithm is then applied to determine the amount of water received from and directed to adjacent grid squares through the stream network. WATClass is then able to calculate water levels in the stream network.

Soil moisture is an important state variable in the WATClass model. The moisture content of the soil is a major factor in the infiltration rate. The infiltration rate, when coupled with the precipitation rate, determines how much runoff occurs. During a rainfall event, water initially begins to pool on the surface and begins to drain into the soil column. WATClass then calculates the soil moisture values as the wetting front descends into the soil column. WATClass also calculates the amount of water flowing into stream network from surface overland flow, interflow and water that



**Figure 1 – WATClass soil column (Soulis, 2000, Fig. 6)**

has drained through the soil column into the groundwater. WATClass stores the average soil moisture estimates at each discrete time step of three distinct vertical layers: 0-10 cm in depth, 10-35 cm in depth and 35-410 cm in depth.

Figure 1 illustrates the soil column as modeled by WATClass. On a rainfall event, the soil moisture values for the three soil layers,  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ , are tracked as well as the volume of water passing from the surface of the soil column to the stream network (overland flow),  $q_{over}$ ; the volume of water passing from the soil column to the stream network (interflow),  $q_{int}$ ; and the volume of water passing from the soil column into the water table,  $q_{drain}$ . The depth of pooled water,  $d_p$ , is also tracked.

The initial soil moisture content of the soil,  $\theta_0$ , can be estimated by the Antecedent Precipitation Index (API),  $I_a(t)$ , which is a soil moisture estimate based on the time since the last rainfall (Kouwen, 2006):

$$\theta_0 = \frac{I_a(0)}{100}. \quad (\text{Equation 2-1})$$

API is calculated as

$$I_a(t) = \alpha \cdot I_a(t) + P(t), \quad (\text{Equation 2-2})$$

where  $I_a(t)$  is the API at discrete time  $t$ ,  
 $\alpha$  is an optimised decay constant (value 0.985-998 if  $t$  is discretised by hour), and  
 $P(t)$  is the precipitation at time  $t$ .

In general,  $\alpha$ , the decay constant, varies from day to day and is a function of the potential evapotranspiration, which itself is a function of plant type and atmospheric conditions, and soil moisture (Teng, 1993).

### 2.3 Soil moisture detection

Soil moisture is measured in several different ways. The most direct approach to determine the water content of a soil is to take a soil sample and calculate the difference between the weight of the wet soil sample and the weight soil sample after it has spent sufficient time in a drying oven. There are also several instruments that make use of the physical characteristics of moist soil, such as electrical conductivity or capacitance. Typically in these instruments, probes are inserted into the soil, some sort of signal is applied to the probes and the water content is calculated as a function of some return signal. For example, the ThetaProbe generates an electrical signal using an array of four probes inserted into the soil (Delta Devices, 1999). The impedance of the soil is then measured using the array. Since the impedance of the soil is highly dependent on the soil's water content, the instrument can then calculate the volumetric water content of the soil. The ThetaProbe is accurate to

$\pm 0.01 \text{ m}^3 \cdot \text{m}^{-3}$  at 0 to 40°C when the instrument is properly calibrated to the specific soil type. Soil moisture measurements range from  $0.00 \text{ m}^3 \cdot \text{m}^{-3}$  to approximately  $0.5 \text{ m}^3 \cdot \text{m}^{-3}$  (Delta Devices, 1999).

The problem with using probe-sensed soil moisture measurements is that soil moisture is highly variable due to local topography, soil heterogeneity and meteorological variability in rain, radiation and temperature (Merz, 1997). Gravimetric or probe measurements provide detailed soil moisture measurements for a region of soil of a few square centimetres, while hydrological models require soil moisture estimates for areas in the range of square kilometres. Gathering sufficient measurements to properly characterise a watershed using a probe would be extremely time-consuming.

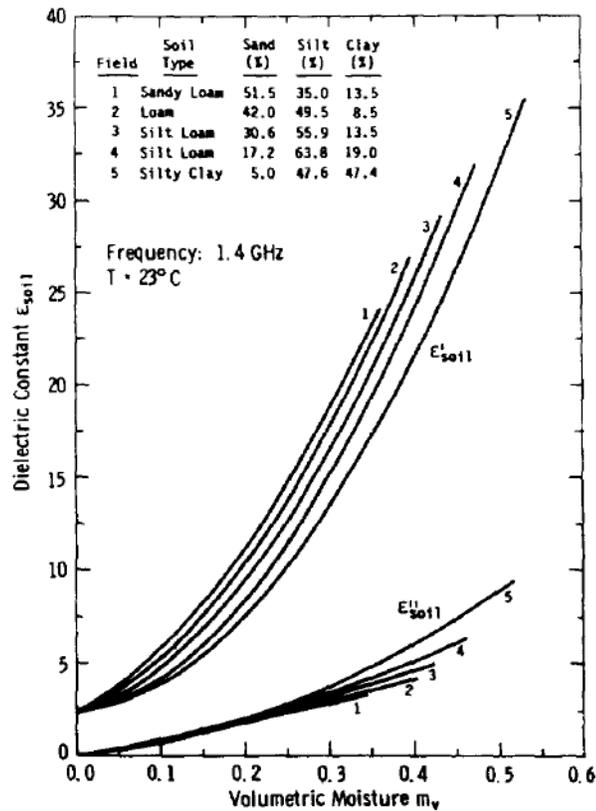
### **2.3.1 Soil moisture estimates by active microwave**

The detection of soil moisture using active microwave, i.e. radar, has been an active area of research for the past twenty years (Ulaby, 1986, Dobson, 1998). Radar images are generated by sending a microwave pulse from an antenna toward a scene and then measuring the strength of the signal that returns to the antenna. For airborne and space borne sensors, the microwave pulse illuminates a long, relatively narrow strip of the Earth's surface perpendicular to the direction of flight. This direction is referred to as the range direction, whereas the direction of flight is referred to as azimuth or along-track direction. When the radar signal strikes the Earth's surface, part of the signal is reflected back to the receiver. This reflected portion of the signal is called the backscatter. The radar instrument can differentiate backscatter from different objects in the range direction by relating the time it takes for the reflected signal to be received back at the antenna to the distance in the range direction. The radar forms two-dimensional images by combining the signals from successive pulses as the radar travels in the azimuth direction.

For real-aperture radar (RAR), the azimuth resolution is limited by the microwave pulse's beam width. Beam width can be decreased by focusing the beam using a larger antenna size or by lowering the altitude of the radar, neither of which are achievable by space borne radar. Therefore, space borne

imaging radars do not use RAR, but rather use synthetic-aperture radar (SAR) (Dobson, 1998). SAR systems simulate large antennae by illuminating the scene with a coherent pulse and analyzing the Doppler frequency shifts of the returned signal. Due to the relative velocity of the radar antenna to the ground, objects ahead of the receiver in the azimuth direction will experience an upshift in frequency, whereas objects behind the receiver will experience a downshift. Furthermore, several sub-images or looks of a scene can be created by partitioning the bandwidth of the received signal. These sub-images can be averaged to increase the signal-to-noise ratio of the image. The trade-off of SAR is the added complexity of the processing of the backscatter signal.

The dielectric constant of a substance, which is a relative measure of the permittivity of a substance to an electromagnetic signal, is one of the main factors in the level of backscatter when that substance is illuminated by electromagnetic energy (Ulaby, 1986). The dielectric constant of water is significantly higher than the dielectric constant of dry soil, enabling soil moisture to be measured by active microwave. Figure 2 shows the relationship between soil moisture and dielectric constant for five different compositions of soils. Note that while soil composition has an effect on dielectric constant, the main contributor to the variability is moisture content.



**Figure 2 – Measured dielectric constant,  $\epsilon_{soil}$ , for five soils at 1.4GHz.**  
 $\epsilon'_{soil} = \text{Re}(\epsilon_{soil})$  and  $\epsilon''_{soil} = \text{Im}(\epsilon_{soil})$  (Ulaby, 1996, Fig 9)

Backscatter levels from active microwave imaging of soil are further affected by surface roughness, the amount and type of vegetation covering the soil, and soil composition. Surface roughness is the geometry of the air-soil boundary. The penetration depth of a radar signal in moist soil is proportional to the wavelength of the signal and inversely proportional to the soil's moisture content. For moist soil, the penetration depth is approximately equal to the wavelength of the signal; therefore, radar platforms are only capable of detecting soil moisture in the top layer of soil (5-20 cm depending on signal wavelength). For C-band SAR instruments, such as ERS-1 and RADARSAT, the penetration depth for moist soil is less than 5 cm (Ulaby, 1986). Due to the backscatter from vegetation over the soil, the best results are produced when detecting soil moisture from bare soil or light vegetation. Taller crops and forest canopies prevent the estimation of soil moisture underneath.

Ulaby demonstrates that the parameters that produce the highest correlation between radar backscatter and soil moisture is C-band signal with an incidence angle of  $10^\circ$  (Ulaby, 1986). At higher incidence angles, surface roughness begins to have a greater effect on backscatter. Space borne SAR systems typically have incidence angles of  $20^\circ$  or higher. SAR can image the surface of the Earth with lower incidence angles, but the range resolution decreases as a consequence.

While much of the research into the use of single-polarisation space borne active microwave for soil moisture detection has failed to produce sufficiently accurate soil moisture estimates due to the inability to account for surface roughness, recent research into the use of multi-frequency and multi-polarisation SAR has shown much promise (Dobson 1998); by comparing several different SAR images across different frequencies and different polarisations the effect of surface roughness on the soil moisture estimate can be mitigated. With single-polarisation SAR, the antenna transmits and receives the signal using one of either a vertically (VV) or horizontally (HH) polarised signal. Dual-polarisation allows the antenna to transmit either a horizontally or vertically polarised signal and receive both horizontally and vertically polarised signals (HH/HV or VV/VH), or, to transmit and receive alternatively with vertically and horizontally polarised signals (HH/VV). Quad-polarisation allows the antenna to transmit alternatively horizontal and vertical signals, and receive horizontally and vertically polarised signals (HH/HV/VH/VV).

Multi-polarised SAR datasets have only recently been available from orbital satellite borne systems. SIR-C/X-SAR, a quad-polarisation and multi-frequency C-, L- and X-band SAR, was shown to accurately estimate soil moisture within 5% (Bindlish, 2000); however, it was flown as part of two space shuttle missions and was only operable each mission for ten days. In 2002, Envisat ASAR was the first multi-polarisation SAR system to be put into orbit (ESA, 2007). RADARSAT-2, which is scheduled to launch in March 2007, will also provide quad-polarisation imagery in the C-band (McNairn, 2004). Table 1 describes several current space borne SAR platforms.

**Table 1 – Orbital SAR systems launched since 1994 (Raney, 1998, ESA, 2007, Ali, 2004)**

	<b>SIR-C/X-SAR</b>	<b>ERS-2</b>	<b>Envisat ASAR</b>	<b>RADARSAT-1</b>	<b>RADARSAT-2</b>
Country	USA Germany, Italy	Europe	Europe	Canada	Canada
Agency	NASA/DLR/DARA	ESA	ESA	CSA/USA	CSA/USA
Spacecraft	Space Shuttle	ERS-2	Envisat	RADARSAT-1	RADARSAT2
Launch Date	Apr 94, Oct 94	1995	Feb 2002	Nov 95	Mar 2007
DesignLifetime	10 days	2-3 years	5 years	5 years	5 years
Band	L, C, X	C	C	C	C
Frequency (GHz)	1.25, 5.3, 9.6	5.3	5.3	5.3	5.3
Polarisation	Quad (L,C), VV (X)	VV	Dual	HH	Quad
Incidence angle	15-55°	23°	15-45°	<20->50°	20-41° (quad)
Repeat Cycle	nil	35 days	35 days	24 days	24 days

Another technique to reduce the effect of unknown surface roughness and vegetation cover is to compare the temporal variability of several images (Moeremans, 2000, Moran, 2000). This technique involves examining the change in backscatter coefficients over time instead of the backscatter coefficients themselves. Since surface roughness and vegetation cover for a particular field remain relatively constant over time, the variability in the change in backscatter coefficients is more dependent on soil water content.

A drawback to the high spatial resolution of active microwave imagery from orbital satellites is the low temporal resolution. For example, once RADARSAT-1/2 images a particular swath of the Earth's surface, the satellite will not revisit the same swath from the same vantage point for another twenty-four days. The repeat cycle for Envisat is thirty-five days. Both RADARSAT-1/2 and Envisat ASAR offer several beam modes within each imaging mode which require the radar beam to be redirected, allowing more frequent sampling of a particular area on the earth (ESA, 2007, Ali, 2004). Note that there are higher levels of noise in finer resolution imaging modes, due to a lesser number of "looks". The number of looks is defined as the number of independent samples taken of the same scene from the imaging radar.

Furthermore, since both RADARSAT and Envisat follow a polar orbit, the swath width overlap increases at increased latitude (ESA, 2007, Ali, 2004). Therefore, larger latitudes will be able to be imaged more frequently. However, these latitudes will be imaged from a slightly different vantage

point each time they are imaged within the same repeat cycle. Tables of the beam modes and beam positions for RADARSAT and ENVISAT are shown in Appendix C.

### **2.3.2 Soil moisture estimates by passive microwave**

Passive microwave is becoming an established method of estimating soil moisture values on a global scale (Kerr, 2001, Njoku, 1996, Shi, 2006). While active microwave uses the reflectance of a signal produced by the device, passive microwave measures the brightness temperature from the soil surface. However, since the amount of microwave energy emitted from the Earth is so small, the resolution for passive microwave imagery is quite low. L-band passive microwave has been shown to be most sensitive to soil moisture from passive radiometers (Njoku, 1996). Furthermore, due to the longer wavelength of L-band signal compared to C- or X-band, the microwave signal emitted from the soil originates from deeper in the soil column.

There are two current missions to provide space borne near-daily global soil moisture monitoring: NASA's Aqua mission and ESA's Soil Moisture and Ocean Salinity (SMOS) mission. The Aqua satellite is equipped with the Japanese Advanced Microwave Scanning Radiometer (AMSR) to measure soil moisture for watershed-scale hydrology purposes (Njoku, 2003). The AMSR instrument has two sensors that are sensitive to soil moisture: 6.925 GHz (C-band, 4.3 cm) and 10.65 GHz (X-band, 2.8 cm). It provides soil moisture estimates of the same location on earth and the same vantage point from space once every three days, i.e. its repeat cycle is three days. The resolution of this sensor at these two frequencies is 60 km. The Aqua satellite has been operational since 2002.

The SMOS satellite, scheduled to launch in early 2008, will also provide frequent global soil moisture estimates (Kerr, 2001). The SMOS satellite will carry the Microwave Imaging Radiometer using the Aperture Synthesis (MIRAS), an interferometric passive microwave radiometer. MIRAS uses an array of smaller sensors to simulate a larger antenna to estimate soil moisture from the most

sensitive L-band with suitable resolution for meso-scale hydrological purposes. The resolution of the MIRAS imagery is <50 km and its repeat cycle is three days.

The effect of surface roughness in space borne microwave radiometers is much less than that of microwave radars (Njoku, 2003). However, the resolution of passive radiometers is constrained by the size of the antenna that can be placed on the satellite, unlike active radiometers where a large antenna is simulated by controlling the phase of the emitted microwave signal and examining the phase of the reflected signal. Therefore, passive microwave sensors have much lower spatial resolution than active microwave sensors.

Active research is also being conducted into combining the accuracy of passive microwave sensors and the spatial resolution of active microwave sensors. NASA's Hydrosphere State (HYDROS) mission, which was scheduled to launch in late 2009 or 2010, combines an active L-band SAR and passive L-band radiometer on the same platform. The resolutions of the active and passive microwave sensors are 3 km and 40 km respectively (Entekhabi, 2004). Using a Bayesian approach, the active and passive microwave images are combined to form an aggregate image with a resolution of 10 km. Although the mission was cancelled due to budget cuts, the mission is still open to be revived in the future.

### **2.3.3 Summary of remote sensing soil moisture estimates**

The most accurate means to estimate soil water content is the gravimetric method, where the soil is collected, weighed, dried in an oven and weighed again. Other ground based methods, such as measurement by probes or truck-based microwave sensors also provide accurate estimates of soil moisture. However, these methods are only practical for determining soil moisture for small areas and are consequently unfeasible for the monitoring of soil moisture at a watershed level. Microwave remote sensing provides a means to monitor soil moisture at a watershed level. While aerial microwave sensors provide enhanced spatial resolution, orbital satellite borne microwave sensors can provide ongoing near daily temporal resolution of soil moisture of the entire surface of the Earth.

Space borne active microwave sensors, such as RADARSAT and ERS-2 provide soil moisture estimates at high spatial resolution (30 m). However, the temporal resolution of these SAR systems is several weeks unless the beam is redirected. Furthermore, the accuracy SAR-based soil moisture estimation is degraded by surface roughness and vegetation cover, which are difficult to estimate from single-polarisation, single-frequency SAR. By examining multi-frequency, multi-polarisation and/or multi-temporal SAR images, errors in soil moisture values caused by surface roughness and vegetation cover can be reduced.

Active research in both active and passive microwave sensors is continually improving the quality of soil moisture estimates. The newest generations of orbital satellite borne microwave sensors aim to provide researchers, hydrologists and meteorologists with near daily global soil moisture estimates.

## **2.4 Soil moisture assimilation into hydrological models**

With the advance in soil moisture detection by space borne microwave sensors, the use of this data in hydrological models has become an active area of research. Due to the insufficient spatial resolution of current passive microwave radiometers and the insufficient temporal resolution and inaccuracy caused by surface roughness and vegetation cover of active microwave radars, direct substitution of the remotely-sensed soil moisture estimates into hydrological models is not currently feasible. A popular technique to assimilate microwave-derived soil moisture estimates into models is the use of a Kalman filter (Galantowicz, 1999, Hoeben, 2000, Crosson, 2002, Reichle, 2002, Aubert, 2003) due to its efficiency and simplicity. The Kalman is the optimal Bayesian filter for estimating the state of a noisy linear system from a set of noisy observations. Since hydrological equations involved with these models are non-linear, a non-linear form of the Kalman filter must be used.

Margulis implemented an ensemble-based non-linear form of the Kalman filter to assimilate airborne passive L-band microwave soil moisture observations during the Southern Great Plains 1997

field experiment into the NOAH hydrological model (Margulis, 2002). The Electronically Scanned Thinned Array Radiometer (ESTAR) was flown at an altitude of 7.5 km above a 10,000 km<sup>2</sup> study area on sixteen days during the thirty day experiment. Margulis's algorithm forms the basis for the algorithm described in this thesis.

Sections 2.4.1 and 2.4.2 discuss the Kalman filter and the extended Kalman filter respectively. The extended Kalman filter is a non-linear form of the Kalman filter. Section 2.4.3 discusses the ensemble Kalman filter, which is another non-linear form of the Kalman filter, and Margulis's use of this algorithm for assimilation of remotely-sensed soil moisture measurements into a hydrological model.

#### **2.4.1 The Kalman filter**

The Kalman filter is an algorithm used to update a linear model of some state variables with periodic observations, with the goal of minimising the error between the model and the true values of those state variables (Kalman, 1960).

In its simplest form, the Kalman filter tracks some real world state variable using a model of that variable and the periodic observations of the variable. It estimates the state variable's value as a function of previous observations and an a priori model of the variable, and also calculates the corresponding estimated variance.

When an observation is made, the algorithm combines the observed value and the model value to produce a better estimation of the state. If it is assumed that there is no error between the observation and the true value, then the model can rightly substitute the observation for the model's estimated value. In general, there is an error associated with the observation; thus the resultant estimation, or a posteriori estimate, is some linear combination of the observation and the model estimate, with the proportions dictated by the corresponding estimated error variances. The a posteriori estimate will also have an estimated error variance calculated from the a priori estimated error variance and the observation estimated error variance. In general, the Kalman filter tracks

several state variables and their corresponding estimated error covariances, and the observations are some linear combination of these state variables.

If the state and observations are defined as

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}, \text{ and} \quad (\text{Equation 2-3})$$

$$z_k = Hx_k + v_k, \quad (\text{Equation 2-4})$$

where  $x_k \in \mathfrak{R}^n$  is the state being modeled at discrete time  $k$ ,  
 $A_{n \times n}$  is the transition matrix between  $x_{k-1}$  and  $x_k$ ,  
 $u_k \in \mathfrak{R}^m$  is a control input at time  $k$ ,  
 $B_{n \times m}$  is the transformation matrix between  $u_{k-1}$  and  $x_k$ ,  
 $w_{k-1} \sim N(0, Q)$  is the process noise,  
 $z_k \in \mathfrak{R}^p$  is the observation at time  $k$ ,  
 $H_{p \times n}$  is the transformation matrix between  $z_k$  and  $x_k$ , and  
 $v_{k-1} \sim N(0, R)$  is the measurement noise,

and the estimates, errors and estimated error covariances are defined as

$$\hat{x}_k^- \in \mathfrak{R}^n \text{ as the a priori state estimate,}$$

$$\hat{x}_k \in \mathfrak{R}^n \text{ as the a posteriori state estimate,}$$

$$e_k^- \equiv x_k - \hat{x}_k^- \text{ as the a priori estimate error,}$$

$$e_k \equiv x_k - \hat{x}_k \text{ as the a posteriori estimate error,}$$

$$P_k^- = E \left[ e_k^- e_k^{-T} \right] \text{ as the a priori estimate error covariance, and}$$

$$P_k = E \left[ e_k e_k^T \right] \text{ as the a posteriori estimate error covariance,}$$

then the Kalman filter predict equations can be described as

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}, \text{ and} \quad (\text{Equation 2-5})$$

$$P_k^- = AP_{k-1}A^T + Q. \quad (\text{Equation 2-6})$$

These equations are used to predict the state in the model during the time between observations.

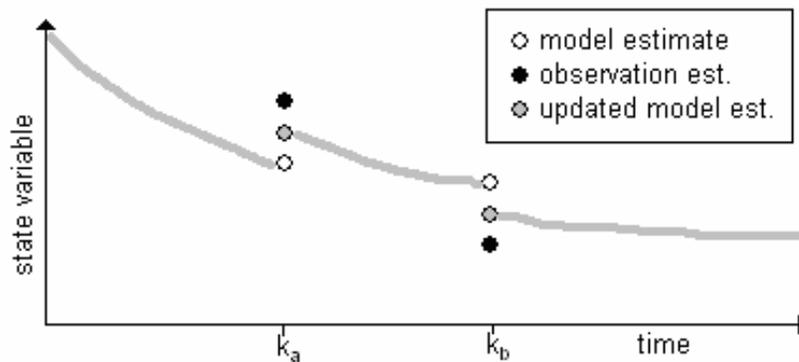
Therefore, they are called the predict equations in the context of the Kalman filter.

When an observation is made at time  $k$ , the following equations are used to update the estimate:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}, \quad (\text{Equation 2-7})$$

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-), \text{ and} \quad (\text{Equation 2-8})$$

$$P_k = (I - K_k H)P_k^-. \quad (\text{Equation 2-9})$$



**Figure 3 – Kalman filter example**

The Kalman gain,  $K_{nxl}$ , dictates how much of the priori estimate and how much of the observation make up the a posteriori estimate. These equations are called the update equations in the context of the Kalman filter.

Figure 3 demonstrates a simple one-dimensional example of the filter. Before time  $k_a$ , the state is estimated solely by the model. The estimated error covariance, which in the one-dimensional case is just the estimate error variance of the state, is also tracked. At time  $k_a$ , an observation is made and the Kalman gain is calculated based on the estimated observation error variance and the estimated error variance of the model-predicted estimate. The Kalman gain determines the ratio of the observation and the model estimate that will make up the new estimate. The model is then propagated forward using this new estimate.

### **2.4.2 The extended Kalman filter**

The Kalman filter is designed to estimate the state of a system controlled by linear stochastic difference equations from a series of noisy and incomplete measurements; it is not designed to estimate non-linear systems. The extended Kalman filter is a modification of the Kalman filter for non-linear stochastic difference equations (Julier, 1997). In essence, the extended Kalman filter attempts to linearise the system, for the purposes of determining the error covariance matrices. The linearisation is done by taking the partial derivatives of the non-linear state control functions with

respect to the state variables and the noise to form the transition matrices. While the model remains non-linear, the estimated error distribution is modeled as a normal distribution, using the transition matrices in the same way as the linear form of the Kalman filter to determine the Kalman gain. However, the error distributions of a non-linear system are not normally distributed. Furthermore, the state control equations must be in explicit form such that the partial derivatives can be taken. The ensemble Kalman filter is designed to solve these problems.

### **2.4.3 The ensemble Kalman filter**

Instead of parameterising the estimate error distribution as per the extended Kalman filter, the ensemble Kalman filter uses an ensemble of state variables at each time step to represent the estimated error distribution (Evensen, 2003). Each point in the ensemble is individually propagated through the model until an observation is made. Then, the estimated error distribution is approximated as a normal distribution for use in the Kalman gain equation. This differs from the extended Kalman filter in that the ensemble Kalman filter maintains the estimated error distribution throughout the propagation of the model, whereas the extended Kalman filter approximates the estimated error distribution as normally distributed throughout the model. Therefore, the ensemble filter provides a better representation of the error statistics than the extended Kalman filter. The trade-off between the ensemble Kalman filter and the extended Kalman filter is an increase in computation time; while the extended Kalman filter only needs to run one set of state variables through the model, the ensemble Kalman filter requires many state variables to be propagated through the model. As the system being modeled increases in complexity, the set of state variables, called the ensemble, must increase in size to properly represent the error distribution.

Another advantage of the ensemble Kalman filter is that the model can be treated as a “black box”, with only its input controlled and outputs examined. No access to the underlying equations of the model is needed.

In Margulis's soil moisture assimilation algorithm, the state variable estimated by the Kalman filter is soil moisture (Margulis, 2002). The soil moisture values are assumed to be spatially independent; therefore, the state is one-dimensional. In WATClass, soil moisture is also spatially independent, in that the soil moisture in one grid square does not affect the soil moisture of a neighboring grid square; interaction between grid squares is reserved to water in the stream network. Therefore, the soil moisture of an individual grid square can be examined without examining the grid squares around it; altering the soil moisture in a particular grid square cannot affect the soil moisture of another grid square in WATClass. Note that the inputs to the WATClass model, such as rainfall and temperature, are spatially correlated due to the low spatial variability of meteorological conditions. Thus, it is possible that running all the grid squares concurrently might increase accuracy.

To assimilate soil moisture estimates into the hydrological model, the ensemble Kalman filter creates  $n$  individual replications of the model, i.e. an ensemble of state variables. In each replicate, the uncertain inputs are randomly assigned from normal distributions such that collectively, the set of inputs approximate the estimate error distribution. For example, the Margulis implementation of the ensemble Kalman filter on their hydrological model randomised initial soil moisture measurements, hydraulic conductivity of the soil, upper and lower limits to soil moisture (porosity and permanent wilting point) and precipitation levels (Margulis, 2002).

In this implementation, the model state is assumed to be a function of the previous state, the time-independent parameters, such as soil-type, vegetation type, topography, etc., and time-dependent parameters, such as atmospheric conditions and precipitation levels,

$$\theta(t) = f(\theta(t-1), b_1(t), \dots, b_p(t), \alpha_1, \dots, \alpha_q) \quad (\text{Equation 2-10})$$

where  $\theta(t)$  is the soil moisture at time  $t$   
 $b_1(t) \dots b_p(t)$  is the set of  $p$  time dependent parameters  
 $\alpha_1(t) \dots \alpha_q(t)$  is the set of  $q$  time independent parameters  
 $f(\dots)$  is the equation that governs the soil moisture

However, neither the initial soil moisture value nor the parameters are known exactly. Therefore, the initial soil moisture values and each of the parameter values are represented by an ensemble of sample points, or replicates, drawn from a normal distribution with mean equal to the estimated value of the parameter and a variance equal to the estimated error of the parameter. It is assumed that the error variances are known for all of the inputs. Therefore,

$\boldsymbol{\theta}(0)$  is an ensemble of  $k$  soil moisture estimates at time 0,

$\theta_j(0)$  drawn from  $N(\mu_{\theta 0}, \sigma_{\theta 0}^2)$  is the  $j^{\text{th}}$  replicate in  $\boldsymbol{\theta}(0)$

$\mu_{\theta 0}$  is the estimated initial soil moisture value

$\sigma_{\theta 0}^2$  is the estimated error in the initial soil moisture value

$\mathbf{b}_i(t)$  is the ensemble of  $k$  time dependent parameters for  $b_i$

$b_{ij}(t)$  drawn from  $N(\mu_{b_i}(t), \sigma_{b_i}^2)$  is the  $j^{\text{th}}$  replicate in  $\mathbf{b}_i(t)$

$\mu_{b_i}(t)$  is the estimated value of parameter  $b_i$  at time  $t$

$\sigma_{b_i}^2$  is the estimated error in parameter  $b_i$

$\boldsymbol{\alpha}_i$  is the ensemble of  $k$  time dependent parameters for  $\alpha_i$

$\alpha_{ij}$  drawn from  $N(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)$  is the  $j^{\text{th}}$  replicate in  $\boldsymbol{\alpha}_i$

$\mu_{\alpha_i}$  is the estimated value of parameter  $\alpha_i$

$\sigma_{\alpha_i}^2$  is the estimated error in parameter  $\alpha_i$

$k$  is the number of replicates in the ensemble.

Each replicate is propagated forward in time until using,

$$\boldsymbol{\theta}_i(t) = f(\boldsymbol{\theta}_i(t-1), b_{i_1}(t), \dots, b_{i_{p_i}}(t), \boldsymbol{\alpha}_{i_1}, \dots, \boldsymbol{\alpha}_{i_{q_i}}) \quad (\text{Equation 2-11})$$

until a observation is made.

The a priori estimate error covariance is estimated by the sample covariance of the ensemble,

$$C_{\theta\theta}(t) = \frac{1}{k} \left( \begin{matrix} \boldsymbol{\theta} \\ (t) \end{matrix} - \hat{\boldsymbol{\mu}}_{\theta}(t) \right) \left( \begin{matrix} \boldsymbol{\theta} \\ (t) \end{matrix} - \hat{\boldsymbol{\mu}}_{\theta}(t) \right)^T \quad (\text{Equation 2-12})$$

where  $\mu_{\theta}(t)$  is the sample mean of the soil moisture values at time  $t$ .

The Kalman gain, the a posteriori soil moisture estimate and the a posteriori estimate error covariance is then calculated using Equations 2-7 to 2-9. The a posteriori estimate and error covariance are used to create the replicates in the ensemble for the model until the next observation is made.

## Chapter 3

### Data

There are two sources for the data used in this thesis. The first set of data was collected by the hydrology lab at the University of Waterloo. The dataset consists of ground soil moisture measurements and RADARSAT-1 imagery of the Grand River watershed of southern Ontario during 1996 and 1997 (Seglenieks, 1998).

The second set of data was collected by the Canadian Centre of Remote Sensing. This dataset consists of ground soil moisture measurements made in the Roseau River watershed in southern Manitoba between September 2002 and June 2003. The Roseau dataset is part of the dataset used in the Deschamps study (Deschamps, 2004).

#### 3.1 Grand River watershed data

The Grand River watershed data is composed of seventeen pasture fields selected in the Grand River watershed in Ontario. Ten soil samples (depths 0-5 cm and 5-10 cm) were taken from each field on thirteen days during 1996 and 1997, and the gravimetric soil moisture values were calculated. The grass height and soil composition were also measured. Soil composition was used to determine field capacity. Field capacities ranged from 16.6% to 32.4% for the fields. Field capacity, as previously stated, refers to the water content a soil can maintain against the force of gravity. Therefore, this variation indicates that the fields will drain at different rates. Table 2 describes the field locations and field capacities. Figure 4 shows the placement of the fields. Note that the Grand River watershed's dimensions are approximately 200×200 km. Therefore, the fields chosen cover only a small part of the watershed.

**Table 2 – Grand River watershed field data**

The 'x' symbols in the table denote fields that were unable to be sampled on a given date.

Field name	UTM Easting	UTM Northing	Field Capacity	8-Sep-96	15-Sep-96	2-Oct-96	26-Oct-96	19-Nov-96	6-Jun-97	13-Jun-97	7-Jul-97	17-Jul-97	24-Jul-97	31-Jul-97	10-Aug-97	3-Sep-97
W1	531	4864	0.2831													
W2	530	4867	0.2726													
W3	538	4863	0.3031													
W4	538	4854	0.3075		x											
W5	537	4852	0.3054													
W6	532	4838	0.3051						x				x			
E1	554	4843	0.1662	x									x			
E2	560	4864	0.2446													
E3	548	4853	0.3004													
E4	546	4852	0.3243													
E5	543	4847	0.2726	x												
S1	558	4839	0.2069													
S2	554	4836	0.2642													
S3	555	4832	0.1940		x											
S4	555	4827	0.2598													
S5	551	4830	0.2641		x	x										
S6	545	4830	0.2363													
<b>Radarsat Beam Mode</b>				S2	S1	S2	S2	S2	S1	EL1	EL1	S2	S1	EL1	S2	S2

UTM Zone 17.

RADARSAT-1 imagery is also provided for the area. Since the repeat cycle for RADARSAT-1 is twenty-four days, different beam modes are used to image the fields more frequently. The beam modes for RADARSAT are shown in Table 2. S1 and S2 are both the standard beam modes; however the images are taken from different points on the RADARSAT orbit path, i.e. different vantage points. The incidence angles for the fields imaged in S1 range from 20.1-22.7°. The incidence angles for the fields imaged in S2 range from 25.1-27.8°. The extended – low incidence angle beam mode is also used. The incidence angles for the fields imaged in extended low incidence beam mode (EL1) range from 12.7-16.1°.

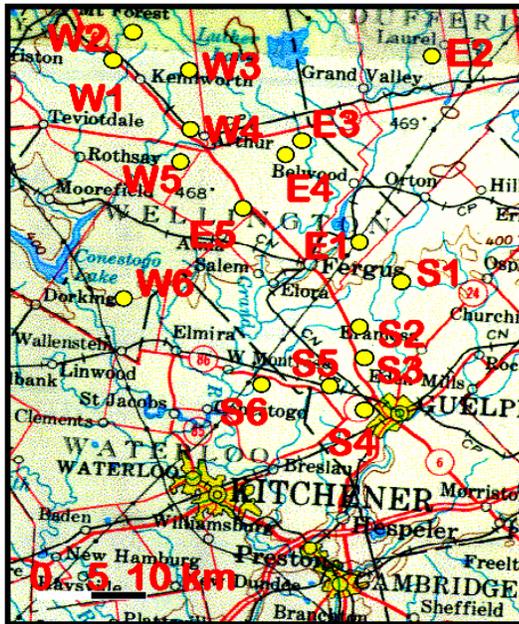


Figure 4 – Site placement in Grand River watershed (Seglenieks, 1998, Fig. 1)

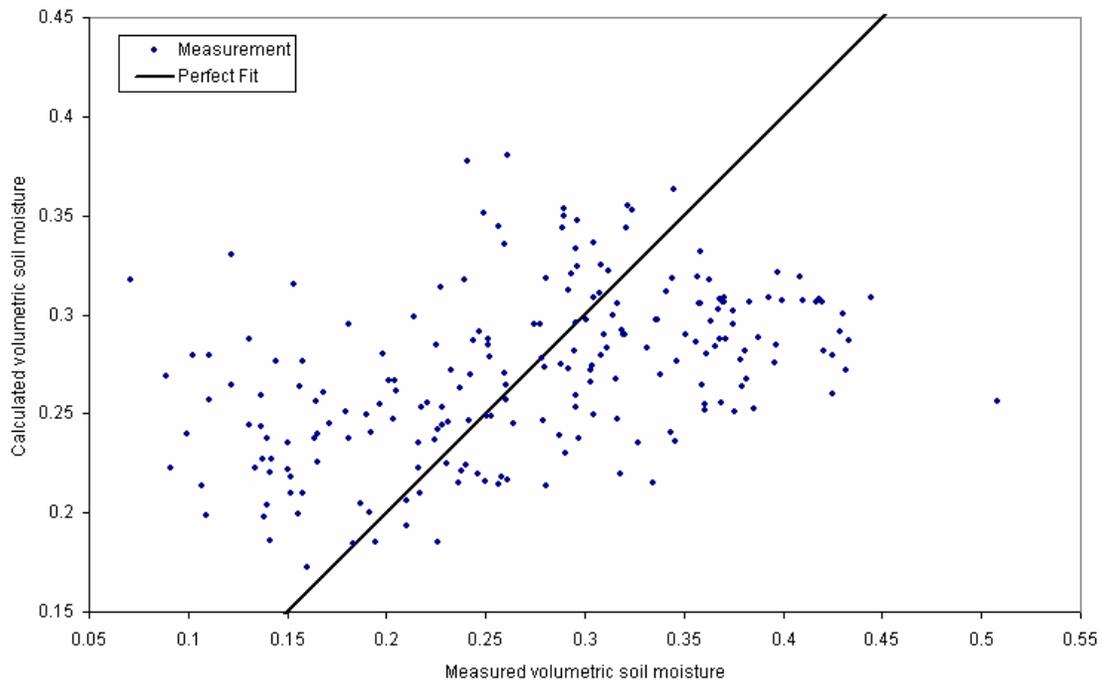
When a linear regression is performed with two independent variables (backscatter value and incident angle) and one dependent variable (volumetric water content), there is little correlation found ( $R^2 = 0.207$ , standard error = 0.0817) (Seglenieks, 1998). Using the resultant equation from the regression, soil moisture estimates are calculated from the remotely-sensed data. Figure 5 shows the ground sampled soil moisture measurements (measured) versus these calculated estimates. When a field dependent constant is added to the regression equations,

$$\theta = a + a_{\text{field}} + b \text{ backscatter} + c \text{ incidence angle} , \quad (\text{Equation 3-1})$$

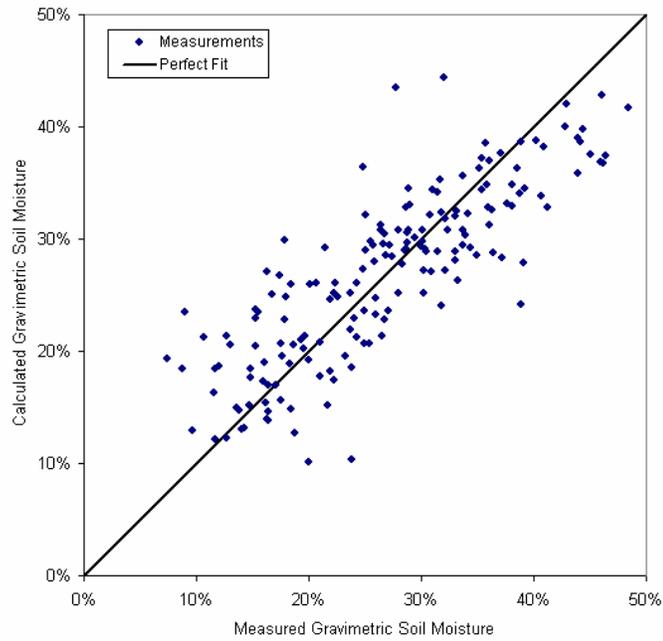
where  $\theta$  is the volumetric water content,  
 $a, b, c$  are constants, and  
 $a_{\text{field}}$  is a field dependent constant,

the correlation increased ( $R^2 = 0.702$ , standard error = 0.0499) (Figure 6). One possible reason for the increased correlation is the local incidence angle, which was not taken into account in the regression. Finally, when volumetric water content values are averaged across all fields for each day and the same regression is performed, the correlation increases further ( $R^2 = 0.941$ , standard error = 0.0255) (Figure 7). However, the spatial information is lost when averaging is performed.

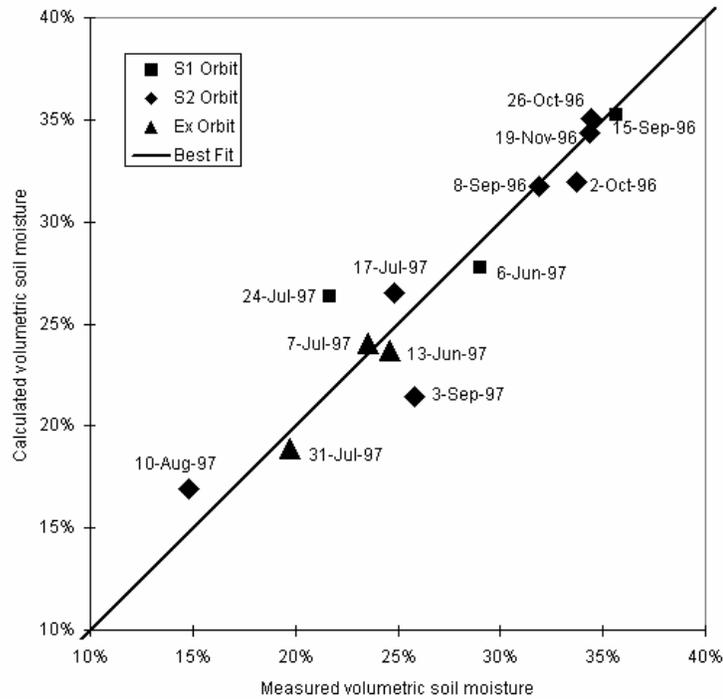
The WATClass data is also included in the dataset. The WATClass study period runs from January 1, 1996 to December 31, 1997 and covers the entire Grand River watershed. Meteorological data is included for each of the 226 10×10 km grid squares sampled every hour throughout the study period. Using this meteorological data and land class information, WATClass can estimate the soil moisture of each grid square every half hour throughout the study period.



**Figure 5 – Measured vs. calculated volumetric soil moisture values for regression using all data**



**Figure 6 – Measured vs. calculated volumetric soil moisture values for regression with field dependent constant**

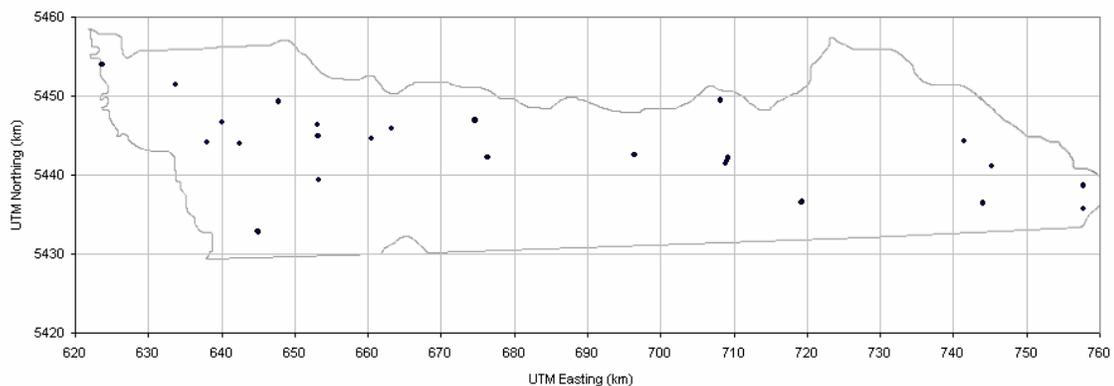


**Figure 7 – Measured vs. calculated volumetric soil moisture values for regression using daily sums**

### 3.2 Roseau River data

The Roseau River data is composed of twenty-two bare fields sampled in the fall of 2002 and the spring of 2003 (Deschamps, 2004). The soil moisture was measured by the ThetaProbe soil moisture detector at three times at five locations per site, totaling fifteen measurements per site per date. Furthermore, surface roughness was measured at each field using the SRM-200 surface roughness meter. Local environmental conditions were measured from three meteorological stations situated throughout the watershed. Figure 8 shows the locations of the fields as well as an outline of the Canadian portion of the Roseau River watershed. Universal Transverse Mercator (UTM) coordinates are for zone 14. Several fields on the eastern portion of the watershed are found in zone 15; their corresponding UTM coordinates as if these fields were in zone 14 are shown. Table 3 presents the field UTM coordinates and the days that each field was sampled.

Coincident SAR from both RADARSAT-1 and Envisat ASAR was collected on the dates that the ground sampling was taken. The instrument used, beam mode and polarisation are shown in Table 3. Linear regression was performed using volumetric soil water content as the dependent variable and backscatter coefficient, incidence angle and surface roughness as independent variables, for each instrument and polarisation. The regression  $R^2$  values vary from 0.2606 for Envisat HH/VV to 0.8032 for Envisat HH (Deschamps, 2004). When daily averages across basin are computed,  $R^2$  values increase to 0.993 for RADARSAT and 0.911 for RADARSAT and Envisat ASAR combined.



**Figure 8 – Roseau sample site and met-station locations.**

**Table 3 – Roseau River dataset field data**

Field Name	UTM Easting	UTM Northing	Oct 1, 02	Cot 6, 02	Oct 25, 02	Oct 28, 02	Apr 11, 03	Apr 15, 03	May 4, 03	May 5, 03	May 8, 03
A1_FOR_1	645	5433		x				x	x	x	x
A1_GLE_1	638	5444		x				x	x	x	x
A1_GLE_2	640	5447		x				x	x		x
A1_KIR_2	642	5444		x		x		x	x		x
A1_SAB_2	626	5453		x				x	x	x	x
A1_TER_1	640	5450		x				x	x		x
A2_DYC_1	662	5446				x		x			
A2_GRI_1	660	5446						x		x	x
A2_HIR_1	694	5442				x					
A2_LOE_1	674	5442						x			
A2_NED_1	657	5442				x		x		x	x
A2_PAL_1	649	5449			x			x			x
A2_REI_1	657	5445				x		x		x	x
A2_SMO_1	685	5446						x		x	
A3_ECK_1	747	5440				x					x
A3_EWA_2	755	5439				x				x	x
A3_GOB_2	742	5443				x				x	x
A3_GOT_1	740	5437				x					x
A3_MCC_2	709	5442		x							
A3_NOR_1	718	5437									
A3_TKA_1	709	5442				x				x	
A3_WAN_1	708	5449									
Radarsat beam mode			W1	-	W1	S4	W1	-	-	W1	-
Envisat-ASAR beam mode			-	IS1	-	IS3	-	IS1	IS1	-	IS2
Polarisation			HH	VV	HH	HH*	HH	HH/ VV	HH/ VV	HH	HH

UTM Zone 14 – x symbols indicate where no data is available.

\*both Radarsat and Envisat-ASAR have HH polarisation.

## Chapter 4

### Methodology

#### 4.1 Objectives

The objective of this thesis is to determine the feasibility of using space borne microwave remote sensing for assimilation into the WATClass hydrological model. More emphasis is put on active microwave, since the dataset provided contains RADARSAT backscatter values. However, the theoretical use of passive microwave radiometers will also be examined.

Currently, the WATClass soil moisture values are initialised using the API, which is solely determined by rainfall gauges; that is, the soil moistures values only indirectly observed. The WATClass model continually updates the soil moisture values throughout the computation of the model. However, WATClass does not use real-world observations of the soil moisture to correct the values in the model. While outside the scope of this thesis, the future goal of this project is to be able to update the WATClass soil moisture values with observed soil moisture values taken periodically from satellites.

The first part of the methodology examines the spatial variability of soil moisture in the WATClass model. The spatial correlation between soil moisture fields of the two datasets is examined. This spatial correlation will be used to estimate the observation error covariance matrix. This part of the methodology is explained in Section 4.2.

The second part of the methodology examines the temporal variability of soil moisture in WATClass, as explained in Section 4.3. The Kalman filter is implemented using a model of WATClass. The purpose of the implementation is to determine the effectiveness of the Kalman filter for assimilating space borne soil moisture estimates.

The microwave remote sensing techniques of soil water content studied in this thesis are only valid for liquid soil moisture; frozen soil moisture or snow covered soil, though modeled in WATClass, is not studied. Therefore, the algorithm described in this thesis is only applicable during the months during which temperatures remain above freezing. In the Grand River watershed datasets examined in this thesis, the ground remained thawed between the beginning of May and the end of September. Unfortunately, soil moisture estimates cannot be estimated during the spring thaw, when there is still snow concealing the soil surface.

Furthermore, since soil moisture can only be estimated in bare soil or light vegetation fields (i.e. agriculture crops and grasses), this method can only be used to update one land class in WATClass, namely the agriculture land class. Therefore, this algorithm will be most applicable to watersheds that are primarily agricultural.

## **4.2 Scale and spatial correlation of soil moisture**

When assimilating soil moisture measurements into a hydrological system, the scale of the measurements and the estimates must be considered. The spatial variability of soil moisture is quite different at different scales. When examining soil moisture at the micro-scale ( $< 1$  km), variations in soil moisture are typically determined by local topography, which determines where pooling will occur when it rains. At a slightly coarser scale where the average soil moisture of fields are examined, differences in the fields' tillage, the amount and type of vegetation and the soil composition contribute to the variability of soil water content value. At a larger scale where soil moisture averages are taken over large areas such as WATClass grid square, coarse-scaled trends in soil variation and topography, and meteorological conditions are the main contributors to the variability in the soil moisture measurements.

Consider the spatial resolutions of the different means of soil moisture measurements. For soil moisture values that are calculated gravimetrically in the field, cylinders of earth approximately 10

cm in diameter are taken. The resolution of space borne active microwave is approximately 30×30 m. Passive microwave images have resolutions of about 50×50 km. Finally, the WATClass grid squares modeling the Grand River watershed are 10×10 km in size. To downscale the measurements, the soil moisture values from the finer scale can simply be averaged to determine the average soil moisture measurement for the coarser scale. In general, however, the soil moisture at the finer scale is not available for every point at the finer scale. For example, to sample every point in a field gravimetrically, the entire field would have to be dug up and measured. This is clearly unfeasible. Instead, ten to fifteen soil samples in the field are taken and the average of these measurements is used to determine the soil moisture of the field. The number of the samples needed to correctly quantify the average is determined by the spatial variability at the finer scale. If the soil moisture at the scale in question does not vary much, then fewer measurements are needed. Conversely, if the soil moisture at the scale is high, then many measurements are needed. In either case, it is important to know the spatial variability of the soil moisture at the scale in question.

Space borne active microwave soil moisture estimates must be downscaled for use in WATClass. If locations of all the fields that fall under the agriculture land class in WATClass and their soil moisture values can all be measured from the radar backscatter, then the simplest way to downsample the radar image is to average the soil moisture estimate for all radar image's pixels that represent agriculture fields. However, sampling every field over the entire watershed may not be possible. For example, if the measurement technique requires the backscatter coefficient to soil water content inversion formula to be calibrated using ground measurements, it would be infeasible to use every field in the entire watershed. Therefore, a subset of the watershed's fields would have to be used to represent all the fields.

This discussion raises the question of at what scale is remotely-sensed soil moisture measurements most useful. It has been shown that the given a particular model resolution, the optimal scale for remotely-sensed soil moisture is the scale of the model or finer (Walker, 2004).

WATClass, however, is capable of spatially discretising at varying resolutions; there requirement being that each grid square in a WATClass model watershed must be connected to a stream network, such that the streamflow routing algorithm can function correctly (Kouwen, 1993). Therefore, it is possible to run the WATClass model in such a way that each grid cell corresponds to a pixel in an active microwave remotely-sensed image (~30 m). While this method would be preferred for exercising the assimilation algorithm (i.e. the ensemble Kalman Filter) since no downsampling algorithm is required. Whether or not the extra spatial acuity will favourably affect the overall model accuracy is unknown since no averaging between pixels can be performed. WATClass model data at this resolution is not available for the purposes of this thesis; therefore, this aspect of the project is recommended for future study.

As previously stated in Section 2.4, in the Kalman filter, the observations are noisy measurements of the state. If the radar observations at time  $k$  are  $\{z_{1k}, z_{2k}, \dots, z_{pk}\}$  and the soil moisture value of the grid cell is  $x_k$ , then

$$z_{jk} = Hx_k + v_{jk}, \quad (\text{Equation 4-1})$$

For simplicity, the measurements are the soil moisture estimates derived from radar backscatter values, such that

$$H = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{p \times 1} \quad (\text{Equation 4-2})$$

$$v_{jk} \sim N(0, R)$$

$$R = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1p}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2p}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}^2 & \sigma_{p2}^2 & \cdots & \sigma_{pp}^2 \end{bmatrix} \quad (\text{Equation 4-3})$$

where  $\sigma_{ii}^2$  is the expected variance between the observation,  $z_i$ , and the state,  $x$ , and,  
 $\sigma_{ij}^2$  is the expected error covariance between  $z_i$  and  $x$ , and,  $z_j$  and  $x$

That is, the remotely sensed measurements at the finer scale are simply noisy observations of the average soil moisture value at the coarser scale. The elements of the measurement covariance matrix will be functions of the inherent measurement error in the inversion of the backscatter coefficients, the spatial variability of the soil moisture measurements at the finer scale and the variance of the soil moisture measurements at the coarser scale.

It is proposed that the variances of the measurement error, i.e. the diagonal components of the error covariance matrix, can be estimated by the sum of the estimated variance between of the soil moisture measurements at the fine scale and the coarse scale and the estimated measurement error. The estimated variance of the difference between the observation,  $z_i$ , and the state,  $x$ , is estimated as

$$\sigma_{ii}^2 = \sigma_{x,zi}^2 = (1 - \rho_{x,zi}) \cdot \sigma_x^2, \quad (\text{Equation 4-4})$$

where  $\sigma_{x,zi}^2$  is the expected variance between the observation,  $z_i$ , and the state,  
 $\sigma_x^2$  is the variance in the state, and,  
 $\rho_{x,zi}$  is the correlation coefficient between  $x$  and  $z_i$ .

The correlation coefficient is a measure of the correlation between two variables. The correlation coefficient as well as the calculation of the sample correlation coefficient is described in Appendix B.1. A correlation coefficient equal to 1 indicates that the variation in state and the observation are completely correlated; that is variations in the estimate are completely correlated with the observation. Then the error variance of that observation can be assumed to be entirely attributed to the measurement error. A correlation coefficient equal 0 indicates that the variables are independent, therefore, the variance in the error of that measurement will be equal to the sum of the variance of the state and the error in the measurement. This formula is an approximate method to estimate the error between the measurement and the state.

The off-diagonal components of the measurement error covariance matrix indicate the covariance of the noise in the measurement. The values of the covariances will also be functions of the spatial variability of the soil moisture values at the fine scale. It is proposed that the value of these components be the expected covariance of the two measurements calculated as the product of the expected correlation coefficient between the two measurements and the variance of the state.

$$\sigma_{ij}^2 = \rho_{zi,zj} \sigma_x^2 \quad (\text{Equation 4-5})$$

When these off-diagonal components are low, then the correlation between the observations is low, resulting in the observations containing more information about the state. Therefore, the Kalman gain equation will use more of the observations as it forms the a posteriori estimate. If the off-diagonal components are higher, then the observations are more correlated, resulting in less information from the observations. Therefore, the Kalman filter will use less of the observations in forming the a posteriori estimate.

To determine the correlation coefficients for use in the Kalman filter, the sample correlation coefficients between all pairs of fields in the Grand River watershed dataset and the sample correlation coefficients between all pairs of fields in the Roseau River watershed are calculated. An exponential regression is used to develop a relationship between distance between fields and the correlation coefficient. Then to calculate the covariance between two measurements, the correlation coefficient can be calculated to substituting the distance between the two fields as the distance. Since there are insufficient fields to correctly calculate the average soil moisture for the state, i.e. the entire grid squares, covariance between the measurement and the observation is taken from the correlation coefficient equation using the average distance between the observation and the state, whose area is much larger than the resolution of the radar image. The average distance calculation is explained in Appendix B.3.

Similarly, when upscaling a soil moisture measurement, spatial variability is also pertinent. Regardless of the spatial variability, the finer scale's estimate will simply be the coarser scale's

estimate. If, soil moisture values at the finer scale do not vary much within the area covered by a single pixel of the coarser scale, then the finer scale's estimate will have a high degree of confidence. However, if the soil moisture values are highly variable at the finer scale within the area covered by the coarser scale, then estimated error of the finer scale's estimate will be higher.

Since no passive data is available, the correlation between WATClass model and the imagery from passive radiometers is not established. If this data were available, the correlation coefficient or covariance between the model and the image as a function of the position of the WATClass grid square inside the passive microwave pixel would be established. Theoretically, grid squares in the middle of the pixel would have a higher degree of correlation; however, the difference in correlation might be too low, resulting in a single correlation coefficient or covariance between the grid square soil moisture value and the passive microwave value.

### **4.3 Soil moisture assimilation**

For this thesis, the WATClass model is simulated so that error levels for all parameters and truth-values are known and can be varied. The purpose of the simulation is to determine how much the soil moisture estimates can be improved by assimilating soil moisture observations using the ensemble Kalman filter. Only the soil moisture aspect of WATClass is simulated; streamflow is not part of the simulation. The simulation model uses a discrete time unit, i.e. a day, since soil moisture observations are likely to occur less frequently than daily.

While ground samples for the WATClass data were taken, these gravimetric measurements cannot be considered to be completely representative of the soil moisture value for a WATClass grid square. A WATClass soil moisture value is an abstract value, whose value can only be approximated by taking the average soil moisture value over the entire grid square. The optimal value of this variable is the soil moisture value that produces the best outputs, e.g. heat retention and surface

runoff, which are measurable. Therefore, a few ground samples can not accurately estimate the soil moisture value of the grid square.

WATClass's model of soil moisture is approximated by API. That is, in absence of precipitation, the soil moisture can be calculated as some fraction of the previous day's soil moisture:

$$\theta(t) = \alpha(t) \cdot \theta(t-1) \quad (\text{Equation 4-6})$$

where  $\theta(t)$  is the soil moisture estimate for day  $t$ , and  $\alpha(t) \sim N(\mu_\alpha, \sigma_\alpha^2)$  is the drying factor of the soil on day  $t$ .

The drying factor varies from day to day in the simulation. The mean and variance of the drying factor are estimated based on the Grand River watershed data. The API-based drying takes the place of the drainage, interflow and evapotranspiration processes. The unit for the soil moisture in the simulation is percent saturation, where  $\theta = 0$  represents no saturation and  $\theta = 1$  represents saturated soil.

Rainfall is produced randomly using a simple Markov Chain to determine which days rain occurs and which days rain does not occur. That is, the probability that it will rain today is dependent on whether it rained yesterday. If it rained yesterday, then the probability that it will rain today is defined as

$$P_{\text{rain,rain}} = P(x(t) = \text{rain} | x(t-1) = \text{rain}). \quad (\text{Equation 4-7})$$

Conversely, if it did not rain yesterday, then the probability that it will rain today is defined as

$$P_{\text{no rain,rain}} = P(x(t) = \text{rain} | x(t-1) = \text{no rain}). \quad (\text{Equation 4-8})$$

The values of these probabilities are estimated from the Grand River watershed WATClass data. Generally, these probabilities vary depending on the time of year; it tends to rain more in spring than it does in the summer or fall.

On rainy days, the amount of rainfall for the model is determined randomly by selecting a value from a gamma distribution (Selvalingam, 1978, Coe, 1982). The parameters of the gamma distribution are estimated using the Grand River WATClass data. The amount of rainfall value in the model represents the increase in soil moisture associated with the rainfall, not the amount of rainfall

in millimetres. This variable can be extracted from the WATClass data and is practical for the simulation. Rainfall data is not included in the dataset. If the rain causes the soil moisture to increase past a threshold, the saturation level, then the soil moisture is capped at the saturation level. The saturation level is not time dependent.

Therefore, the soil moisture during a rainfall event can be described as:

$$\begin{aligned} \theta'(t) &\equiv \alpha_d(x(t-1)) \cdot \theta(t-1) + r(d), \\ \theta(t) &= \begin{cases} \theta'(t) & \text{if } \theta'(t) < \theta_{\max}(t) \\ \theta_{\max}(t) & \text{if } \theta'(t) \geq \theta_{\max}(t), \end{cases} \end{aligned} \quad (\text{Equation 4-9})$$

where  $\theta(t)$  is the soil moisture on day  $t$ ,  
 $\theta'(t)$  is the soil moisture on day  $t$  if there is no saturation level,  
 $\alpha_d(t) \sim N(\mu_\alpha, \sigma_\alpha^2)$  is the drying factor,  
 $r(t)$  is the rain on day  $t$ , and  
 $\theta_{\max}(t)$  is the saturation level on day  $t$ .

The surface runoff level in the simulation is assumed to be the excess rain that is not absorbed by the soil,

$$\theta_{\text{runoff}}(t) = \theta'(t) - \theta(t). \quad (\text{Equation 4-10})$$

This runoff level has the same unit as rainfall, i.e. percent saturation of soil moisture.

Error in the model is simulated as three separate entities: the error in the drying factor,  $\alpha$ , the error in the rainfall,  $r$ , and the error in the saturation level,  $\theta_{\max}$ . The error in the drying factor of the simulation is related to errors in WATClass's hydraulic conductivity of the soil, meteorological conditions that affect exfiltration such as temperature, humidity and incident solar radiation, and the transpiration rate of the vegetation. The error in the saturation levels of the simulation is related to porosity due to soil composition.

In addition to the "true" values of these three parameters, "estimated" values are produced by selecting a value from a normal distribution with mean equal to the true value and variance equal to the error variance. These estimated parameters are essentially noisy versions of the true parameters. The estimated error level for each of the three parameters is also assigned a value. If the model is

perfect, the estimated error level is equal to the true error level; however, often the actual error level is unknown and must be estimated.

Observations are also simulated. The observations period – the number of days between observations – and the observation measurement error level are two of the major variables of interest in this thesis. Observations are modeled as direct observations of the soil moisture values of the entire squares. That is, the radar backscatter values are not being simulated, but rather, observations of abstract true values of the entire square are made. The observation is defined as

$$y(t) \sim N(\theta(t), \sigma_o^2) .$$

For the simulation, the observation is a single measurement of the soil moisture value.

This model, which simulates WATClass, is used as the model or predict component of the ensemble Kalman filter. An ensemble of soil moisture estimates is created from a random uniform distribution to create the initial soil moisture measurement. This ensemble is run through the model. For each sample in the ensemble, the drying factor, saturation level and rainfall level are assigned values from a normal distribution, with the mean and variance specified by the expected value and expected error variances in these three parameters. The ensemble of soil moisture is recalculated every day.

If there is an observation of a particular day, then the process error variance is estimated by calculating the variation in the model. The estimated process error variance and the estimated observation (or measurement) error variance are used to calculate the Kalman gain. The a posteriori estimate as well as the a posteriori error variance is then calculated. A new ensemble of soil moisture estimates is created by randomly selecting values from a normal distribution whose mean is the a posteriori estimate of soil moisture and the error variance is a posteriori estimated error variance.

Given this simulated data, the model can be run to determine how much improvement in soil moisture estimates can be made, given

- the frequency of soil moisture observations,
- the measurement error level in soil moisture observations,
- any bias in the soil moisture observations,
- the error in the estimated measurement error
- the model error level, which includes errors in the saturation level, drying factor and rainfall,
- the error in the model error levels, and,
- ensemble size.

This model is an abstraction of WATClass. The purpose of the simplification of the underlying equations and the change in discretisation in time for calculations is to reduce the run-time of the model, rendering the running of a large number of datasets more practical. Furthermore, this model affords better control over the inputs and model errors. While the model may be heavily simplified, the model is able to directly control the state, the error in the state, the errors in the inputs and the errors in the observation by simulating them.

It should be noted that the model errors inherent in WATClass are not incorporated into the model used in this thesis. That is, WATClass is a model of the hydrologic cycle and will therefore have errors caused by temporal and spatial discretisation and numerical approximations of the differential equations. This inherent model error is difficult to quantify; it is assumed to be negligible for the purposes of this thesis.

#### 4.4 Summary of proposed algorithm

Firstly, the parameters are defined.

```
AssignValueTo ProbRainIfRainYesterday  
AssignValueTo ProbRainIfNoRainYesterday  
AssignValueTo GammaDistShapeParametersForRainAmt  
AssignValueTo DryMean, DryStd `Drying constant distribution  
AssignValueTo ErrSat `Error in Saturation measurement  
AssignValueTo ErrRain `Error in rainfall measurement  
AssignValueTo ErrDry `Error in drying constant  
AssignValueTo ObsFreq `Time between Observations
```

```
AssignValueTo ObsErr `Error in observations
```

For the sake of the simulation, before running the ensemble Kalman filter, the “truth” values of the state and input parameters, the “measured” values of the input parameters, and, the observations are simulated. These values are:

1. The days during which rain occurs are selected using a Markov chain.

```
RainyDays(1) = NoRain
For Day := 2 to NumDays
  RainyDays(Day) = NoRain `by default
  If RainyDays(Day-1) = Rained Then
    RainyDays(Day) := Rained if RandUniform(0,1) > ProbRainIfRainYesterday
  Else
    RainyDays(Day) := Rained if RandUniform(0,1) > ProbRainIfNoRainYesterday
```

2. The amount of rain on each of these rainy days is simulated by randomly assigning values taken from a gamma distribution.

```
For Day := 1 to NumDays
  If RainyDays(Day) = Rained Then
    TrueRain(Day) := RandGamma(GammaDistShapeParametersForRainAmt)
  Else
    TrueRain(Day) := 0
```

3. The drying constant (decrease in soil moisture due to drainage and evapotranspiration) for each time step is determined by randomly selecting a value from a Gaussian distribution.

```
For Day:= 1 to NumDays
  TrueDry(Day) := RandGaus(DryMean, DryStd)
```

4. The saturation level (time-independent) for the grid cell is randomly assigned a value from a Gaussian distribution.

```
TrueSatLevel := RandGaus(1, ErrSat)
```

5. The initial “truth” soil moisture value is assigned a value selected from a uniform distribution ranging from 0 to 1.

```
TrueSoil(1) := RandUniform(0,1)
```

6. The “truth” soil moisture values for each time step are simulated using Equation 4-8.

```
For Day := 2 to NumDays
  TrueSoil(Day) := TrueSoil(Day - 1) * TrueDryConst(Day) + TrueRain(Day)
  If TrueSoil(Day) > TrueSatLevel
    TrueSoil(Day) := TrueSatLevel
```

7. Measured values of the rain and drying constants are simulated by selecting values from Gaussian distributions with mean equal to the “truth” values and standard deviations equal to the expected error in the parameters.

```
For Day := 1 to NumDays
  EstRain(Day) = RandGaus(TrueRain(Day), ErrRain)
  EstDryC(Day) = RandGaus(TrueDry(Day), ErrDry)
```

8. The observations are simulated by selecting values from Gaussian distributions with mean equal to the “truth” soil moisture values and standard deviation equal to the expected observation error.

```
For Day := ObsFreq to NumDays Step ObsFreq
  Obs(Day) := RandGaus(TrueSoil(Day), ErrObs)
```

The ensemble Kalman filter is implemented into the simulation as follows. The state variable is soil moisture of a single grid cell.

1. The estimated errors of the input parameters and the observation measurements are established. These error levels are not necessarily equal to the expected errors used in the simulation of the input parameters. When running the algorithm in WATClass, these error levels are estimated from real world data.

```
AssignValueTo [EstErrDecay] `Error in Drying decay constant error
AssignValueTo [EstErrSat] `Error in Saturation level
AssignValueTo [EstErrRain] `Error in Rain
AssignValueTo [EstErrObs] `Error in observation measurement
```

2. The initial ensemble is determined by randomly selecting  $n$  points from a uniform distribution ranging from 0 (dry soil) to 1 (saturation). The initial estimated soil moisture value is set as 0.5.

```
For Replicate := 1 to EnsembleSize
  EnsSoil(Replicate) := RandUniform(0,1)
EstSoil(1) := 0.5
```

3. At each time step:

- a. The estimated soil moisture value is propagated through the model using Equation 4.8.

```
EstSoil(Day) := EstSoil(Day - 1) * EstDecay(Day) + EstRain(Day)
If EstSoil(Day) > 1 then `Saturation
  EstSoil(Day) := 1
```

- b. The time-dependent parameters (rain, drying constant and saturation level) are assigned a separate value for each point in the ensemble. The values are randomly selected from Gaussian distributions with mean equal to the estimated values of the parameter and standard deviation equal to the expected error of the parameter.

```
For Replicate := 1 to EnsembleSize
  EnsRain(Replicate) := RandGaus(EstRain(Day), EstErrRain)
  EnsDry(Replicate) := RandGaus(EstDry(Day), EstErrDry)
  EnsSat(Replicate) := RandGaus(1, EstErrSat)
```

- c. The each point in the ensemble is propagated through the model using Equation 4-8.

```
For Repl := 1 to EnsembleSize
  EnsSoil(Repl) := EnsSoil(Repl) * EnsDry(Repl) + EnsRain(Repl)
  If EnsSoil(Repl) > EnsSat(Repl)
    EnsSoil(Repl) := EnsSat(Repl)
```

d. If an observation of the state occurs at this time step then

- i. The estimated a priori estimated error is found by calculating the standard deviation of the ensemble.

```
ErrSoilAPriori := StDev(EnsSoil)
```

- ii. Using the model's soil moisture estimate as the a priori state, the a posteriori estimate is determined from Equations 2-7, 2-8, 4-2 and 4-3 (i.e. Kalman update equations)

```
K := ErrSoilAPriori / (ErrSoilAPriori + EstErrObs)  
EstSoil(Day) := EstSoil(Day) + K * (Obs(Day) - EstSoil(Day))
```

- iii. The a posteriori estimated error is determined from Equation 2-9.

```
ErrSoilAPost := (1 - K) * ErrSoilAPriori
```

- iv. A new ensemble is created by selecting points randomly from a Gaussian distribution with mean equal to the a posteriori estimate of the state and standard deviation equal to the a posteriori estimated error.

```
For Repl := 1 to EnsembleSize  
  EnsSoil(Repl) = RandGaus(EstSoil(Day), ErrSoilAPost)
```

## Chapter 5

### Results

This section presents the findings from the examination of the data. Section 5.1 examines the spatial correlation between the soil moisture estimates for each square in a watershed in WATClass as well as the correlation between the soil moisture measurements sampled from the fields studied in the Grand River Watershed and the Roseau River Watershed datasets. These findings are used to develop a general correlation function between the expected soil moisture levels of two locations in a watershed as a function of the distance between the two fields.

Section 5.2 presents the findings from the implementation of the ensemble Kalman filter into the hydrological model. Firstly, the derivation of the parameters from the data is presented. The effect of the ensemble Kalman filter on the error in soil moisture is then examined as a function of

- Ensemble size
- Observation frequency
- Measurement error level
- Model error level
  - Error in saturation level
  - Error in drying decay constant
  - Error in rain measurement

#### 5.1 Spatial variation of soil moisture content

This section examines the spatial variation of the soil moisture values in the different datasets used in the thesis. Section 5.1.1 describes the spatial variation of the WATClass modeled soil moisture estimates. These soil moisture estimates are a function of the meteorological data and the land cover data. Furthermore, since WATClass is gridded, the soil moisture estimates are regularly spaced and cover the entire watershed.

Section 5.1.2 presents the spatial variation of the soil moisture values sampled from the fields in the Grand River Watershed study and the Roseau River Watershed study. These soil moisture values are actual real-world measurements. They are irregularly spaced and, in the case of the Grand River Watershed data, only cover a small portion of the watershed.

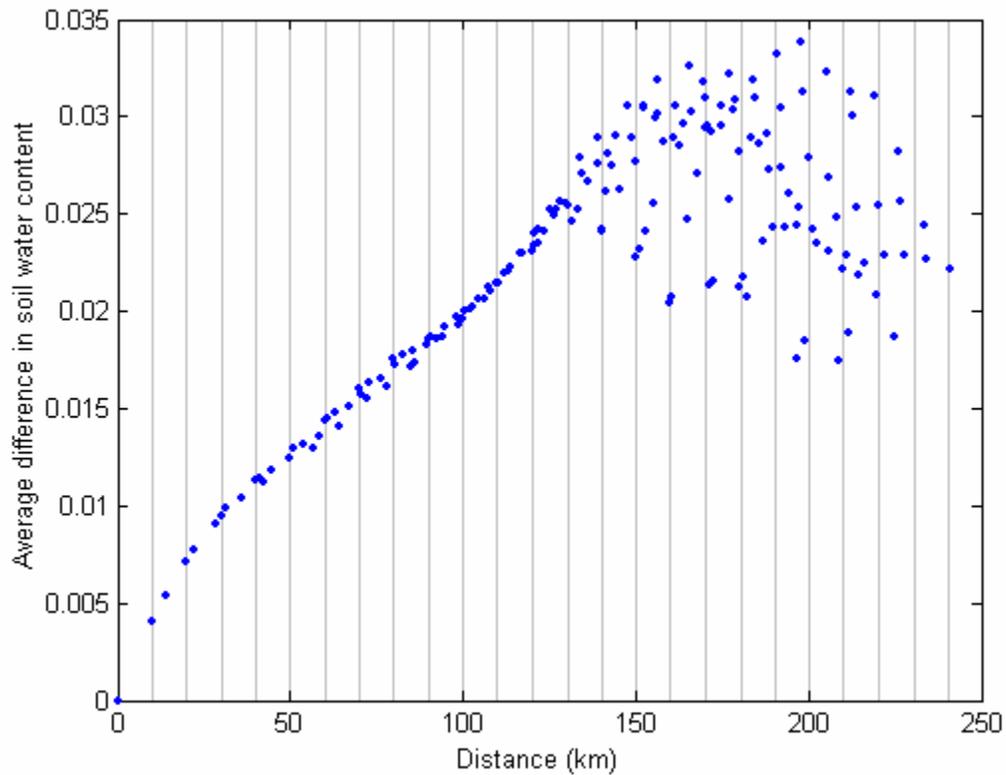
### **5.1.1 Spatial variation of soil moisture content in WATClass**

The WATClass data includes meteorological data for each grid square for each hour in the study period (Seglenieks, 2001). The meteorological data includes

- temperature,
- humidity,
- long and short wave radiation,
- pressure,
- wind speed, and,
- precipitation.

The proportions of each land class in each grid square are also included. However, the soil properties are assumed to be homogeneous for each type of land class across all grid squares. The only other parameter affecting soil moisture is internal slope of the grid square, whose value is included for each grid square. Therefore, when the WATClass model is run, the spatial variation in soil moisture values is almost entirely dependent on meteorological conditions.

Figure 9 shows the variogram for the WATClass estimated soil moisture values for the Grand River watershed for the summer months of 1996 and 1997. The variogram is produced by finding the difference in soil moisture value between all pairs of grid squares and then calculating the average difference as a function of distance between fields. For this variogram, spatial variation is considered to be isotropic; therefore, distance direction is irrelevant.



**Figure 9 – Variogram for WATClass estimated volumetric water content of the Grand River watershed during the summer months of 1996 and 1997**

At distances less than 120 km, the average difference in volumetric soil moisture is linearly dependent on distance. The cause of the linear dependency is the method by which the meteorological conditions for each grid square are created. If the meteorological conditions are collected only at particular points in the watersheds such as meteorological stations, WATClass linearly interpolates between these points to create meteorological condition estimates for each grid square (Kouwen, 2006). Since the variability of soil moisture estimates in WATClass is almost entirely dependent on meteorological conditions, the soil moisture estimates also appear linearly interpolated. Therefore, a linear relationship between average difference in soil moisture and distance between grid squares occurs.

At distances greater than 120 km, the squares begin to be far enough apart that the two squares no longer lie between two meteorological stations. Therefore, the difference in soil water content is no longer linearly related.

WATClass is capable of using ground radar as an input to determine rainfall levels for each grid square (Kouwen, 2006). In this case, the radar rainfall levels are calibrated using the rainfall gauges located at the meteorological stations. For the Grand River watershed dataset, however, radar rainfall does not appear to have been used. Note that all other meteorological inputs to WATClass are interpolated from the point measurements made from meteorological stations. Since the soil moisture measurements are only derived from a few meteorological stations, this dataset does not provide relevant information on the correlation between soil moisture measurements at different locations.

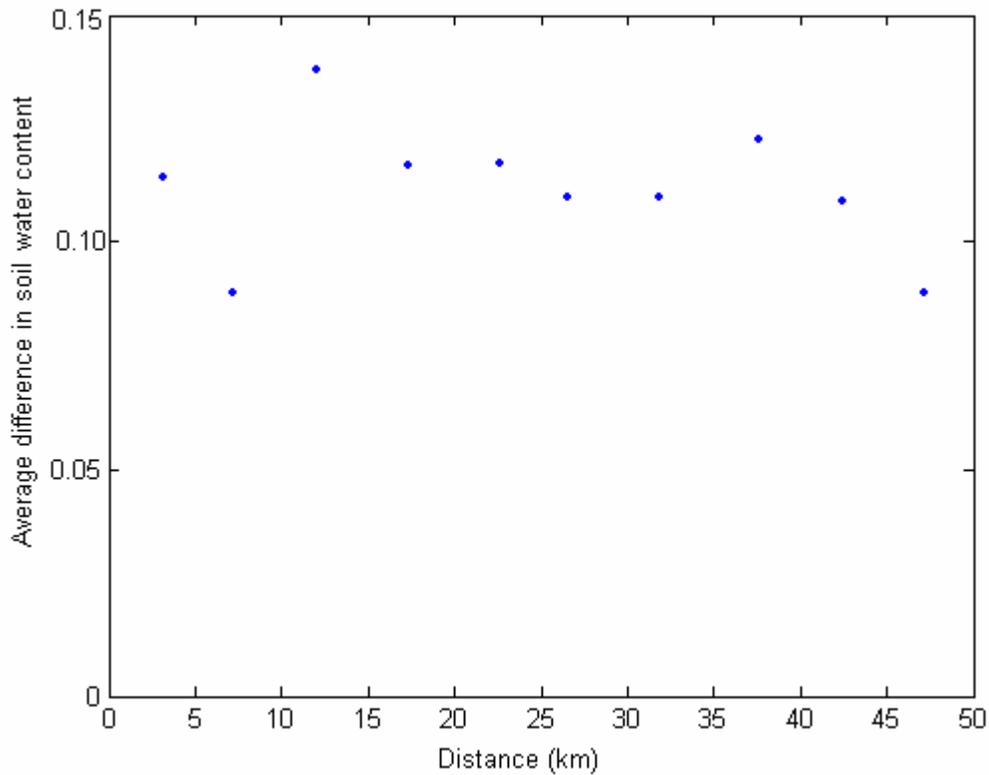
### **5.1.2 Spatial variation of soil moisture content in ground data**

The soil moisture measurements from the Grand River Watershed study and the Roseau River Watershed study represent true ground measurements and thus provide a better representation of the relationship between distance between fields and correlation of soil moisture estimates of the two fields.

#### **5.1.2.1 Grand River watershed data**

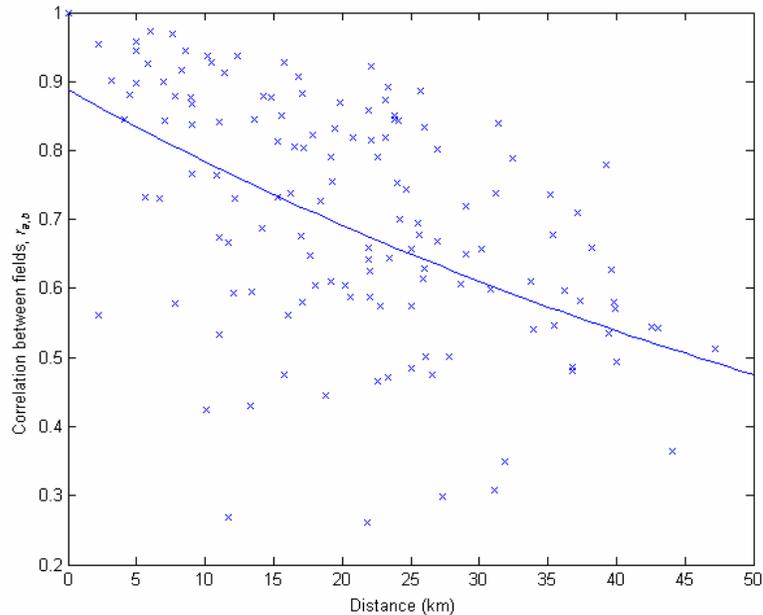
The variogram computed for the Grand River watershed ground sampled data is shown in Figure 10. For this variogram, pairs are grouped by distance and the average difference is computed for each group. The distance between fields assigned to each average difference is also the average distance between the pairs of fields in that group. The average difference in Grand River watershed shows little dependency on distance. This is due to the large variability in field capacity between fields, as shown in Table 2. A large variation in field capacity would cause the fields to have different soil water content values even if the rain level is constant over all fields.

Figure 11 displays the correlation coefficients between fields in the Grand River dataset as a function of distance. The correlation coefficient displays a better representation of how closely the fields correlate, since the correlation coefficient examines the relative changes in soil moisture measurements and how synchronized these changes are between fields. That is, the correlation coefficient describes how in tune the temporal changes in soil moisture for one field are with those of another field. The solid line is an estimate of the correlation coefficient as a function of distance produced from a logarithmic regression of the samples ( $r_{a,b} = 0.889 \cdot e^{-0.0125d}$ ). As distance between fields increases, the correlation between field decreases. The y-intercept of the line is less than 1, indicating that even the soil moisture values of two fields that are close to one another might not be highly correlated with one other.

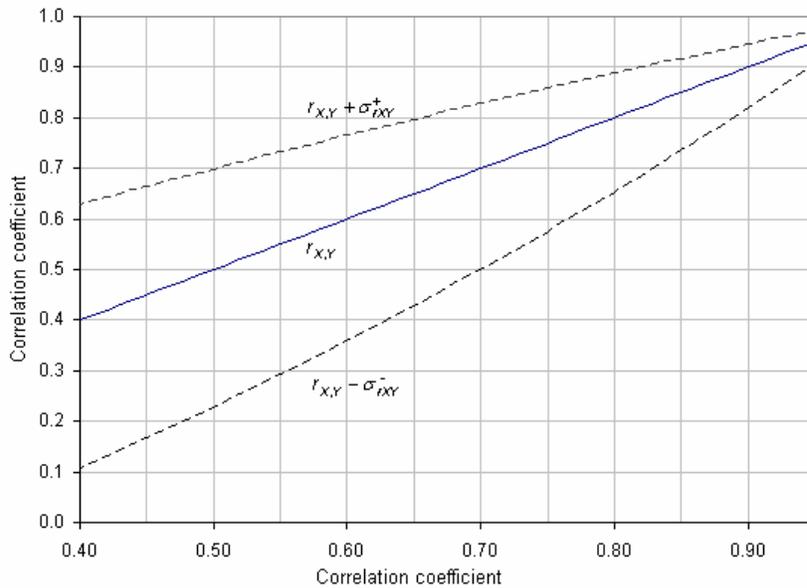


**Figure 10 – Variogram for Grand River watershed ground-sampled soil moisture content**

The small sample size for calculating the correlation coefficients – thirteen dates are used for each correlation coefficient between fields – leads to a high standard error for the correlation coefficients. This plot is shown in Figure 12 shows the standard error as a function of correlation coefficient. The standard error is much higher for lower correlation coefficients. Subsequently, the decay constant for the correlation coefficient will have a higher error. The root mean square of the residuals of the regression, which is the expected error in the calculation of the correlation coefficient from the sampled correlation coefficients, is 0.146. Combining these two errors, the expected error in the estimated correlation coefficient between two points given a distance is very high for the correlation coefficient function derived from this dataset.



**Figure 11 – Correlation of soil water content between fields as a function of distance in the Grand River dataset**



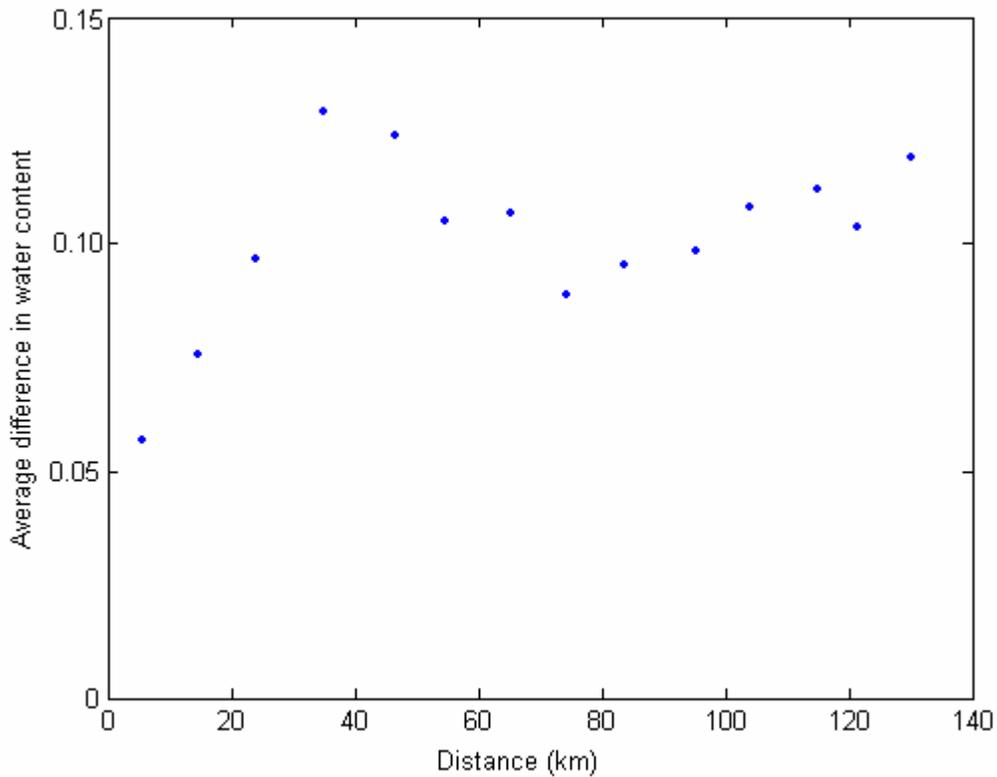
**Figure 12 – Standard error for correlation of soil water content between fields in the Grand River dataset (N = 13)**

#### 5.1.2.2 Roseau River watershed data

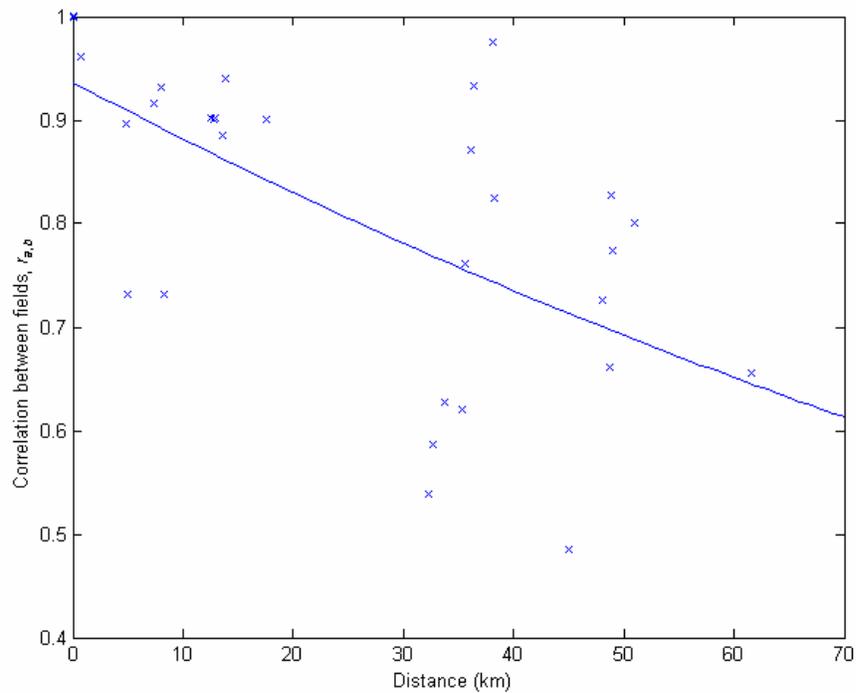
Figure 13 presents the variogram for Roseau River dataset. The average difference in soil moisture measurements increase linearly in distances under 40 km. Above 40 km, the average difference remains constant.

Figure 14 presents the correlation coefficient as a function of distance. The equation of the correlation coefficient line is  $r_{a,b} = 0.937e^{-0.00606d}$ . Due to the sparseness of the measurement matrix, only part of the Roseau River dataset is used to create the correlation estimates; only eight fields from seven dates are used, creating only sixty-four samples for the regression. Given that the number of samples in the calculation of the correlation coefficient is seven, the standard error level is high. The standard error as a function of correlation coefficient is shown in Figure 15. The root mean square of the residual of the regression between distance and correlation coefficient is 0.113. As with the Grand River dataset, the high level of these two errors results in a high expected error in the estimated correlation coefficient between two points given a distance from the correlation coefficient function derived from the Roseau River dataset.

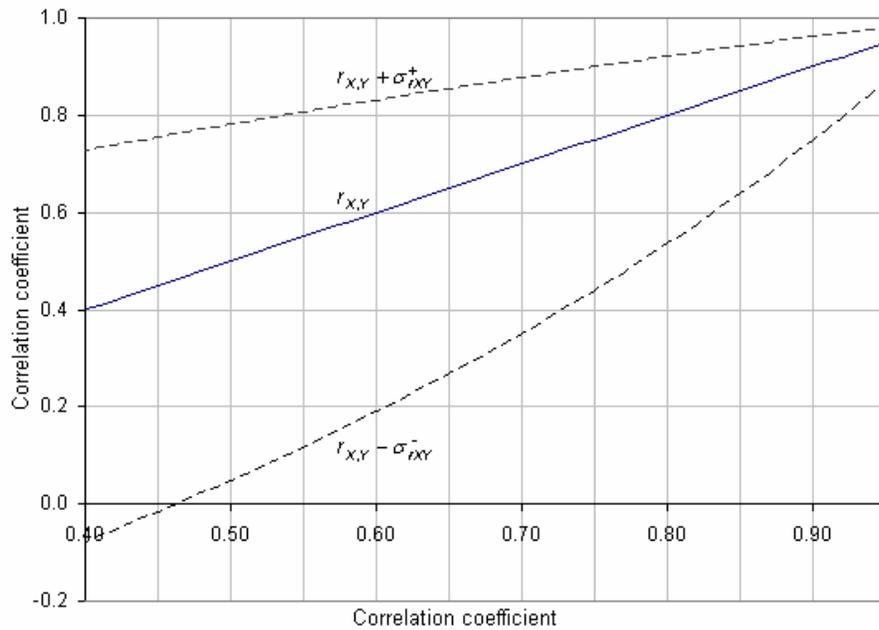
For each field on each day for Roseau River ground sampling, volumetric water content was measured at five measurement sites spaced at least 5 m apart. At each measurement site, three samples were taken within a 1 m radius. Therefore, a total of fifteen measurements were taken at each field on each day. Though the samples were all taken from the same field within 50 m of each other, a high degree of variation is apparent in the measurements. Table 4 displays the standard deviation of the fifteen measurements as a fraction of the mean of the measurements. The average standard deviation is 26.5% of the measurement mean. This is an indicator of the high spatial variability of soil moisture within a field.



**Figure 13 – Variogram for Roseau River watershed ground-sample soil water content**



**Figure 14 – Correlation of soil water content between fields as a function of distance in Roseau River dataset**



**Figure 15 – Standard error for correlation of soil water content between fields in the Roseau River dataset (N = 7)**

**Table 4 – Soil moisture estimate standard deviation normalised by mean for Roseau River dataset**

Field name	Sep 30, 02	Oct 5, 02	Oct 24, 02	Oct 27, 02	Apr 10, 03	Apr 14, 03	Apr 16, 03	May 3, 03	May 4, 03	May 7, 03
A1_FOR_1	0.149		0.368	0.200	0.175					
A1_GLE_1	0.529		0.257	0.179	0.203					
A1_GLE_2	0.521		0.359	0.197	0.224				0.450	
A1_KIR_2	0.313		0.189		0.092				0.191	
A1_SAB_2	0.314		0.244	0.084	0.167					
A1_TER_1	0.220		0.575	0.161	0.302				0.122	
A2_DYC_1	0.161	0.469	0.537		0.232			0.269	0.188	0.184
A2_GRI_1	0.497	0.521	0.276	0.242	0.133			0.128		
A2_HIR_1	0.154	0.512	0.432		0.124	0.184	0.240	0.209	0.197	0.105
A2_LOE_1	0.257	0.162	0.260	0.409	0.211			0.316	0.105	0.147
A2_NED_1	0.192	0.162	0.153		0.161			0.247		
A2_PAL_1	0.175	0.473		0.392	0.105			0.212	0.165	
A2_REI_1	0.260	0.480	0.116		0.231			0.241		
A2_SMO_1	0.042	0.183	0.402	0.523	0.323			0.298		0.203
A3_ECK_1	0.203	0.202	0.224		0.172	0.139	0.201	0.192	0.156	
A3_EWA_2	0.231	0.260	0.413		0.081	0.169	0.114	0.155		
A3_GOB_2	0.248	0.209	0.862		0.112	0.076	0.103	0.197		
A3_GOT_1	0.224	0.294	0.722		0.190	0.144	0.198	0.277	0.156	
A3_MCC_2	0.366		0.766	0.723	0.247	0.155	0.213	0.264	0.136	0.209
A3_NOR_1	0.162	0.333	0.377	0.485	0.171	0.149	0.212	0.136	0.236	0.124
A3_TKA_1	0.367	0.275	0.722		0.208	0.341	0.270	0.233		0.299
A3_WAN_1	0.090	0.200	0.350	0.431	0.326	0.299	0.388	0.559	0.271	0.221

### 5.1.3 Producing soil moisture measurement estimates

Due to the high variability of the soil moisture estimates, the error in the correlation coefficient functions derived in the previous sections are too high to be useful in the determining of the elements of the Kalman measurement error covariance matrix. Furthermore, the data in the Grand River dataset is too sparse both spatially and temporally to be assimilated into WATClass. For example, there are only a few fields sampled in a grid square, typically one or two, and the variance between the measurements of the fields is very high. Therefore, for the testing of the algorithm, simulated data is used instead of the real world data. The simulated data is described in Section 4.3.

If more dates are sampled, then the expected error in the correlation coefficients would reduce, making the result of the regression more valid, allowing the function to be used in conjunction with the ensemble Kalman filter. To better quantify the variance in the measurement error, the correlation coefficients could be another parameter randomised as part of the ensemble. The mean of the correlation would be the value derived from the regressed equation and the standard deviation of the error would be the root-mean-square error of the regression. If there are a sufficient number of samples used in the regression, the RMS error could also be a function of distance.

## **5.2 Ensemble Kalman filter implementation**

### **5.2.1 Parameter selection**

Parameters and variables for the simulation of WATClass are described in Table 5.

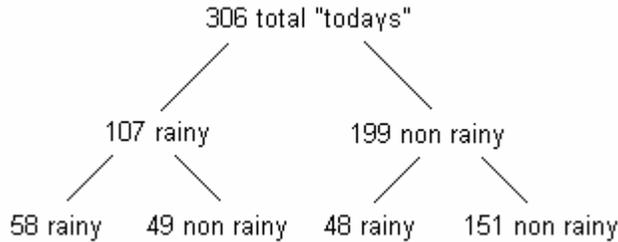
The rain parameters are estimated from the Grand River watershed WATClass data. The WATClass model is run for a single square where the soil moisture estimate is zeroed four times per day. Therefore, any increase in the soil moisture estimate in any of the four periods per day would indicate that the precipitation occurred during that day. The precipitation level for each day is determined by summing the maximum soil moisture estimate in each period. This rain estimate will underestimate the actual value because drainage and evapotranspiration will cause the soil moisture estimates to drop during the period. For example, if the rate of loss of water content due to drainage and transpiration at a given time is equal to the precipitation rate, then this rain estimation method will determine that no precipitation had occurred. Therefore, the precipitation parameters are modified slightly to create slightly higher precipitation level.

These “rain” measurements are used because they indicate the increase in soil moisture due to the rain on a given day, rather than the amount of precipitation falling on the ground. Therefore, these values are more suitable for the simulation.

**Table 5 – Simulation parameters**

Function	Parameter	Value	Explanation
<i>Simulation parameters</i>			
Rain simulation	$P_{rain,rain}$	0.55	Probabilities that it will rain today given that either it rained and did not rain yesterday
	$P_{no\ rain,rain}$	0.24	
	$\alpha_{rain}$	2	Shape and scale parameters for the gamma distribution, which determines the amount of rain on rainy days.
	$\beta_{rain}$	6	
Soil moisture	$\mu_{\alpha}$	0.85	Parameters for the normal distribution for $\alpha$ , the decay constant.
	$\sigma_{\alpha}$	0.05	
	$\theta_0$	0-1	Initial soil moisture value at the beginning of the simulation
	$\mu_{\theta_{max}}$	1	Parameters for the normal distribution for $\theta_{max}$ , the soil moisture value at saturation
	$\sigma_{\theta_{max}}$	0.1	
<i>Simulation variables</i>			
Measurements	$T_{obs}$		Number of days between observations
	$\sigma_{eobs}$		Observation error level (% error)
Rain	$\sigma_{erain}$		Rain measurement error level (% error)
Soil moisture	$\sigma_{edecay}$		Error level of the decay constant (% error)
	$\sigma_{esat}$		Error level of the soil moisture saturation level (% error)

The precipitation is determined for the summer months only, i.e. when WATClass determines that there is no soil water moisture, i.e. ice, in the soil. In 1996, there were 170 days between the last occurrence of solid soil moisture in the spring and the first occurrence of solid soil moisture in the fall. In 1997, there were 138 days of this nature. These days are divided into two categories: days in which it rained and days in which it did not rain. The dependency of the occurrence of rain on one day on the occurrence of rain on the next day is determined. A summary of the findings is as follows:



The rain probabilities are determined from the sample.

The gamma distribution parameters for the amount of rain on rainy days are determined by maximum likelihood estimation. If there are  $N$  sample points  $\{x_1, \dots, x_N\}$  to fit to the gamma distribution

$$g(x; \alpha, \beta) = x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)}, \text{ for } x > 0. \quad (\text{Equation 5-1})$$

The parameters,  $\alpha$  and  $\beta$  are estimated numerically (Choi, 1969).

The parameters for the API drying factor are taken from ranges specified in the literature (Dingman, 2002, Kouwen, 2006).

Since no information is known about the initial soil moisture, the initial soil moisture for the simulation is set to be a random variable selected from a uniform distribution between 0 and 1. The initial a priori estimate for soil moisture is thus 0.5.

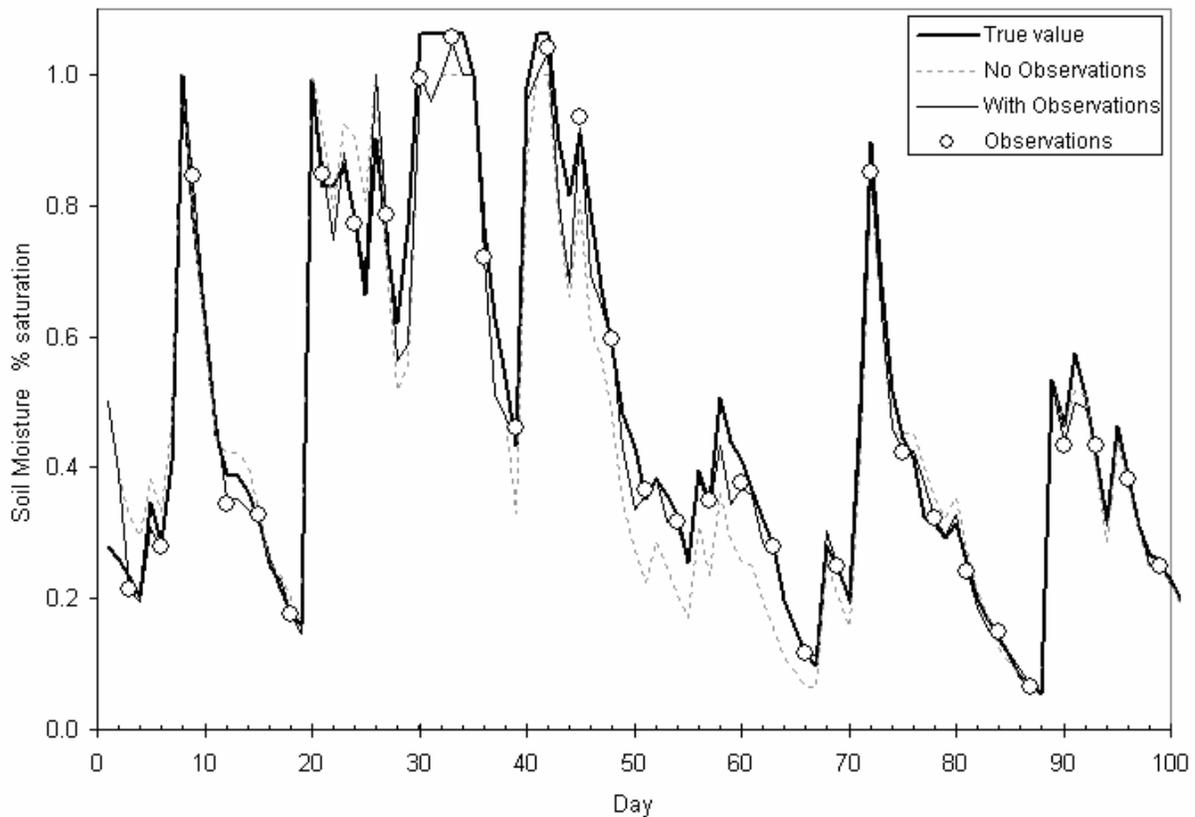
The mean soil moisture value at saturation is set at 1, as WATClass assumes all soils of the same land type have the same saturation point. The error in saturation is arbitrarily chosen to be 0.1. The variability in soil moisture saturation levels is higher, as shown in the variation in field capacity of the Grand River dataset as evident in Table 2, however, the WATClass saturation level is representative of a large area; therefore, averaging lowers the variation.

### 5.2.2 Simulation

For each test, the simulation is executed with nominal values used for the variables not under study. The standard deviations,  $\sigma$ , are error levels and are normalised by the variable value such that they are percentage errors, not absolute errors. These values are shown in Table 6.

**Table 6 – Nominal variable values for initial simulation**

Variable	Value
$T_{\text{obs}}$	3 days
$\sigma_{\text{eobs}}$	10%
$\sigma_{\text{erain}}$	5%
$\sigma_{\text{edecay}}$	10%
$\sigma_{\text{esat}}$	10%



**Figure 16 – Soil moisture value and estimates for nominal values**

Several general comments can be made examining the outputs of the simulation using these nominal parameter values. Figure 16 shows the true soil moisture value, the estimated soil moisture value if no observations are assimilated (i.e. estimates based on meteorological and land class data only), the estimated soil moisture value with assimilated observations and the observations created by the simulation using the nominal values for the parameters. Using the nominal parameters, the absolute RMS error in soil moisture decreased to approximately 0.071 from 0.095. Note that the soil moisture values range from 0 for completely dry to approximately 1 for completely saturated. Qualitatively, the soil moisture estimates made with observations tracks the true soil moisture value better than those made without observations. However, the soil moisture estimates without observations do not vary much from the true soil moisture value. Since all the estimated parameters are randomly taken from probability distributions with means equal to the true value of the

parameters, the errors in the soil moisture tend to even out. Quantitative results are included in the sections that follow.

Furthermore, the absolute error for the estimates is generally lower for dry soils than wet soils. For example, if there is no rain for long periods of time, the estimated value of the soil moisture will tend to converge with the true value of the soil moisture as the soil moisture value decreases. The estimate without observation also tends to converge with the estimate with observation when precipitation causes the soil moisture to reach the saturation level, since the soil moisture value cannot exceed the saturation level. However, if the saturation level has an error, then both the estimate with observations and the estimate without observations will be the same, and both will have an error equal to the saturation level error.

### **5.2.3 Ensemble size**

The ensemble used in the simulation should be representative of the distribution of the estimated error. If the ensemble is adequately large, then simulation results are repeatable even though the ensemble points are selected at random. The ensemble should also not be indiscriminately large either since an increased ensemble size will lead to increased computation time.

For several ensemble sizes, the simulation is executed twenty-five times using exactly the same parameters and variables. The only differences between simulations are the points inside the ensemble, which could possibly affect the estimated model error variance. The standard deviation of the estimated error level is calculated for each day in the study period. Each daily standard deviation is normalised by the mean of the twenty-five estimated error levels. Then the daily standard deviations are divided by the daily means, producing the daily percent error of the estimated error level. The entire test is performed twenty times and the averages of the statistics of the estimated error for all these simulations are presented in Table 7.

**Table 7 – Error in soil moisture value (%saturation) as a function of several ensemble sizes**

<b>Size</b>	<b>Avg SD</b>	<b>Avg SD %</b>
<b>2</b>	0.0288	62.1%
<b>5</b>	0.0157	26.8%
<b>10</b>	0.0106	17.3%
<b>20</b>	0.00730	11.7%
<b>50</b>	0.00453	7.21%
<b>100</b>	0.00321	5.07%
<b>200</b>	0.00226	3.56%
<b>500</b>	0.00143	2.24%
<b>1000</b>	0.00100	1.58%
<b>2000</b>	0.000725	1.13%
<b>5000</b>	0.000449	0.707%
<b>10000</b>	0.000319	0.503%

Avg SD: Average standard deviation of the expected error over all dates  
Avg SD %: Percent error of expected error

As ensemble size increases, the error in the estimated error decreases. However, the error level decreases at a rate much slower than the ensemble size increases. An ensemble size of 1000 is used for the remainder of the tests, as simulations run with this ensemble size had manageable run time and the estimated error in the estimated error is low (<2%).

#### **5.2.4 Observation frequency**

The effect of observation frequency is also examined. Several time lengths for time between observations are chosen: 1, 2, 3, 7, 14, 24 and 35 days. The shortest three time periods represent the observation frequencies for passive microwave remote sensing of soil moisture, whereas the longer time periods represent the observation frequencies for active remote sensing. The error in the soil moisture estimates is found for five different sets of observations for each time period in one hundred different simulation datasets. The average error over all days and the a priori error estimate for days with observations for each period are shown in Table 8. The a priori error estimate is the estimated error as predicted by the ensemble Kalman filter just before the observation is assimilated.

**Table 8 – RMS error in soil moisture value (%saturation) as a function of time between measurements**

<b>Period</b>	<b>Total error</b>	<b>A priori error</b>
<b>1</b>	0.052	0.063
<b>2</b>	0.063	0.078
<b>3</b>	0.071	0.088
<b>7</b>	0.085	0.106
<b>14</b>	0.091	0.116
<b>24</b>	0.093	0.120
<b>35</b>	0.094	0.120
$\infty$	0.096	---

The error in soil moisture is approximately halved when observations are assimilated daily than when no observations are applied. Conversely, when observations are made once every thirty-five days, the improvement in soil moisture estimate accuracy is only 2% better than if no observations were made. The reason why only a slight increase is found is that once the observation is assimilated, the errors inherent in the model creep back into the soil moisture estimate. The a priori error estimate is within 10% for 14, 24 and 35 days, implying that the improvement of the soil moisture estimate is essentially lost approximately fourteen days after an observation is assimilated.

It should be noted that the error levels are very dependent on the model error levels and the measurement error levels. That is, if the measurement error is decreased, the total errors would decrease significantly for the shorter time periods. Whereas, if the model errors are decreased, the total errors for all time periods would decrease, resulting in less of an improvement of shorter time periods over longer time periods.

### **5.2.5 Measurement error**

The observation measurement error is the error in the conversion of the remote sensing images to the soil moisture estimates to be applied into WATClass as outlined in Section 0. Obviously, any increase in the accuracy of the observation will cause the soil moisture estimates from the ensemble Kalman filter to increase. Furthermore, assimilating soil moisture remotely sensed soil moisture

estimates into the hydrological system will not worsen the estimated error. Therefore, on average, the soil moisture estimates from the ensemble Kalman filter will be better than the soil moisture estimates than if the ensemble filter had not been applied, regardless of the level of error in the signal. Table 9 presents the actual error levels for several different observation measurement errors run through the simulation. The a priori and a posteriori estimated errors are also shown. The a priori estimate error is the average of the soil moisture estimate error on the day of the measurement prior to the assimilation of the measurement. Hence, this value represents the maximum soil moisture estimate error in the observation period. The a posteriori estimate is the average of the soil moisture estimate error on the day of the measurement immediately after the assimilation of the measurement. Hence, this value represents the minimum soil moisture estimate error in the observation period. The last row presents the estimate error if no observations are used.

The error in soil moisture estimates is also included if the soil moisture estimate is directly replaced by the observation. If the observation error is low, then the method of direct replacement performs just as well as the ensemble Kalman Filter. However, as the observation error increases, the ensemble Kalman Filter begins to significantly outperform direct replacement. The ensemble Kalman Filter is able to leverage the information from the model, whereas if the soil moisture estimate is simply replaced by the observation, all the information gained from the model is discarded.

The statement that the Kalman filter cannot worsen the soil moisture estimate is valid under the assumption that the measurement error level is accurate and the measurement error has a zero mean, i.e. the observation measurements are not generally higher or lower than the actual soil moisture value.

Table 10 shows the root-mean-square (RMS) error of the soil moisture over all the dates if the measurement error is erroneous. In the leftmost column, the estimated measurement error is shown as a fraction of the actual measurement error. Two measurement error levels, 10% and 20%, are tested.

**Table 9 – Error in soil moisture value(%saturation) as a function of estimated measurement error level**

<b>Observation error</b>	<b>RMS error</b>	<b>A priori est. error</b>	<b>A posteriori est. error</b>
<b>0.01</b>	0.058	0.082	0.006
<b>0.03</b>	0.060	0.083	0.017
<b>0.05</b>	0.063	0.084	0.028
<b>0.10</b>	0.071	0.087	0.048
<b>0.15</b>	0.077	0.090	0.062
<b>0.20</b>	0.081	0.093	0.071
<b>0.30</b>	0.087	0.097	0.083
<b>0.40</b>	0.090	0.099	0.090
<b>0.50</b>	0.093	0.101	0.094
<b>0.60</b>	0.094	0.102	0.097
<b>0.70</b>	0.094	0.103	0.099
$\infty$	0.095	---	---

For both error in measurement error levels, the smallest error in soil moisture occurs when the estimated measurement error is the actual measurement error, which is expected. For the smaller error in measurement error level ( $\sigma_{eobs}/\mu_{obs} = 10\%$ ), the soil moisture estimates remained more accurate than the estimates without observation, even when the measurement error levels are five times lower or five times higher than the actual error level. However, the actual measurement error level is much lower than the actual process error level. Therefore, even if the Kalman gain equation gravitates more to the observation, since the observation error is generally lower than the process error, an improvement in error is made. However, if the actual observation measurement error level is higher, then there is a chance that the observation has a higher error than the process. When the Kalman filter overly trusts the observation error, i.e. when the estimated measurement error level is too low, then the average error with the observations can increase over the average error if no observations are assimilated. If the estimated measurement error level is too high, then the result is that the Kalman filter does not trust the observation as much as it should, resulting in a loss of improvement over the non-observation case; however, it is not possible for the average error to increase above the error level if no observations are made.

Table 11 displays the average error in the soil moisture estimates if there is a bias in the observation, i.e. the observations tend to be higher (or lower) on average than the state that the state being observed. Bias was simulated by simply multiplying the observation value by a multiplier,  $x$ , prior to assimilating it. Therefore, the observation mean,  $\mu_z$ , will become

$$\mu_z = x\mu_\theta.$$

**Table 10 – Error in soil moisture (%saturation) as a function of error in measurement error level**

<b>Multiplier <math>x</math> (<math>\hat{\sigma}_{eobs} = x\sigma_{eobs}</math>)</b>	<b>RMS error (<math>\sigma_{eobs} = 0.05</math>)</b>	<b>RMS error (<math>\sigma_{eobs} = 0.10</math>)</b>	<b>RMS error (<math>\sigma_{eobs} = 0.20</math>)</b>
0.200	0.063	0.075	0.105
0.250	0.063	0.074	0.101
0.333	0.063	0.073	0.096
0.500	0.063	0.072	0.088
0.707	0.063	0.070	0.084
1.000	0.062	0.069	0.082
1.414	0.063	0.070	0.083
2.000	0.065	0.073	0.085
3.000	0.068	0.078	0.089
4.000	0.072	0.082	0.091
5.000	0.076	0.085	0.092
<i>No observations RMS error 0.095</i>			

Note:  $\sigma$  are actually  $\sigma/\mu$

**Table 11 – Error in soil moisture (%saturation) as a function of bias in the observation (%of true value)**

<b>Bias</b>	<b>RMS error</b>
0.5	0.194
0.6	0.161
0.7	0.129
0.8	0.101
0.9	0.079
1.0	0.070
1.1	0.077
1.2	0.095
1.3	0.118
1.4	0.142
1.5	0.167
<i>No obs. 0.095</i>	

As a consequence of the multiplier, the measurement error level of the observation will also be multiplied by the factor. However, as shown in Table 10, a small difference in measurement error level will not greatly affect the simulation. The effect of the bias is much greater than the effect of errors in the estimated measurement error. Therefore, for observations to be beneficial, it is imperative that there be little bias in the observations.

## 5.2.6 Model error

### 5.2.6.1 Saturation Level

In WATClass, soil properties of a particular land cover type are assumed to be constant across all grid squares. However, as previously shown in the variability of the field capacity, there is a marked spatial variability in soil properties. Since the saturation level is constant across the agriculture land cover type in all grid squares, the model's error variance in the saturation level will be the spatial variance in the saturation level of the soil. It should be noted that the variance in the saturation level is the variance between grid squares, not between fields or points in a field; therefore, it is possible that much of the micro-scale variability will even out, such that the variance in the saturation level at the meso-scale is quite a bit smaller than that of the micro-scale.

**Table 12 – Error in soil moisture (% saturation) as a function of actual saturation level (as a fraction of estimated saturation level)**

<b>Saturation Level</b>	<b>RMS Error in soil moisture</b>
0.70	0.113
0.80	0.086
0.90	0.067
0.95	0.063
1.00	0.063
1.05	0.068
1.10	0.075
1.20	0.096
1.30	0.119

Table 12 displays the error in soil moisture as a function of saturation level. The estimated saturation level for the model is 1.000 and the estimated saturation error level for the ensemble Kalman filter is 0.1. Therefore, the errors in saturation represented in the test are  $\pm 3$ ,  $\pm 2$ ,  $\pm 1$ ,  $\pm 0.5$  and 0 standard deviations away from the mean. The error in the soil moisture estimates improves for all levels of the actual saturation level.

#### 5.2.6.2 Drying decay constant

The decay constant simulates drainage and evapotranspiration in the WATClass simulation. The error in soil moisture generally decreases as the error variance in the decay constant decreases. The correlation between the decay constant error variance and the error is weakened when the observation error variance is small because the soil estimate is mostly made up of the observation. When the observation measurement error variance is higher, the level of error in soil moisture is more dependent on the decay constant error variance. Table 13 shows the average error in soil moisture as a function of decay constant error variance and observation measurement error variance. The last column in the table displays the RMS error in the soil moisture if no observations are made (or if observations are not in any way reliable).

**Table 13 – Error in soil moisture (%saturation) as a function of decay constant error variance**

$\sigma_{decay}$	RMS error $\sigma_{eobs} = 0.05$	RMS error $\sigma_{eobs} = 0.10$	RMS error $\sigma_{eobs} = 0.20$	RMS Error $\sigma_{eobs} = \infty$
0.05	0.051	0.057	0.064	0.072
0.06	0.053	0.059	0.067	0.077
0.07	0.057	0.062	0.071	0.081
0.08	0.058	0.065	0.075	0.086
0.09	0.060	0.068	0.078	0.091
0.10	0.063	0.071	0.082	0.096
0.12	0.068	0.076	0.089	0.106
0.14	0.073	0.081	0.096	0.115
0.16	0.078	0.087	0.102	0.126
0.18	0.084	0.093	0.109	0.136
0.20	0.090	0.098	0.116	0.146

Note:  $\sigma$  are  $\sigma/\mu$

Table 14 presents, the RMS error in soil moisture as a function of error in the estimated decay constant error level. The decay constant estimated error is presented as fraction of the actual decay constant error. As with the observation measurement error variance, the best soil moisture estimates are produced when the estimated error variance is equal to the actual error variance. When the decay constant error variance is underestimated, the Kalman filter overestimates the accuracy of the a priori model and consequently underestimates the value of the observation. Therefore, the improvement in soil moisture estimates from the observations is lessened as the ratio of estimated decay error level to actual decay error lessens. However, the error in soil moisture level will not drop below the error level if no observations were made.

When the decay error is overestimated, the Kalman filter will underestimate the value of the model and consequently underestimates the value of the model. The effect of the overestimation of the decay constant on the accuracy of the soil moisture measurement increases as the accuracy of the decay constant increases.

**Table 14 – Error in soil moisture (%saturation) as a function of error in decay constant error level**

<b>Multiplier <math>x</math> (<math>\hat{\sigma}_{edecay} = x\sigma_{edecay}</math>)</b>	<b>RMS error (<math>\sigma_{edecay} = 0.05</math>)</b>	<b>RMS error (<math>\sigma_{edecay} = 0.1</math>)</b>	<b>RMS error (<math>\sigma_{edecay} = 0.2</math>)</b>
0.200	0.062	0.082	0.122
0.250	0.061	0.080	0.118
0.333	0.061	0.078	0.112
0.500	0.060	0.074	0.104
0.707	0.058	0.071	0.100
1.000	0.057	0.070	0.097
1.414	0.057	0.071	0.097
2.000	0.058	0.072	0.097
3.000	0.061	0.074	0.098
4.000	0.063	0.075	0.098
5.000	0.064	0.075	0.098
No observations	0.071	0.095	0.145

### 5.2.6.3 Rain

In the simulation, rain is represented by the increase in the soil moisture value caused by rain on a given day. Theoretically, this value is proportional to the volume of rain falling on a grid square on a given day. The percentage error in the measurement of the simulated rain is assumed to be constant with respect to the level of rain. Several different levels of percentage error in the measurement of rain are run through the simulation. The average RMS error in daily soil moisture and RMS error in yearly runoff level for a hundred simulated parameter sets with five separate simulated observation sets are shown in Table 15. For this test, the estimated error in rain measurement is assumed to be equal to the actual error in rain measurement.

The error in rain measurement has less effect on the soil moisture estimates than the error in the decay constant, which can be attributed to the infrequency of rainy days. That is, error in rain measurement is only applicable on days when rain occurs, which is approximately 35% of the time.

The error in estimated rain measurement error is also examined. The results are shown in Table 16. The rain measurement estimated error is presented as a fraction of the actual rain measurement error. As with the other error variances, the best soil moisture estimates are produced when the estimated error variance is equal to the actual error variance. When the rain measurement error level is low, there is little dependence on the estimated error level because the majority of the model error is caused by the decay constant error. As the rain measurement error increases, the error in rain measurement error has more of an effect on the soil moisture error; however, the decay constant still makes up the majority of the error.

**Table 15 – Error in soil moisture as a function of error in rain measurement**

<b>Error in rain measurement %</b>	<b>RMS error soil moisture</b>	<b>RMS error soil moisture (no obs)</b>
0.00	0.071	0.096
0.05	0.071	0.097
0.10	0.073	0.099
0.15	0.074	0.102
0.20	0.076	0.105
0.25	0.079	0.110
0.30	0.083	0.115

**Table 16 – Error in soil moisture as a function of error in estimated rain measurement error**

<b>Multiplier <math>x</math> (<math>\hat{\sigma}_{rain} = x\sigma_{rain}</math>)</b>	<b>RMS error <math>\sigma_{rain} = 0.05</math></b>	<b>RMS error <math>\sigma_{rain} = 0.10</math></b>	<b>RMS error <math>\sigma_{rain} = 0.20</math></b>
0.200	0.0707	0.0718	0.0767
0.250	0.0707	0.0718	0.0767
0.333	0.0707	0.0718	0.0764
0.500	0.0707	0.0717	0.0761
0.707	0.0706	0.0716	0.0759
1.000	0.0706	0.0716	0.0759
1.414	0.0706	0.0717	0.0760
2.000	0.0706	0.0721	0.0766
3.000	0.0708	0.0728	0.0775
4.000	0.0712	0.0735	0.0781
5.000	0.0716	0.0742	0.0785
No observations	0.0956	0.0969	0.1041

## Chapter 6

### Conclusions and Recommendations

#### 6.1 Conclusions

When the Grand River watershed and Roseau River watershed ground sampled data are examined, a spatial relationship is found in soil moisture. An exponential relationship is established between the distance between samples and the correlation coefficient between the soil moisture values of those fields. Due to the limited number of dates sampled, however, the error in the sample correlation coefficients is high, rendering the correlation coefficient function ineffectual.

The ensemble Kalman filter is found to reduce the error in the soil moisture estimates from approximately 9.5% to 7.1% of percentage of saturation level. An ensemble size of 1000 is found to be appropriate for the WATClass model.

Soil moisture value estimate error reduces as observation frequency increases. Obviously daily measurements are ideal. The improvement due to the observations is negligible approximately fourteen days after the observations are assimilated. Furthermore, as observation measurement error decreases, soil moisture estimates will decrease. If there is error in the estimated measurement error, then the accuracy of the soil moisture estimates will reduce. However, only if the measurement error level is high, and the measurement error is grossly underestimated will the soil moisture estimates with the assimilation of the observations be worse off than the soil moisture estimates without the assimilation of the observations. Bias in the observations, i.e. if the observations are consistently too high or too low, has more of an effect on the accuracy of the estimated soil moisture values. If the measurements are consistently 20% higher or 20% lower than the actual soil moisture value, then the soil moisture estimates will be worse off than if no observations are made.

Elevated error in the saturation level of the model increases the error in the soil moisture estimates. However, the improvement due to the observations is more pronounced when the saturation level error is higher.

Errors in the decay constant, which will be affected by WATClass drainage, interflow and evapotranspiration rates also have a large effect on the soil moisture values. If the error in the decay constant is low then the improvement in soil moisture is minimal. Error in the error level of the decay constant does not have a great deal of an effect on the soil moisture estimates.

Error in the measurement of rain has less of an effect than the error in decay constant. Subsequently error in error in measurement of rain also has little effect.

## **6.2 Recommendations**

To accurately assimilate remotely sensed soil moisture measurements into WATClass, the spatial variability of soil moisture values must be further examined. While the spatial variability analysis study presented in this thesis performed some of this analysis, the sample size is too small to be produce accurate results. While the ground sampling of more fields on more dates is impractical, the knowledge of the exact spatial variability at several different resolutions will be highly beneficial. As the observation of remote sensing of soil moisture becomes more accurate, remotely sensed soil moisture maps can be used in place of ground sample data to determine the spatial relationship. In addition to the spatial variability, the error variances for the parameters and input variables into WATClass should also be determined.

The ensemble Kalman filter should also be implemented into WATClass proper. Simulated data should be used as inputs initially to determine the efficiency of the algorithm. In addition to determining how much soil moisture estimates improve with a more complete hydrological model, a WATClass simulation would also be able to determine the effect of the assimilation of soil moisture on streamflow water levels. An improvement in the estimation of soil moisture content will improve

the estimation of surface runoff levels. Subsequently, the estimates of streamflow levels will be more accurate the prediction of flood occurrences will be better. If possible, the ensemble Kalman filter should be exercised on the WATClass at different spatial scales. There may be some advantage in running WATClass with square size equal to the remotely-sensed images pixel size.

Since remotely sensing of soil moisture currently can only be performed on bare soil or light vegetation, some study should be done to determine how much of WATClass's agriculture land class is made up of bare soil or light vegetation fields. Furthermore, some study should be undertaken to determine whether the soil moisture estimates of other land types can be updated using the data from soil estimates of the bare soil and light vegetation fields.

Finally, once remotely sensed soil moisture assimilation is properly implemented into WATClass, the remotely sensed soil moisture estimates can be used to correct the underlying parameters in WATClass. For example, if it is found that WATClass is consistently underestimating the rate at which the soil drains after a rainfall event, the hydraulic conductivity parameter for that grid square should be increased. With this in mind, soil parameters can be made to vary spatially from grid square to grid square, instead of just varying between WATClass land types.

## References

- Z. Ali, G. Kroupnik, G. Matharu, J. Graham, I. Barnard, P. Fox and G. Raimondo, 2004. "RADARSAT-2 space segment design and its enhanced capabilities with respect to RADARSAT-1", *Canadian Journal of Remote Sensing*, **30**(3), 235-245.
- R. Bindlish and A.P. Barros, 2000. "Multifrequency Soil Moisture Inversion from SAR Measurements with the Use of IEM", *Remote Sensing of Environment*, **71**(1), 67-88.
- S.C. Choi and R. Wette, 1969. "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias", *Technometrics*, **11**(4) 683-690.
- R. Coe and R.D. Stern, 1982. "Fitting Models to Daily Rainfall Data", *Journal of Applied Meteorology*, **21**(7), 1024-1031.
- W.L. Crosson, C.A. Laymon, R. Inguva and M.P. Schamschula, 2002. "Assimilating remote sensing data in a surface flux-soil moisture model", *Hydrological Processes*, **16**, 1645-1662.
- Delta Devices. 1999. *Thetaprobe Soil Moisture Sensor*. User manual, MI1-UM-1.21, Delta Devices.
- A. Deschamps, T.J. Pultz, A. Pietroniro and K. Best, 2004. "Temporal Soil Moisture Estimates from Radarsat-1 and Envisat ASAR for Flood Forecasting", *Proceedings of the IGARSS04*, September 20-24, Anchorage, **3**, 2119-2122.
- S.L. Dingman, 2002. *Physical Hydrology*. 2nd Edition, Upper Saddle River: Prentice-Hall.
- M.C. Dobson and F.T. Ulaby, 1998. "Mapping soil moisture distribution with imaging radar", in *Principles and Applications of Imaging Radar, Manual of Remote Sensing*. Third Edition, **2**. F.M. Henderson and A.J. Lewis eds. New York: John Wiley & Sons, 407-433.
- D. Entekhabi, E.G. Njoku, P. Houser, M. Spencer, T. Doiron, Y. Kim, J. Smith, R. Girard, S. Belair, W. Crow, T.J. Jackson, Y.H. Kerr, J.S. Kimball, R. Koster, K.C. McDonald, P.E. O'Neill, T. Pultz, S.W. Running, J. Shi, E. Wood and J. van Zyl, 2004. "The Hydrosphere State (Hydros) Satellite Mission: An Earth System Pathfinder for Global Mapping of Soil Moisture and Land Freeze/Thaw", *IEEE Trans. on Geoscience and Remote Sensing*, **42**(10), 2184-2195.
- ESA, "Envisat ASAR: Science and Applications", SP-1225, European Space Agency, 1998.
- ESA, "ASAR Product Handbook Issue 2.2", European Space Agency, 2007.
- G. Evensen, 2003. "The Ensemble Kalman Filter: theoretical formulation and practical implementation", *Ocean Dynamics*, **53**, 344-367.
- J.F. Galantowicz, D. Entekhabi and E.G. Njoku, 1999. "Tests of Sequential Data Assimilation for Retrieving Profile Soil Moisture and Temperature from Observed L-Band Radiobrightness", *IEEE Trans. on Geoscience and Remote Sensing*, **37**(4), 1860-1870.
- R. Hoeben and P.A. Troch, 2000. "Assimilation of active microwave observation data for soil moisture profile estimation", *Water Resources Research*, **36**(10), 2805-2819.
- S.J. Julier and J.K. Uhlman, 1997. "A New Extension of the Kalman Filter to Non-linear Filters", *Proceedings SPIE*, **3068**, 182-193.
- R.E. Kalman, 1960. "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME - Journal of Basic Engineering*, **82**, 35-45.

- Y.H. Kerr, P. Waldteufel, J.-P. Wigneron, J.M. Martinuzzi, J. Font and M. Berger, 2001. "Soil Moisture Retrieval from Space: The Soil Moisture and Ocean Salinity (SMOS) Mission", *IEEE Trans. on Geoscience and Remote Sensing*, **39**(8), 1729-1735.
- N. Kouwen, E.D. Soulis, A. Pietroniro, J. Donald and R.A. Harrington, 1993. "Grouped Response Units for Distributed Hydrologic Modeling", *Journal of Water Resources Planning and Management*, **119**(3), 289-305.
- N. Kouwen, 2006. *WATFLOOD/WATROUTE: Hydrological Model Routing & Flow Forecasting System*. Computer program manual, University of Waterloo.
- S.A. Margulis, D. McLaughlin, D. Entekhabi and S. Dunne, 2002. "Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment", *Water Resources Research*, **38**(12), 1299.
- H. McNairn and B. Brisco, 2004. "The application of C-band polarimetric SAR for agriculture: a review", *Canadian Journal of Remote Sensing*, **30**(3), 525-542.
- B. Merz and E.J. Plate, 1997. "An analysis of the effects of spatial variability of soil and soil moisture on runoff", *Water Resources Research*, **33**(12), 2909-2922.
- B. Moeremans and S. Dautrebande, 2000. "Soil moisture evaluation by means of multi-temporal ERS SAR PRI images and interferometric coherence", *Journal of Hydrology*, **234**, 162-169.
- M.S. Moran, D.C. Hymer, J. Qi and E.E. Sano, 2000. "Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland", *Agriculture and Forest Meteorology*, **105**, 69-80.
- E.G. Njoku and D. Entekhabi, 1996. "Passive microwave remote sensing of soil moisture", *Journal of Hydrology*, **184**, 101-129.
- E.G. Njoku, T.J. Jackson, V. Lakshmi, T.K. Chan and S.V. Nghiem, 2003. "Soil Moisture Retrieval From AMSR-E", *IEEE Trans. on Geoscience and Remote Sensing*, **41**(2), 215-229.
- A. Quesney, S. Le Hégarat-Masclé, O. Taconet, D. Vidal-Madjar, J.P. Wigneron, C. Loumagne and M. Normand, 2000. "Estimation of Watershed Soil Moisture Index ERS/SAR Data", *Remote Sensing of Environment*, **72**, 290-303.
- RSI, "RADARSAT Data Products Specification", RSI-GS-026, RADARSAT International, 2000.
- R.K. Raney, 1998. "Radar Fundamentals: Technical Perspective", in *Principles and Applications of Imaging Radar, Manual of Remote Sensing*. Third Edition, **2**. F.M. Henderson and A.J. Lewis eds. New York: John Wiley & Sons, 9-130.
- R.H. Reichle, D.B. McLaughlin and D. Entekhabi, 2002. "Hydrologic Data Assimilation with the Ensemble Kalman Filter", *Monthly Weather Review*, **130**(1), 103-114.
- F. Seglenieks, E.D. Soulis and N. Kouwen, 1998. "Establishment of a relationship between RADARSAT backscatter and soil moisture at hydrological modelling scales", *Proceedings of the Fourth International Workshop on Applications of Remote Sensing in Hydrology*, November 4-6, Santa Fe, 375-386.
- F. Seglenieks and B. Davison, 2001. *WATClass Users Guide*. Version 6.0. Computer program manual, University of Waterloo.
- S. Selvalingam and M. Miura, 1978. "Stochastic modeling of monthly and daily rainfall sequences", *Journal of the American Water Resources Association*, **14**(5), 1105-1120.

- J. Shi, L. Jiang, L. Zhang, K.S. Chen, J.P. Wigneron, A. Chanzy and T.J. Jackson, 2006. "Physically Based Estimation of Bare-Surface Soil Moisture With the Passive Radiometers", *IEEE Transactions on Geoscience and Remote Sensing*, **44**(11), 3145-3153.
- E.D. Soulis, N. Kouwen, A. Pietroniro, F.R. Seglenieks, K.R. Snelgrove, P. Pellerin, D.W. Shaw and L.W. Martz, 2005. "Prediction in Ungauged Basins: Approaches for Canada's Cold Regions", in *A Framework for Hydrological Modelling in MAGS*. Spence, C., J.W. Pomeroy and A. Pietroniro eds. Cambridge: Canadian Water Resources Association, 119-138.
- E.D. Soulis, K.R. Snelgrove, N. Kouwen and F. Seglenieks, 2000. "Toward Closing the Vertical Water Balance in Canadian Atmosphere Models: Coupling of the Land Surface Scheme CLASS with the Distributed Hydrological Model WATFLOOD", *Atmosphere-Ocean*, **38**(1), 251-269.
- W.L. Tang, J.R. Wang and P.C. Doraiswamy, 1993. "Relationship between satellite microwave radiometric data, antecedent precipitation index, and regional soil moisture", *International Journal of Remote Sensing*, **14**(13), 2283-2500.
- F.T. Ulaby, R.K. Moore and A.K. Fung, 1986. *Microwave Remote Sensing Active and Passive*. Delham: Artech House.
- F.T. Ulaby, P.C. Dubois and J. van Zyl, 1996. "Radar Mapping of Surface Soil Moisture", *Journal of Hydrology*, **184**, 57-84.
- J.P. Walker and P.R. Houser, 2004. "Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution", *Advances in Water Resources*, **27**, 785-801.

## **Appendix A**

### **Glossary/Acronyms**

API: Antecedent Precipitation Index: an index derived from rainfall measurements, which can be compared with or used to estimate soil moisture.

C-band: a portion of the microwave electromagnetic range of frequencies ranging from 4 to 8 GHz (3.75-7.5 cm in wavelength). C is for “compromise”, as in a compromise between L- and X-band

CLASS: Canadian Land Surface Scheme: a land surface parameterization scheme for use in large scale climate models developed by the Canadian Atmospheric Environmental Service.

Degree of saturation: a physical property of soil indicating the percentage of void space in the soil filled with water.

Envisat ASAR: ESA’s ENVIRONMENTAL SATellite Advanced Synthetic Radar launched in March 2002.

ERS-1/2: European Remote Sensing satellites: satellites launched by ESA in July 1991 and April 1995 respectively. Both satellites have SAR sensors.

ESA: European Space Agency: an intergovernmental organisation dedicated to the exploration of space, currently with 17 member states. It is headquartered in Paris, France.

Evapotranspiration: the process by which water leaves the soil into the atmosphere.

Gravimetric water content: soil moisture value derived from the weight ratio of water to soil material.

Grid square: the basic computational unit of WATClass.

Incidence angle: the angle between the line of sight from the radar to an element of an imaged scene, and the vertical direction characteristic of the scene.

Infiltration: the process by which water enters the soil.

Interflow: the lateral movement of water in the soil.

L-band: a portion of the microwave electromagnetic range of frequencies ranging from 1 to 2 GHz (15-30 cm in wavelength). L is for “long” wavelength.

Meso-scale: a scale of meteorological phenomena ranging from 1 to 100 km in horizontal extent. It is synonymous with regional-scale or watershed-scale for hydrology.

Micro-scale: a scale of meteorological phenomena less than 1 km in horizontal extent. It is synonymous with local-scale.

NASA: National Aeronautics and Space Administration: an agency of the United States government responsible for the public space program.

Permanent wilting point: the soil moisture value at which a plant wilts and can no longer recover its turgidity.

RADARSAT-1/2: Canadian commercial Earth observation satellites imaging the Earth with a C-band (5.3 GHz) SAR. RADARSAT-1 was launched in November 1995, whereas RADARSAT-2 will be launched in summer 2007.

RAR: Real-Aperture Radar: a type of imaging radar.

Repeat cycle: the period between times that the imaging device can view the same location on the Earth.

SAR: Synthetic-Aperture Radar: a type of imaging radar.

SMOS: Soil Moisture and Ocean Salinity: a space mission run by ESA to observe soil moisture over land and salinity over oceans. The SMOS satellite contains the Microwave Imaging Radiometer with Aperture Synthesis (MIRAS), a passive microwave radiometer.

Soil porosity: a measure of void spaces in soil.

Stream network: the geomorphologic patterns of hierarchical connections formed by streams as they drain the watershed. Also called drainage network.

Surface roughness: an indication of surface irregularity measured by the root-mean square (rms) of the surface variations.

ThetaProbe: a soil moisture monitoring device developed by Delta-T Devices.

Transpiration: the evaporation of water into the atmosphere from the leaves and stems of plants.

UTM: Universal Transverse Mercator coordinate system: a grid-based method of specifying locations on the surface of the Earth.

Volumetric water content: soil moisture value derived from the volume ratio of water to soil material.

Watershed: the region of land whose water drains into a specified body of water.

Wetting front: the interface between soil that is unchanged from the initial state and the saturated zone caused by the infiltration of pooled surface water.

X-band: a portion of the microwave electromagnetic range of frequencies ranging from 8 to 12 GHz (2.5-3.75 cm in wavelength). Named X because it was secret during WWII.

## Appendix B

### Spatial Statistics

#### B.1 Correlation Coefficient

The correlation coefficient between the ground sampled fields indicates the extent to which the fields' soil moisture values tend to vary in unison (Dingman, 2002).

Given soil moisture estimates,  $x_{kt}$ ,  $k = 1 \dots n$ ,  $t = 1 \dots T$  for  $n$  fields sampled on  $T$  dates. The average soil moisture,  $\mu_{xk}$ , and the variance of the soil moisture,  $\sigma_{xk}^2$ , for a particular field are

$$\mu_{xk} = \frac{1}{T} \sum_{t=1}^T x_{kt}, \text{ and}$$

$$\sigma_{xk}^2 = \frac{1}{T-1} \sum_{t=1}^T (x_{kt} - \mu_{xk})^2 .$$

There are a total of  $\frac{n(n+1)}{2}$  pairs of fields. For each pair, say  $k = a$  and  $k = b$ , the sample correlation between the fields is

$$r_{a,b} = \frac{\sum_{t=1}^T (x_{at} - \mu_{xa})(x_{bt} - \mu_{xb})}{(T-1)\sqrt{\sigma_{xa}^2 \sigma_{xb}^2}} .$$

Standard error can be computer for a correlation coefficient by considering the transformed variable,  $\hat{W}_{X,Y}$ ,

$$\hat{W}_{x,y} \equiv 0.5 \cdot \ln \left[ \frac{1+r_{X,Y}}{1-r_{X,Y}} \right].$$

The standard error of  $\hat{W}_{X,Y}$ ,  $\sigma_{WXY}$ , is

$$\sigma_{WXY} = \frac{1}{\sqrt{N-3}},$$

where  $N$  is the sample size.

Therefore, the standard error for  $r_{X,Y}$ ,  $\sigma_{r_{XY}}$  can be equated as a function of  $r_{X,Y}$ .

## B.2 Variogram

A variogram is a function describing the degree of spatial dependence of a random field of stochastic process,  $Z(x)$ . It is defined as the expected squared difference between the values between locations  $x$  and  $y$

$$2\gamma(x, y) = E\left((Z(x) - Z(y))^2\right).$$

If the process is stationary, then the variogram can be expressed as a function of the difference in locations,  $h=x-y$ . Therefore, the variogram can be defined as

$$\gamma(x, y) = \gamma(h = x - y).$$

Note that  $h$  is a vector.

If the process is isotropic, then the variogram can be expressed as a function of the distance between locations. Therefore, the variogram can be defined as

$$\gamma(x, y) = \gamma(\|x - y\|).$$

To calculate a variogram, the squared differences between all pairs of locations are determined and grouped by distance between locations. The average squared difference of each distance is used to estimate the variogram value at that distance. The variogram can be downsampled by averaging over ranges of distances.

## B.3 Average distance

Average distance,  $d_{\text{avg}}$ , from point  $p$  to area  $A$  can be defined as

$$d_{\text{avg}} = \frac{\int_A r \, dA}{\int_A dA},$$

where  $r$  is the distance from  $p$  to the infinitesimal area  $dA$ .

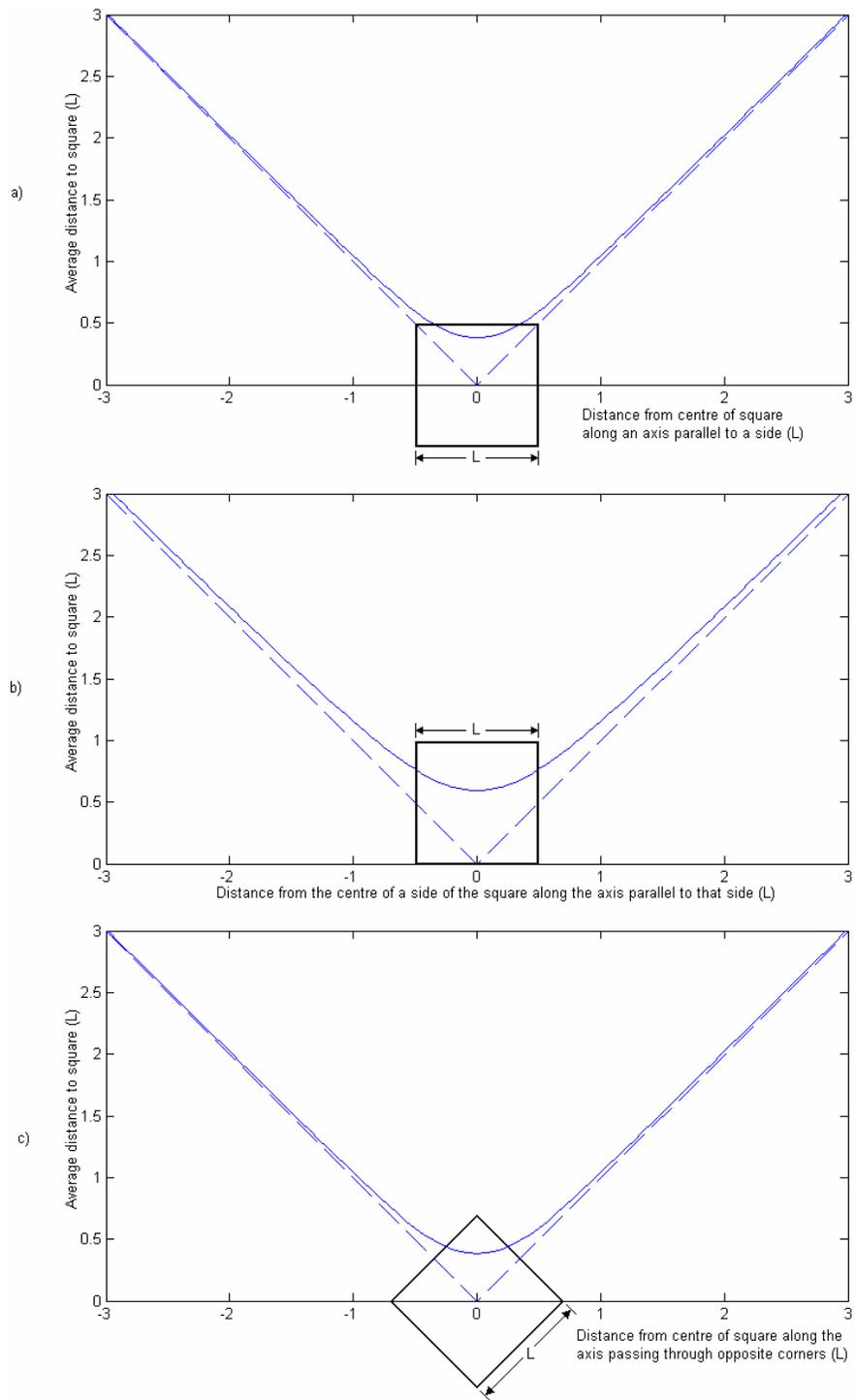
This distance can be computed numerically or explicitly. The numerical solution involves breaking the area up into smaller squares, computing the distance between the point and the centre of each of these smaller areas and then averaging the distance to each of these small areas. The explicit solution involves integrating over the area of the square using polar coordinate system. Then the integral to calculate the average distance becomes

$$d_{\text{avg}} = \int_{r_1}^{r_2} \theta r \cdot r \cdot dr ,$$

where  $\theta$  is the portion of the circle centred at  $p$  and of radius  $r$  that lies within  $A$ , in radians,  
 $r_1$  is the shortest distance between  $p$  and  $A$ , and  
 $r_2$  is the farthest distance between  $p$  and  $A$ .

The angle,  $\theta$ , can be related to  $r$  and the distances from the point to the edges of  $A$  through trigonometry. The integral is broken up into smaller integrals as different edges come into play. Figure 17 displays the average distance away from a square side length  $L$  along several different axes.

Table 17 presents the average distance from a point to a square of side length  $L$  from several locations. Note that the average distance away from the square approaches the distance away from the centre of the square as the distance increases.



**Figure 17 – Average distance from a square**

**Table 17 – Average distance to square of side length  $L$**

Point location		Average Distance	Integral
Centre of square		0.3826 $L$	 $\int_0^{L/2} 2\pi r^2 dr$
			 $4 \int_{L/2}^{L/\sqrt{2}} \left( \frac{\pi}{2} - 2\text{Sec}^{-1} \frac{2r}{L} \right) r^2 dr$
Centre of edge of square		0.5932 $L$	 $\int_0^{L/2} \pi r^2 dr$
			 $2 \int_{L/2}^L \left( \text{Csc}^{-1} \frac{2r}{L} \right) r^2 dr$
			 $2 \int_L^{\sqrt{5}/2 L} \left( \text{Csc}^{-1} \frac{2r}{L} - \text{Sec}^{-1} \frac{r}{L} \right) r^2 dr$
Corner of square		0.7652 $L$	 $\int_0^L \frac{\pi}{2} r^2 dr$
			 $\int_L^{\sqrt{2}L} \left( \frac{\pi}{2} - 2\text{Sec}^{-1} \frac{r}{L} \right) r^2 dr$

## Appendix C

### Imaging Radar Beam Modes

**Table 18 – RADARSAT-1 beam modes (Raney, 1998)**

<b>Mode</b>	<b>Resolution</b> <i>r × a</i> (m × m)	<b>Looks</b>	<b>Swath Width</b> (km)	<b>Incidence angle</b> (degrees)
Standard	25×28	4	100	20-49
Wide (1)	48-30×28	4	165	20-31
Wide (2)	32-35×28	4	150	31-39
Fine Resolution	11-9×9	1	45	37-48
ScanSAR (N)	50×50	2-4	305	20-40
ScanSAR (W)	100×100	4-8	510	20-49
Extended (H)	22-19×28	4	75	50-60
Extended (L)	63-28×28	4	170	10-23

**Table 19 – RADARSAT-1 standard beam positions (RSI, 2000)**

<b>Beam Position</b>	<b>Incidence Angle (°)</b>	<b>Resolution</b> <i>r × a</i> (m × m)
S1	20-27	27.9-20.5×27.0
S2	24-31	22.8-18.0×27.0
S3	30-37	27.5-23.2×27.0
S4	34-40	25.1-21.8×27.0
S5	36-42	23.6-20.7×27.0
S6	41-46	21.5-19.2×27.0
S7	45-49	19.8-18.4×27.0

**Table 20 – RADARSAT-2 beam modes (Ali, 2004)**

Mode	Resolution $rg \times az$ (m $\times$ m)	Looks $rg \times az$	Swath Width (km)	Incidence angle (degrees)
<i>Selective polarisation (transmit H or V; receive H and (or) V)</i>				
Fine	10 $\times$ 9	1 $\times$ 1	50	37–49
Standard	25 $\times$ 28	1 $\times$ 4	100	20–49
Low incidence	40 $\times$ 28	1 $\times$ 4	170	10–23
High incidence	20 $\times$ 28	1 $\times$ 4	70	50–60
Wide	25 $\times$ 28	1 $\times$ 4	150	20–45
ScanSAR narrow	50 $\times$ 50	2 $\times$ 2	300	20–46
ScanSAR wide	100 $\times$ 100	4 $\times$ 4	500	20–49
<i>Polarimetric (transmit H and V on alternate pulses; receive H and V on any pulse)</i>				
Fine quad	11 $\times$ 9	1 $\times$ 1	25	20–41
Standard quad	25 $\times$ 28	1 $\times$ 4	25	20–41
<i>Selective single polarisation</i>				
Ultra-fine	3 $\times$ 3	1 $\times$ 1	20	30–40

**Table 21 - Envisat ASAR beam modes (ESA, 1998)**

Mode	Spatial res (m)	Swath width (km)	Polarisation	Equivalent no. of looks	Incidence angle (°)
Image	30	56-105	VV or HH	>3.9	15-45
Alt. Polarisation	30	56-105	VV and HH HH and HV VV and VH	>1.9	15-45
Wide Swath	150	405	VV or HH	~11.5	17-42
Global Monitoring	1000	405	VV or HH	~7-9	17-42
Wave	10	5	VV or HH	1	15-45

**Table 22 – Envisat ASAR image and alternating polarisation mode beam positions (ESA, 1998)**

Image Swath	Swath Width (km)	Incidence Angle Range (°)
IS1	104.8	15.0-22.9
IS2	104.8	19.2-26.7
IS3	81.5	26.0-31.4
IS4	88.1	31.0-36.3
IS5	64.2	35.8-39.4
IS6	70.1	39.1-42.8
IS7	56.5	42.5-45.2