ABSTRACT

Title of Document: SENSITIVITY ANALYSIS OF MACHINE LEARNING IN BRIGHTNESS TEMPERATURE PREDICTIONS OVER SNOW-COVERD REGIONS USING THE ADVANCED MICROWAVE SCANNING RADIOMETER

Yuan Xue, Master of Science, 2014

Directed By: Assistant Professor, Barton A. Forman
Department of Civil and Environmental Engineering

Snow is a critical component in the global energy and hydrologic cycle. Further, it is important to know the mass of snow because it serves as the dominant source of drinking water for more than one billion people worldwide. Since direct quantification of snow water equivalent (SWE) is complicated by spatial and temporal variability, space-borne passive microwave SWE retrieval products have been utilized over regional and continental-scales to better estimate SWE. Previous studies have explored the possibility of employing machine learning, namely an artificial neural network (ANN) or a support vector machine (SVM), to replace the traditional radiative transfer model (RTM) during brightness temperatures (Tb) assimilation. However, we still need to address the following question: What are the most significant parameters in the machine-learning model based on either ANN or SVM? The goal of this study is to compare and contrast sensitivity analysis of Tb with respect to each model input between the ANN- and SVM-based estimates. In general, the results suggest the SVM (relative to the ANN) may be more beneficial during Tb assimilation studies where enhanced SWE estimation is the main objective.

SENSITIVITY ANALYSIS OF MACHINE LEARNING IN BRIGHTNESS TEMPERATURE PREDICTIONS OVER SNOW-COVERD REGIONS USING THE ADVANCED MICROWAVE SCANNING RADIOMETER

By

Yuan Xue

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2014

Advisory Committee:
Assistant Professor Barton A. Forman, Chair
Professor Richard H. McCuen
Associate Professor Kaye L. Brubaker

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Forman for his guidance, assistance and patience during my two-year-study. I would also like to thank him for providing me with the opportunity to come to the U.S., for introducing supercomputer and data assimilation techniques to me, and for his generous support, and thoughtful consideration. I am grateful that I joined his research group and continue to learn more from him as I continue my research.

I would like to thank Dr. McCuen for the great modeling techniques I have learned in his class. I am grateful for his valuable suggestions on how to write a thesis. I am also grateful for his patience in answering my questions in and out of the class, no matter how weird my questions may seem.

I would like to thank Dr. Brubaker for organizing the Water Resources tubing trip when I first came here. I would like to thank her for lending me an office when my office temporarily flooded. I would also like to thank her for teaching me how to use ArcGIS®.

I would like to thank Saad B. Tarik for reviewing my draft of thesis, and for his help on my academic study over the last four semesters. I would like to thank Yilu Feng and Yan Wang, for sharing their past experiences with me. I would like to thank my officemates in the EGL 0147 for cheerful discussions during the lunchtime. I would like to thank all my friends in the U.S. for their support. I would like to thank my family and friends back in China for their moral support. Special thanks go to Feng Shi, who is always supporting and encouraging me to pursue my dream and helping me out when I am in trouble.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION AND MOTIVATION

The following section describes the basics of snow (e.g., snow formation and snow properties) and the basics of remote sensing. It also explains why it is important to estimate snow parameters across large spatial scales and how to achieve such a goal.

1.1. INTRODUCTION OF SNOW

1.1.1. Definition and Formation of Snow

Snow is a permeable aggregate of ice grains with pores filled with air and water vapor [Bader, 1962]. It can also be defined as a type of winter solid precipitation composed of white or translucent ice crystals, chiefly in complex branch hexagonal form and often agglomerated into snowflakes [Glickman, 2000].

Snow generally originates in low or multi-layer stratiform clouds in cold weather when a minute cloud droplet freezes into a tiny particle of ice [Shuttleworth, 2012]. As water vapor starts condensing on its surface, the ice particle quickly develops facets, thus becoming a small hexagonal prism. As the small crystal becomes larger, branches begin to sprout from the six corners of the hexagon [Libbrecht, 2005]. Finally, a complex, branched and sometimes six-fold symmetric structure is developed (**Figure 1.1-1**). In addition, individual snowflakes all tend to look different since the snow crystal develops from various microscopic supercooled cloud droplets and also follows different forming paths. In principle, it can snow at any temperature below freezing, however, not every place in the world receives

snowfall because snow crystal growth depends on the temperature and pressure conditions in the cloud. One of the essential requirements for getting snow is to cool the air below freezing; orographic lifting is one of the most effective techniques to achieve vertical movement of air and hence rapid cooling below the freezing point of the water.

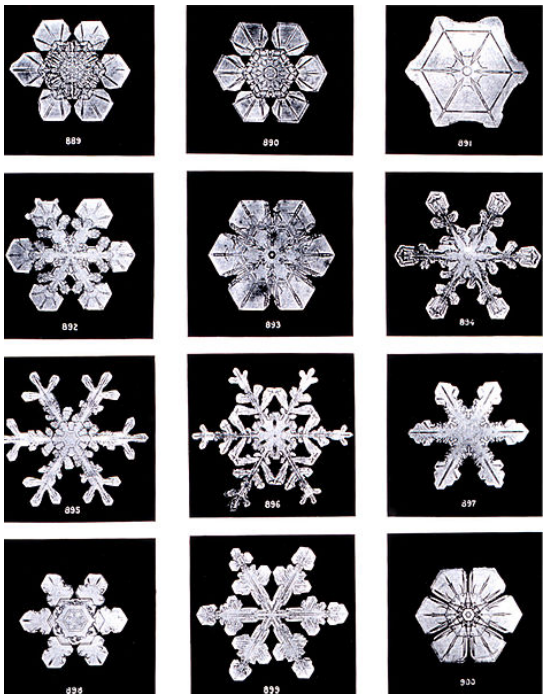

Figure 1.1-1 Six-fold snowflakes [Bentley, 1902].

Snow falling onto the ground can be classified using systems presented by Sommerfeld [1970], or the International Classification for Snow (Canada, National Research Council, 1954). These snow metamorphism classification systems are useful in terms of describing snow at scales ranging from millimeters to centimeters, or slightly larger [Sturm et al. 1995]. Sturm et al. [1995] proposed a technique for global applications based on the unique combinations of textural and stratigraphic characteristics (e.g., physical and thermal properties) of different types of snow, such as tundra, taiga or maritime. As is indicated by **Figure 1.1-2**, under different

combinations of climates and geography, there can be various types of snowpack: in general, tundra snowpack covers the largest portion, followed by the taiga class in the northern hemisphere.

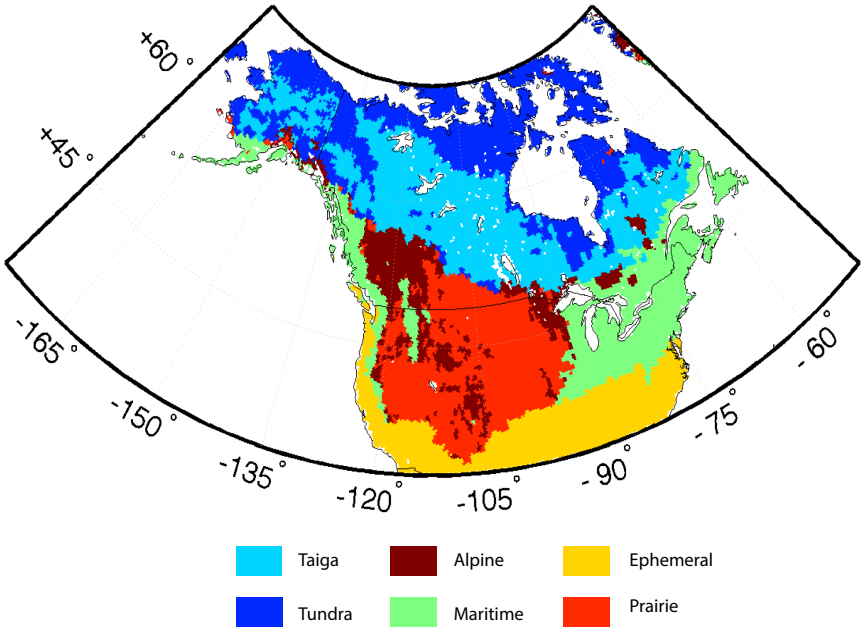

Figure 1.1-2 Snow classification in the study domain.

1.1.2. Importance of Snow

Snow influences a critical component in the global energy and hydrologic cycle by controlling mass and energy exchanges at the land surface [Robinson et al. 1993]. In addition, more than one billion people worldwide are dependent on snow as their main source of terrestrial freshwater supply [Foster et al. 2011]. Seasonal snow is highly variable in space and time and can cover from 7% to 40% of the northern hemisphere annually [Hall, 1985]. However, recent analysis of the updated snow cover extent (SCE) series indicates the northern hemisphere SCE in spring has reduced significantly over the past ~90 years [Brown and Robinson 2011] due to the

effect of global warming and unsteady large-scale atmospheric movement. As global temperature increases, it is estimated that regions currently receiving snowfall will increasingly receive precipitation in the form of rain. For every 1° C increase in temperature, the snowline rises by about 150 meters on average [Bogataj, 2007]. In other words, our virtual reservoir of freshwater – glaciers and snowcapped mountains are disappearing.

At the same time, an earlier onset of spring will induce earlier snowmelt and increases peak stream flow in many mountainous regions, which will increase the likelihood of flooding along the basin areas during the snow melting season. **In order to better understand the hydrologic responses associated with snow melt, we must first determine where and how much snow is found in the natural environment.**

1.1.3. Electromagnetic Attributes of Snow

It is known that every object on Earth emits and reflects radiation across a range of wavelengths [Campbell, 2002] except for objects at absolute zero. Scientists and engineers often compare snow and ice cover to a mirror on the surface of the Earth since snow has a relatively high albedo (a.k.a. reflection coefficient). Fresh snow with small snow grains and low densities could reflect more than 75% of the incident radiation, whereas wet earth may reflect as little as 5% [Lydolph, 1985]. Hence, snow cover presents a good contrast with most other natural land-related surfaces in the visible spectrum.

Snow cover on the ground also emits microwave at relatively low spectral frequency. When a sensor detects microwave radiation naturally emitted by the snow, that radiation is called passive microwave (PMW). Microwaves radiation possesses greater penetration depth through media than does optical (visible) radiation. As a result, microwave radiation is able to penetrate clouds and be used to detect snow during both day and night under all-weather conditions. Thus, passive microwave surveys as measured by space-borne microwave radiometers are particularly effective for detecting snow.

The electromagnetic attributes of snow are constantly changing. For example, the dielectric constant, a measure of the amount of polarization of the matter (e.g., snowpack) upon interaction with the electromagnetic wave, varies as the snow structure and liquid content change [Mulders, 1987; Duguay et al. 2005]. Typically, snow has a dielectric constant between 1.2 and 2.0 when the snow densities range from 0.1 to 0.5g/cm^3 [Hallikainen and Ulaby, 1986]. If the snowpack contains a larger amount of liquid water, it tends to have a higher dielectric constant because liquid water within the snowpack emits rather than scatters PMW radiation [Hall et al. 2004].

The differences in the electromagnetic attributes of snow can be revealed in the recorded radiation as measured by a space-borne radiometer. Tb, a measure of the radiance of microwave radiation travelling upward, is defined as the equivalent temperature of the microwave radiation thermally emitted by an object [Chang et al. 1976]. In general, the Tb increases as the wetness within the snowpack increases until

a saturation threshold of the Tb is reached [Tedesco et al. 2006]. Numerically, Tb is calculated as:

$$Tb = \varepsilon \cdot T_{physical} \tag{1.1-1}$$

where Tb [K] is the brightness temperature of the object; ε is a dimensionless quantity of the emissivity where $\varepsilon \in [0, 1]$; and $T_{physical}$ [K] is the physical temperature of the object (i.e., snow) of interest.

1.1.4. Physical Properties of Snow

Three of the most important properties of snow are snow density, snow depth and SWE [Pomeroy and Gray, 1995]. Once snow reaches the ground surface, the snow density will increase due to gravitational settling, wind compaction, freezing and re-freezing/re-crystallizing processes.

The snow density is the ratio between the snow mass and volume of the snow sample. A freshly fallen snow typically has a density around 100 kg/m^3 [Petrenko and Whitworth, 1999]. As snowpack ages, the snow is compacted, and as a result, its density often increases to greater than 300 kg/m^3 but less than 500 kg/m^3. Sometimes researchers use an equivalent water content (expressed as a percentage) to describe the density of snow as:

$$\rho_r\,(\%) = \frac{SWE}{D} \times 100\% \tag{1.1-2}$$

where ρ_r is the water content within the snow [%]; D is the snow depth; and SWE is the snow water equivalent. SWE and D should have the same units such as [m] or [cm]. For example, a snowpack with 0.5m of SWE and 2.5m snow depth, is specified as 20% of the density of water, or having 20% water content.

Based on the definition given by the National Weather Service (NWS), snow depth is the average depth of snow (including old snow and ice as well as new snow and ice) that remains on the ground at the observation time. It can be measured by a snow ruler or a ultrasonic snow depth sensor (see Chapter 2).

SWE is the amount of water contained within the snowpack, which characterizes the amount of water that could potentially melt and eventually enter neighboring streams. Hence, accurate estimation of SWE is crucial for flood prediction, power generation, and agriculture irrigation. Numerically, the magnitude of SWE is related to the product of snow depth and snow density, which can be expressed as:

$$\mathrm{SWE} = \frac{D \times \rho_{snow}}{\rho_{water}} \tag{1.1-3}$$

where SWE is the snow water equivalent [m]; D is the snow depth [m]; ρ_{snow} is the snow density $\left[\frac{kg}{m^3}\right]$; and ρ_{water} is the water density $\left[\frac{kg}{m^3}\right]$.

The amount of SWE changes with both time and space. The seasonality of SWE (**Figure 1.1-3**) shows that SWE is most likely to achieve its peak in March or April (depending on the latitude), which is a useful indicator of the amount of runoff that could potentially be available in the spring and summer following the cold season [Bohr and Aguado, 2001]. The spatial variability characteristics of SWE can be seen in **Figure 1.1-4**. Areas such as the Cascade Mountains of Washington, Oregon, Central Sierra, eastern Rockies and Regina and Winnipeg regions in Canada, contain greater magnitudes of SWE compared with other regions. In this example, topography plays a critical role in distributing SWE across the North America (NA) domain so that the heaviest accumulations are usually at mountain sites [Cayan, 1996]. Other aspects such as wind orientation, relative humidity, air temperature and large-scale

atmospheric movement will also exert their effects in determining SWE magnitudes. For instance, wind exposure often increases snow density from 10% to 25%, which will possibly result in a change in SWE due to Equation 1.1-3. Due to the highly variable nature of SWE distribution and extent and its complex relationship with synoptic atmospheric conditions, macro-scale prediction of SWE can be relatively inaccurate and contain significant uncertainties [Derksen et al. 2000].

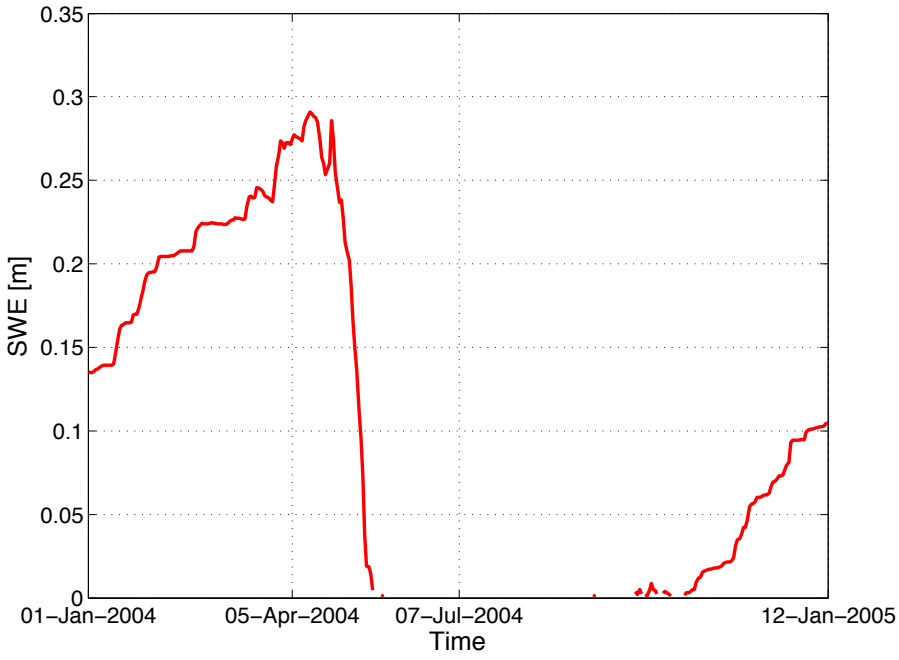

Figure 1.1-3 Annual variability (on a daily basis) of SWE for a location in Canada from 01 Jan 2004 to 12 Jan 2005 when the peak SWE occurred on 05 Apr, 2004.

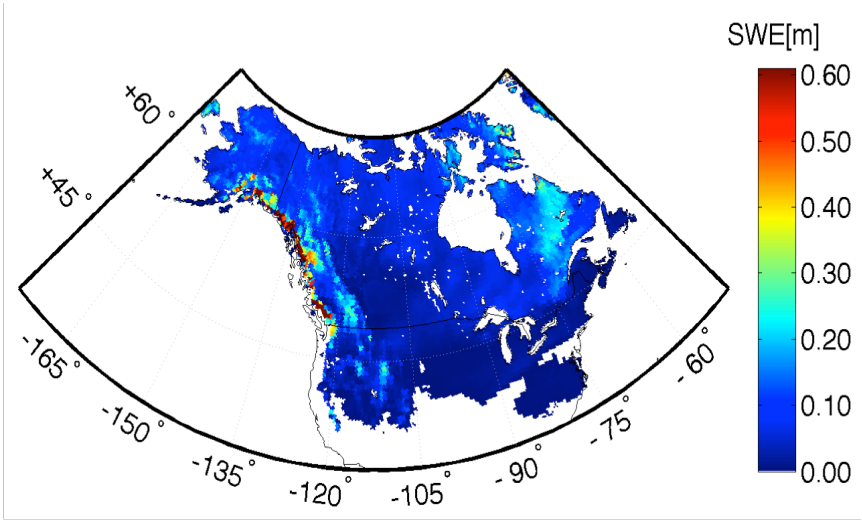

Figure 1.1-4 Spatial distribution of SWE across North America on 11 Jan 2004.

1.2. BASICS OF REMOTE SENSING

By recording emitted or reflected radiation as Tb, snow researchers can infer features of snow cover and snow mass via remote sensing using satellite-based sensors. Remote sensing is the science of acquiring, processing, and interpreting images, and related data via detecting the interaction between matter and electromagnetic radiation [Sabins, 2007]. There are different types of sensors designed to record electromagnetic radiation. For example, a radiometer, which can be either an infrared radiometer or a microwave radiometer, is a device for measuring the radiant flux of electromagnetic radiation emitted by an object. Alternatively, radar is an object detection system using electromagnetic waves to determine range, altitude, direction, or speed of moving objects. Similarly, LIDAR, which stands for light detection and ranging, utilizes visible light from pulsed lasers rather than lower frequency, electromagnetic waves to measure ranges to the Earth.

Passive microwave sensors used in this study, to be more specific, are based on an antenna system used to record the power of an electromagnetic wave emitted by the object and its surrounding environment (**Figure 1.2-1**) (e.g., overlying vegetation or underlying soil) in voltage, which is then converted via a built-in transmitter into Tb such that users are able to calculate the strength of reflected radiation.

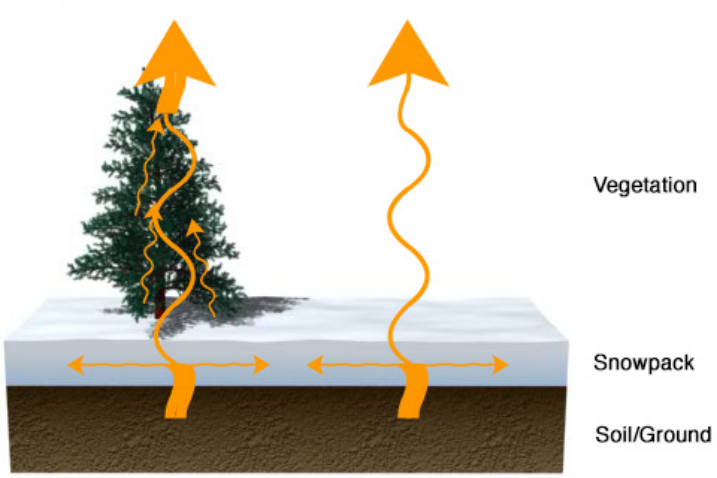

Figure 1.2-1 Electromagnetic wave emitted by each object (vegetation, snowpack and ground) on the surface. The orange arrows indicate the direction of the wave. The width of the arrow indicates the strength of measured radiation [reproduced from University Corporation for Atmospheric Research, the COMET ® Program].

Three main parameters used to design an antenna are: antenna size, frequency and polarization. As Fourier's theorem states, every piece of information in the universe can be completely expressed as a sum of sines and cosines of varying frequencies. Remote sensing analysts typically refer to an antenna in terms of the wavelength or frequency at which it operates. The antenna size should be on the order of one-tenth or more of the wavelength of the signal radiated [Lathi, 1990], but it is typically much larger on space-based sensors in order to achieve a minimum signal to

power ratio. For example, the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) onboard the Aqua satellite has an antenna size of 1.6m.

Antennas are also classified by their polarization, which is defined as the orientation of the electromagnetic wave with respect to the Earth surface [Mott, 1986]. Two types of linear polarizations are offered on AMSR-E: (1) horizontal polarization (H), and (2) vertical polarization (V). Users of AMSR-E measurements must first understand the characteristics of the antenna before collecting the documented data in order to choose the best combination of antenna frequency and polarization from the satellite-based measurements in accordance with the properties of their research target (i.e., snow).

1.3. GOALS AND OBJECTIVES

Since our knowledge of exactly how much SWE is present across the globe is complicated by the difficulty of collecting representative ground-based observations of SWE coupled with complex spatiotemporal uncertainty in snow processes [Dong et al. 2007], the goal of this research is to explore alternative methods to establish the connection between the physical property (e.g., SWE) and the electromagnetic characteristics of snow (in the form of Tb).

This goal was achieved through the following objectives:

1) Understand the basic principles of machine learning techniques for use as a measurement model operator in the prediction of PWM Tb, as originally presented in Forman et al. [2013] and Forman and Reichle [2014];

2) Optimize key parameters within the network set-ups for both artificial neural network (ANN) and support vector machine (SVM) based frameworks;
3) Conduct a sensitivity analysis to compare and contrast the performance of ANN- and SVM-based models;
4) Explain the differences between these two models, and relate the sensitivity results with the physical meaning of each technique;
5) Understand and characterize the limitations of the proposed model based on machine learning.

1.4. IMPLICATIONS

The research proposed here opens up a new avenue for PMW Tb estimation within an advanced land surface model via machine-learning techniques, including an ANN or a SVM. The sensitivity analysis conducted in this study is anticipated to further evaluate and verify the applicability and rationality of the technique. Conclusions drawn from this study will provide future SWE investigation with great research opportunities in terms of utilizing a better measurement model operator, such as SVM (or other machine learning techniques), rather than a traditional radiative transfer model (RTM) that has numerous (and significant) limitations. Therefore, the eventual goal of large-scale estimation of SWE can be achieved within a data assimilation network to be pursued in the future, but only after careful consideration of ANN and SVM sensitivities as conducted in this study.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

The following chapter describes various types of snow measurements and related snow parameter estimation products. It also discusses the similarities and differences between the ANN and SVM techniques.

2.1. IN-SITU SNOW MEASUREMENTS

In-situ techniques obtained from manual survey or ground-based stations provide reasonably accurate measurements of snow states and are not affected by forest cover [Armstrong et al. 2008; Moradkhani, 2008]. Snow measurement techniques at the point-scale include, but are not limited to, the following: (1) a snow ruler used to measure the snowfall, which is the maximum accumulation (or depth) of the freshly-fallen snow prior to settling or melting since the last observation [Ryan et al. 2008]; (2) graduated snow stakes used to measure snow interception, primarily in regions of deep snow; (3) an ultrasonic snow depth sensor to measure total snow depth based on the distance travelled by the emitted ultrasonic impulse [Lea and Lea 1998]; (4) a snow core to sample the snow at the observation time and location and provides information about the snow depth and SWE; and (5) a snow pillow/snow scale to measure the deflection of a pressure transducer and therefore is typically installed to determine the water-content of the overlying snowpack. In terms of large-scale (i.e., on the order of kilometers or more) snow measurements, one of the traditional methods is to estimate SWE/snow depth using an interpolation algorithm (e.g.,

kriging) across a large area based on the ground-based observations only [Dyer and Mote, 2006]. However, direct quantification of snow mass (or snow water equivalent) using interpolation is complicated by significant spatial and temporal variability. Further, the spatial resolution of in-situ measurements is limited by sparsely located stations and their proximity [Bechle et al. 2013] and hence, high quality ground-based measurements are not available everywhere such as mountainous areas or avalanche-prone terrain. Because of these limitations in the point-scale measurements, remote sensing is an attractive alternative for snow measurement across regional- and continental-scales [Foster et al. 1987].

2.2. SNOW REMOTE SENSING

Remote sensing of continental-scale seasonal snow cover has been widely used since the 1980s in obtaining real-time updates and coverage of measurements where ground-based sources of information are not available [Chang et al. 1987; Kelly et al. 2003; Derksen et al. 2010]. Sensors aboard Earth observation satellites are capable of acquiring the strength of reflection and radiation at multiple wavelengths. In terms of snow remote sensing measurements, the sensor type is typically divided into: (1) an optical sensor or (2) a microwave sensor. The former type onboard the satellite is often used to map areal distribution of snow (i.e., snow cover extent), whereas the latter is often used to map snow depth (or SWE).

Since microwaves possess the capability to penetrate deep (the depth of penetration depends on the frequency of microwaves) into the snowpack and to be less affected by vegetation compared with that of shorter wavelengths (e.g., visible

radiation), PMW radiometers are capable of quantifying volumetric storage of snow water (snow depth or SWE) retrieved from Tb [Ulaby and Stiles, 1980]. In other words, the measured Tb contains important information about snow states. Hence, the development of the remote sensing technique is intended to extract useful information, such as snowpack-related properties, from the electromagnetic signal recorded by the space-borne antenna [Foster et al. 1987]. In other words, the PMW remote sensing technique is introduced to establish a relationship between the electromagnetic feature and the physical feature of the target (i.e., snow). SWE retrieval products based on PMW Tb measurements from space-based microwave radiometers such as the Special Sensor Microwave/Imager (SSM/I) [Chang et al. 1982], the Scanning Multichannel Microwave Radiometer (SMMR) [Chang et al. 1987], and AMSR-E [Kelly et al. 2004] have played significant roles in estimating SWE at basin scales. The following study focuses on the utilization of AMSR-E measurements. However, it is hypothesized that the machine learning techniques explored here are equally applicable to both SMM/I and SMMR Tb measurements.

2.3. PROBLEMS WITH EXISTING SNOW PARAMETER ESTIMATION

There are typically three ways to estimate important snow-related properties from space-borne sensors. One of the methods is to merge relatively coarser space-borne observations with in-situ measurements of finer resolutions by spatial interpolation [Cao et al. 2008]. However, this is significantly impacted by sparse spatial coverage of observations particularly in northern regions [Takala et al. 2011] and strong sub-grid scale snow variability in complex terrain (e.g., mountains) [Foppa

et al. 2007]. The second technique is to invert (or retrieve) model states variables from measured Tb at certain frequencies by calibrating regression coefficients in the algorithm. These selected calibrated snow retrieval products are further discussed below.

Chang et al. [1986] presented the first snow depth-Tb relationship for a uniform snowfield with a fixed snow density of 300 kg/m^3 and a mean radius of 0.3mm, which was expressed as:

$$D = 1.59\times(T_{18,H}-T_{37,H}) \quad (2.3\text{-}1)$$

where D is the snow depth [cm]; $T_{18,H}$ denotes the Tb [K] at 18 GHz horizontal polarization; and $T_{37,H}$ is the Tb [K] at 37 GHz horizontal polarization.

Goodingson and Walker [1994] derived another commonly used form of the relationship between SWE and Tb for dry snow as:

$$SWE = a+b(T_{37,V}-T_{19,V}) \quad (2.3\text{-}2)$$

where SWE is the snow water equivalent [mm]; a and b are fixed parameters; a = -20.7 [m], b=-2.74 [K^{-1}]; $T_{37,V}$ is the Tb [K] at 37 GHz vertical polarization; and $T_{19,V}$ is the Tb [K] at 19 GHz vertical polarization.

Kelly et al. [2003] coupled the snow grain radius and volumetric fraction data with a radiative transfer model to estimate snow depth based on SMM/I data at a constant snow temperature of 260 [K] using the following expression:

$$D = b(T_{19,V}-T_{37,V})^2+c(T_{19,V}-T_{37,V}) \quad (2.3\text{-}3)$$

where D is the snow depth [cm]; b and c are coefficients related to the ratio of snow grain size and the volume fraction; $T_{19,V}$ is the Tb [K] at 19 GHz vertical polarization; and $T_{37,V}$ is the Tb [K] at 37 GHz vertical polarization.

Besides snow grain size, forest cover is another important factor to take into consideration in every snow retrieval algorithm [Tedesco and Narvekar 2010]. Overlying vegetation will attenuate the PMW radiation emitted from the underlying snowpack and at the same time, it will add on its own contribution to the signal as measured by the radiometer [Derksen et al. 2005]. Chang et al. [1996] tried to improve the SWE estimation in the forested regions and then came up with another revised form of the algorithm:

$$SWE = \frac{a(T_{19,V}-T_{37,V})}{(1-ff)} \tag{2.3-4}$$

where SWE is the snow water equivalent [mm]; a is a calibration coefficient [dimensionless]; ff is the forest fraction [dimensionless] ranging from 0 to 0.75 [Kelly, 2009]; $T_{19,V}$ is the Tb [K] at 19 GHz vertical polarization; and $T_{37,V}$ is the Tb [K] at 37 GHz vertical polarization.

For the current AMSR-E algorithm, the following expression for calculating snow depth for both forested and non-forested regions [Kelly, 2009] is:

$$D = ff \times \left[p1 \times \frac{(T_{18,V}-T_{36,V})}{(1-b\times fd)}\right](1-ff)\times[p1(T_{10,V}-T_{36,V})+p2(T_{10,V}-T_{18,V})] \tag{2.3-5}$$

where D is the snow depth [cm]; ff is the vegetation fraction [dimensionless]; fd is the forest density; p1 and p2 are two dynamic coefficients ranging from 1 to 2; b is a regression coefficient; $T_{10,V}$ is the Tb [K] at 10 GHz vertical polarization; $T_{36,V}$ is the Tb [K] at 36 GHz vertical polarization; and $T_{18,V}$ is the Tb [K] at 18 GHz vertical polarization.

Certain assumptions, such as uniform snow grain size and constant snow density, have to be made in order to use these empirical equations, many of which are not

reasonable in the real system. Additionally, significant uncertainties are commonly found in space-borne PMW SWE retrievals that impact their estimation accuracy. For example, snow stratigraphy can result in highly nonlinear scattering processes that complicate snow depth estimation [Durand et al. 2011]. Snow grain size is another important (and difficult to characterize) parameter in snow retrieval products that impacts snow albedo [Armstrong et al. 1993]. It is also well known that the increase in depth hoar layer (large loose and cup-like snow grains [Brucker et al. 2011]) thickness will decrease microwave emission [Hall 1987], which will cause measured Tb to decrease. Ice crusts on the surface and within the snowpack also alter the absorption and emission of microwave radiation from the surface by increasing the emissivity at high frequencies relative to low frequencies [Derksen et al. 2010]. However, snow morphology [Kelly et al. 2003] and depth-hoar/ice layer studies [Hall et al. 1986; Foster et al. 2005] have not matured enough for operational use by water resources managers.

Further, wet snow behaves like a blackbody (perfect absorber for all incident radiation [Siegel and Howell, 1992]) at the physical temperature of the snow layer, which makes it hard to distinguish from snow-free soil [Scherer et al. 2005]. Signal saturation for very deep snow (greater than 150mm SWE) can lead to large biases in SWE estimation [Clifford 2010]. In addition, model inputs on snow-related state estimates obtained from land surface hydrologic models (e.g., Variable Infiltration Capacity Model) may contain errors associated with model structure and model parameterization [Andreadis and Lettenmaier, 2006]. Meteorological fields used to force the physical- or empirical-based land surface models may have some

uncertainties such as scaling effects arising from dataset aggregation, disaggregation, extrapolation and interpolation [Blöschl and Sivapalan, 1999].

In an effort to overcome the limitations of the existing satellite-based snow retrieval algorithms, the third alternative of merging measurements of remote sensing observations with estimates from land surface or physical snow models [Reichle 2008] is proposed in SWE/snow depth estimation. Namely, a data assimilation (DA) technique is often implemented to merge measurements with model estimates by weighing their uncertainties, which is anticipated to yield a merged estimate of snow characteristics that is superior to either the measurement or the model alone [Mclaughlin 2002].

Radiative transfer models (RTMs) are widely used [Liang et al. 2008] by researchers to invert PMW Tb measurements into model state variables coupled with a physical snow model in the DA framework [Durand and Margulis, 2007]. The practical utilization of these algorithms is plagued by the complex spatiotemporal uncertainty [Pullianinen et al. 1999] coupled with wet, moderately deep snowpacks (greater than the 100mm) located closer than 200 km to open water [Dong et al. 2007] and the effects of mixed land cover within remotely sensed pixels [Andreadis et al. 2008]. In addition, the complicated inversion of PMW Tb measurements is computationally expensive at regional or continental scales [Durand and Margulis, 2006]. These are the factors that limit the existing PMW SWE retrievals within the DA framework to point-scale or basin-scale applications [Durand et al. 2008].

Therefore, the uncertainties and limitations mentioned above in the existing snow properties characterization motivates the study proposed here to further

investigate another alternative approach of estimating SWE/snow depth at a large-scale (discussed in more detail in Section 2.5).

2.4. INTRODUCTION OF MACHINE LEARNING

Arthur Samuel [1959] first defined machine learning (a.k.a. data mining or supervised learning) as a field of study that gives computers the ability to learn without being explicitly programmed. Another more specific definition is the process of identifying a set of categories (sub-populations) where a new observation belongs on the basis of a training set of data containing observations whose category membership is known [Hastie et al. 2005].

Machine learning, which indicates that the procedure requires analyst-labeled training, develops characteristic class signatures that are then used to assign labels to all other unassigned areas ("unseen" model inputs areas) in the model framework [Campbell, 2002]. It is different from unsupervised algorithms that are self-organizing, iterative models capable of finding "natural" data clusters [Campbell, 2002]. It commonly refers to a field of study about how to automatically learn, acquire and generalize information based on these known examples so as to make accurate predictions in the future.

Machine learning aims to generate classifying/regression expressions and functions simple enough to be understood by a human [Michie et al. 1994]. Unlike traditional statistical approaches, which are characterized by having an explicit underlying probability model, machine learning is an attractive tool in the fields of Web search, spam filters, stock trading and drug design [Domingos, 2012].

There is a plethora of machine learning algorithms to choose from depending on what type of question needed to be addressed. Sections 2.4.1 and 2.4.2 discuss the basics of the ANN and SVM. Reasons for selecting these two techniques will be discussed in Chapter 3.

2.4.1. Artificial Neural Network (ANN)

An artificial neural network (ANN) is a mathematical model inspired by biological neural networks (i.e., human brains). An ANN consists of a series of layers: (1) an input layer of neurons used for receiving information outside the network, (2) one or more hidden layer(s) acting as a bridge to connect the input layer with the output layer with input and output signals remaining within the network, and (3) an output layer to send the data out of the network. The ANN proposed for this study is a feed-forward perceptron network. Without any feedback connections, the signal could only flow in one direction: from the defined input layer to the hidden layer and subsequently propagate into the output layer [Atkinson and Tatnall, 1997].

In a constructed ANN, each layer contains multiple processing units (i.e., neurons) connecting with those in the adjacent (previous and subsequent) layers. An independent weight is attached to each link as indicated by the arrows in between the layers as illustrated in **Figure 2.4-1**. The input to each neuron in the next layer is the sum of all its incoming connection weights multiplied by their connecting input neural activation value [Rojas, 1996; Tedesco et al. 2004]. In general, it is assumed that each processing unit provides an additive contribution to the connected output neuron, which may take on the form as:

$$x_j = \sum_{i=1}^{N_i} w_{ji} I_i \qquad (2.4\text{-}1)$$

where x_j is a single value (a.k.a. "net input" [Bishop, 1995]), calculated via combining all the connected input units for the j^{th} propagated (output) unit; N_i is the total number of inputs; $w_{ji}$is the interconnection weight between the i^{th} input neuron and the j^{th} propagated neuron; and I_i is the i^{th} model input.

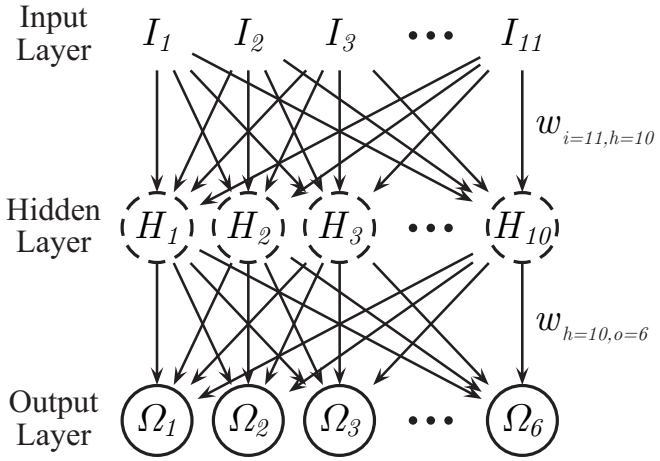

Figure 2.4-1 Schematic of the ANN-based model used in the study [Forman et al. 2013] with 11 model inputs in the ANN input layer, ten (10) hidden neurons, and six(6) model outputs of Tb measurement (see Chapter 3).

The power of an ANN lies in its ability to perform intelligent tasks via applications of different types of neural network algorithms for both unsupervised and supervised learning. It is also one of the highly recommended tools for non-linear statistical data modeling since it has the ability to detect complex non-linear interactions between the input and output neurons [Svozil et al. 1997]. However, an ANN is often referred to as a "black box" algorithm, which indicates it is difficult to gain a thorough understanding and explicitly explain the physical basis behind its performance. In addition, sometimes parameters derived based on ANN learning regularities may not be physically meaningful. Further, an ANN typically requires a

large number of parameters to be tested and established before successful application of the model. Perhaps the greatest shortcoming of the ANN is that it may converge to a local minimum point instead of a global minimum. As illustrated in **Figure 2.4-2**, if the initial estimate happens to fall into the region between a local maximum and a local minimum, then it is likely that the back propagation will stop at the local minimum without searching for other regions possibly with a lower objective function of the mean squared error (MSE) (see Chapter 3). Finally, it is also worth noting that this type of robust machine learning technique is computationally expensive and requires high processing time and numerous iteration steps for solving a complex non-linear model for a large study domain.

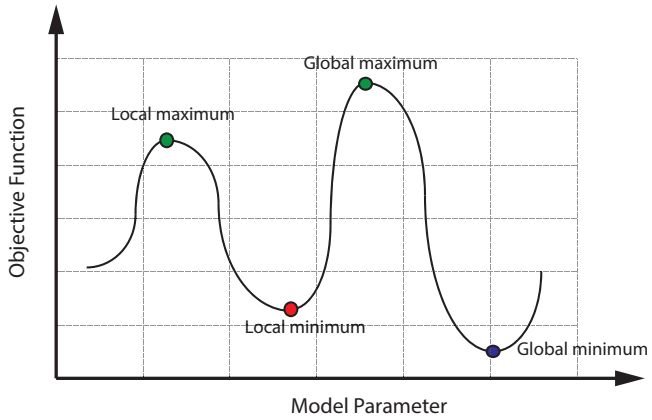

Figure 2.4-2 An example of local minima and global minima in ANN framework in terms of model parameter selection. The red dot is the local minima and it falls into the valley consisting of a local maxima and a global maxima (green dots). The blue dot represents the global minima, which is the optimal target for the minimization procedure.

2.4.2. Support Vector Machine (SVM)

Since the 1980's, machine-learning techniques including decision trees and neural networks have begun to allow for efficient learning of non-linear decision surfaces while achieving reasonable predictive performance. However, as discussed above, it

is difficult for the ANN-based model to explicitly explain (in a physically-based manner) how best to connect the input layer, hidden layer, and output layer with each other using the specified weights. Therefore, Vapnik et al. [1998] proposed another efficient learning algorithm for non-linear functions based on the statistical/computational learning theory called Support Vector Machine (SVM).

Consider an input matrix, **x**, and a vector of training targets, **z**, such that $\{(\mathbf{x_1}, \mathbf{z}_1),\ldots, (\mathbf{x_p}, \mathbf{z}_p)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $\mathbf{z}_i \in \mathbb{R}^m$. A schematic of the SVM framework can be seen in **Figure 2.4-3**. It is assumed that $\phi(\mathbf{x})$ is a nonlinear function that maps the model input space into a feature space. f is a function that is a linear combination of the components of the input $\phi(\mathbf{x})$ such that:

$$f(\phi(\mathbf{x})) = \mathbf{w}^T \phi(\mathbf{x}) + b \text{ with } b \in \mathbb{R} \qquad (2.4\text{-}2)$$

where **w** is a vector of weights and b is an offset (a.k.a., bias) term. Both **w** and b are determined by the SVM during training.

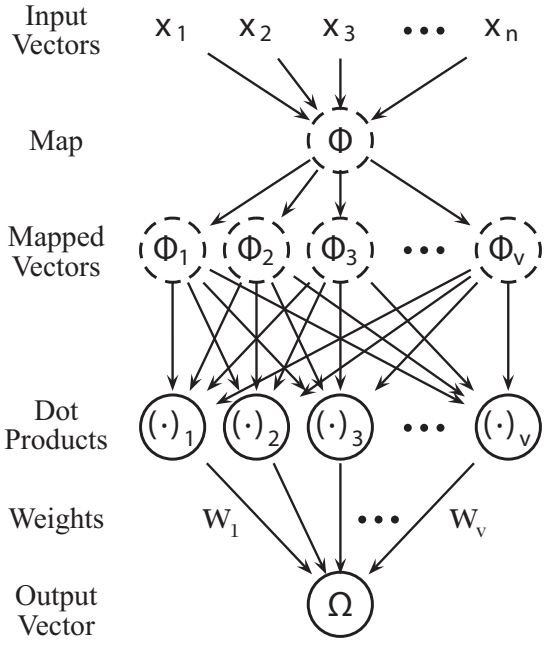

Figure 2.4-3 Schematic of the SVM-based model [Forman and Reichle 2014].

2.4.2.1. Support Vectors

Based on the structural risk minimization concept, the SVM for a given location requires the solution of the following convex optimization problem [Smola and Ikopf, 2004]:

$$\underset{w,b,\xi}{\text{Minimize}} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w}+C\sum_{i=1}^{p}\xi_i \text{ with } C>0$$

$$\text{subject to } \mathbf{w}^T\phi(\mathbf{x_i})_{+b}-\mathbf{z_i}\leq\xi_i, \tag{2.4-3}$$

$$\xi_i \geq 0,\ i=1,2,\ldots,p.$$

where **w** is a vector of weights for a given location in space; p is the total number of measurements in time; $\mathbf{z}_i$ is the set of training targets at time i; ξ is defined as a slack variable, which is intended to relax the constraints to allow outliers to exist or to be misclassified; and C is a trade-off constant (a.k.a., penalty parameter) of the error term.

The weighting vector is defined as [Burges, 1998]:

$$\mathbf{w} = \sum a_i\mathbf{z}_i\phi(\mathbf{x_i}) \tag{2.4-4}$$

$$\text{where, } a_i > 0$$

Therefore, training points with $a_i>0$ are defined as the "support vectors" [Chang et al. 2010]. These support vectors define the decision space to determine the model function.

Parameter C determines how much penalty is to be given for those allowed misclassified points. If C is set to infinity (or a very large number), the number of permitted outliers approaches zero. This "free-of-error" requirement is difficult to achieve since measurement products are not error-free. Hence, a large value of C

would often "overfit" the model with a large number of support vectors, which is not desirable in terms of the computational efficiency as well as the physical expression of the model form. Meanwhile, a small value of C will "underfit" the model with an overly simple model function. Therefore, it is critical to choose a reasonable value of C, to be neither too big nor small, such that there can be enough flexibility for the optimization equation to find its best solution. To find the optimal C, SVM-users are required to vary its value across a wide range and search for the best C by cross-validation.

The objective function (see Equation 2.4-3) is known as a quadratic program (QP) with linear constraints [Potschka et al. 2010]. Time complexity of the original QP often depends on the dimensionality of the target **z** [Fletcher, 1998]. However, such QP problems can be solved more easily in its dual formulation utilizing Lagrange multipliers [Chang and Lin, 2011] where the temporal complexity will be decreased to the number of training examples, which is the key for extending the SVM to better handle nonlinear models. The solution can be written as [Weston 1998]:

$$\underset{w,\,b}{\text{Minimize}} \quad \underbrace{\frac{1}{2}\Big\| \sum_{i=1}^{p} a_i \phi(\mathbf{x_i}) \Big\|^2}_{\text{maximize margin}} + \underbrace{C \sum_i [\mathbf{z}_i (f(\phi(\mathbf{x_i})))]}_{\text{minimize training error}} \tag{2.4-5}$$

where $\|\cdot\|$ is the Euclidean norm operator. Alternatively, it can also be written as [Smola and Ikopf, 2004]:

$$\underset{a_i,\, a_i^*}{\text{Maximize}} \quad \{-\frac{1}{2}\sum_{i,j=1}^{p} (a_i - a_i^*)(a_j - a_j^*) \langle \phi(\mathbf{x_i}).\phi(\mathbf{x_j}) \rangle - \sum_{i=1}^{p} \mathbf{z}_i (a_i - a_i^*)\} \tag{2.4-6}$$

$$\text{subject to } \sum_{i=1}^{p}(a_i - a_i^*) = 0,$$

$$a_i, a_i^* \in [0, C],\ i = 1, 2, \ldots, p$$

As the expression above indicates, the slack variable vanishes from the dual form with only a constant C coefficient modifying the error term where a_i, a_i^* are Lagrange multipliers; $\langle \phi(\mathbf{x_i}) \cdot \phi(\mathbf{x_j}) \rangle$ is the inner (dot) product of $\phi(\mathbf{x_i})$ and $\phi(\mathbf{x_j})$; $\mathbf{x_i}$ and $\mathbf{x_j}$ are two sets of training points; and C is the penalty parameter discussed above.

2.4.2.2. Kernel Functions

Recall that the dual formation of the optimization problem depends on the computation of the form $\langle \phi(\mathbf{x_i}) \cdot \phi(\mathbf{x_j}) \rangle$ where $\mathbf{x_i}$ and $\mathbf{x_j}$ are two sets of training points. The inner (dot) products could be computed in feature space only when the SVM has simpler forms of the mapping function ϕ. Therefore, another technique called "kernel function" (a function of two variables) was used in this study [Chang and Lin, 2011].

In this study, the kernel function is defined as:

$$k(\mathbf{x_i}, \mathbf{x_j}) = \langle \phi(\mathbf{x_i}) \cdot \phi(\mathbf{x_j}) \rangle \qquad (2.4\text{-}7)$$

Hence, the computation was conducted in feature space using the kernel function without explicitly computing $\phi(\mathbf{x})$ or the weighting vector $\mathbf{w}$. Otherwise, the dimensionality of $\phi(\mathbf{x})$ can be very large thereby making $\mathbf{w}$ difficult to represent explicitly in memory and even more difficult for the QP to solve [Weston, 1998].

There are four types of commonly used kernels in both linear and non-linear classification and regression models: (1) linear kernel, the simplest kernel function, which is given by the dot product of the form $\Phi(x_i) \cdot \Phi(x_j)$ with an optional constant c, where the linear kernel usually has the form of $k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) + c$, (2) polynomial kernel that has three parameters - slope parameter p, polynomial degree

of q ($q \in \mathbb{N}$), and a constant c where $c \geq 0$ such that the polynomial kernel has a functional form of $k(x_i, x_j)=\{p[\Phi(x_i)\cdot\Phi(x_j)] + c\}^q$, (3) hyperbolic tangent (sigmoid) kernel (a.k.a. multiplayer perceptron kernel), which can be expressed as:

$$k(x_i, x_j)= \tanh(p\Phi(x_i)\cdot\Phi(x_j)+ c) \tag{2.4-8}$$

with two adjustable parameters in the sigmoid kernel, the slope p and the intercept constant c, and (4) gaussian radial basis function (RBF) kernel.

In this study, an RBF was employed, which is one of the most commonly-used kernel functions. The Gaussian kernel is an example of a radial basis function kernel written as:

$$k(x_i, x_j)=\exp(-\gamma||\Phi(x_i)\cdot\Phi(x_j)||^2) \tag{2.4-9}$$

where $||\Phi(x_i)\cdot\Phi(x_j)||$ represents the Euclidean norm between $\Phi(x_i)$ and $\Phi(x_j)$; and $\gamma > 0$ is an adjustable parameter crucial in the performance of the kernel. It controls the width of the Gaussian distribution and plays a similar role as the degree of the polynomial kernel [Ben-Hur and Weston, 2010]. If γ is overestimated, the exponential function will behave almost linearly and the high-dimensional projection will lose its non-linear power. On the other hand, if γ is underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in the training data [Souza, 2010].

Based on a properly constructed SVM with optimized parameters (see Chapter 3), the SVM has been widely used in different types of classification, regression and pattern distribution estimation. Reasons are as follows: (1) SVMs are able to perform well at regression analysis under either nonlinearity or high dimensionality conditions since the SVM maps the non-linearly separable data into a feature space of higher

dimension where it is linearly separable, (2) SVMs provide a good out-of-sample generalization if the key parameters (e.g., penalty parameter C and adjustable parameter γ in the RBF) are selected properly [Hsu et al. 2003]. Hence, the SVM is a robust algorithm, which is anticipated to work well even when the training examples contain errors, (3) unlike an ANN framework, formulations of SVMs are convex optimization problems and thereby unique global optima will be found and the algorithm will not be affected by the local minima issue, and (4) generally, SVMs can avoid the overfitting issue effectively by implementing the cross-validation method [Hsu et al. 2003] or through Bayesian regulation of the hyper-plane parameters proposed by Cawley and Talbot [2007]. In addition, a form of "early-stopping" [Sarle,1995] can be implemented to prevent overfitting resulting from the direct optimization of the marginal likelihood until convergence [Cawley and Talbot, 2007]. Further SVMs are expected to work well even in cases where limited training data is available since the decision surface of a SVM is comprised of support vectors, which is far less than the number of training data.

However, every machine learning technique has its limitations. Limitations in the SVM approach include: (1) SVMs are sensitive to significant outliers, especially for those playing maximal roles in determining the decision hyper-plane [Xu et al. 2006]; (2) SVMs can be expensive to apply in terms of both computational time and memory. Procedures such as the "grid-search" method used in the LIBSVM (see Chapter 3) can be employed to locate parameters to use during the inner dot product computations.

In summary, based on properly-constructed systems, machine learning algorithms are capable of learning about the regularities present in the training data such that constructing and generalizing rules can be extended to the unknown data [Mathur et al. 2004] during the training phase (see Chapter 3).

2.5. MACHINE LEARNING IN SNOW RETREIVAL

Initial attempts of investigating the possibility of employing a machine learning technique, instead of a RTM, in estimating snow properties were conducted by few studies [Chang and Tsang, 1992; Tsang et al. 1992; Davis et al. 1993; Tedesco et al. 2004; Cao et al. 2008]. They focused on utilizing an ANN to "learn" the pattern of the SWE estimation from a physical snow model or in-situ snow measurements and then try to use this "prior" information to predict SWE in other areas for comparison against observations. Good agreement was obtained from these test areas, such as the Antarctic region [Tsang et al. 1992]; however, these applications are limited to relatively small areas. Additional studies made use of ANN to acquire information from the ground-based measurements [Tedesco et al. 2004]. However, this is not preferred either, since the ANN could not acquire enough information to establish connections between sparsely located stations.

With the eventual goal of SWE or other snow-related properties retrieval, recent research conducted by Forman et al. [2013]; and Forman and Reichle [2014] investigated the possibility of directly estimating Tb's by utilizing machine learning methods of either ANN or SVM. It was concluded that both the ANN and SVM could be used as measurement operators to estimate Tb's for eventual use within the data

assimilation framework for the purpose of SWE estimation at regional and continental scales.

However, we still need to answer some fundamental questions: Do the ANN and SVM reproduce Tb for the right (physically-based) reasons? Further, what are the most significant parameter(s) in the model using either ANN or SVM? In response to these questions, the goal of this study is to compare and contrast sensitivity analysis results between ANN- and SVM-based estimates.

CHAPTER 3: MODEL FORMULATION

The following chapter describes the model inputs and outputs required for use in the ANN and SVM framework. It also discusses how to choose model parameters, how to train the ANN or SVM, and how to conduct cross-validation.

3.1. NETWORK INPUTS

The NASA Catchment land surface model (Catchment) is the land surface component of the NASA Global Modeling and Assimilation Office Land Data Assimilation System (GMAO-LDAS) whose basic computational unit is the hydrological catchment (or watershed) [Koster et al., 2000]. The model encompasses an explicit treatment of spatial variation of snow by dividing the snowpack into three layers including estimation of snow density, snow temperature, SWE, and snow liquid water content (SLWC).

All model inputs to both the ANN and SVM are provided by the land surface state estimates derived from the Catchment model and are listed in **Table 3.1-1** except for the model parameter temperature gradient index (TGI) [Josberger and Mognard, 2002]. TGI is generally defined as the difference between the near-surface soil temperature and the near-surface air temperature divided by the snow depth as:

$$TGI(l) = \frac{1}{C}\int_0^l \frac{T_{p1}(t) - T_{air}(t)}{D(t)}\,dt \qquad (3.1\text{-}1)$$

where t is time at a daily scale; l is the span of time of interest; C is a scaling constant of 20 K $m^{-1}day^{-1}$[Armstrong, 1985; Colbeck, 1987]; T_{p1} is the near-surface soil

temperature [K]; T_{air} is the near-surface air temperature [K]; and D is the snow depth [m]. Armstrong [1985] and Colbeck [1987] showed that thermal gradient metamorphism plays a dominant role within the snowpack in producing different sizes of snow grains. In response, TGI could serve as a proxy for snow grain size in both the ANN- and SVM-model input system as computed from the Catchment output.

Table 3.1-1 Model inputs and output for both ANN and SVM.

Inputs	Symbol	Unit
Top layer snow density	ρ_{sn1}	kg/m^3
Middle layer snow density	ρ_{sn2}	kg/m^3
Bottom layer snow density	ρ_{sn3}	kg/m^3
Snow liquid water content *	SLWC	kg/m^2
Snow water equivalent *	SWE	m
Near-surface air temperature	T_{air}	K
Near-surface soil temperature	T_{p1}	K
Skin Temperature	T_{skin}	K
Top layer snow temperature	T_{sn1}	K
Bottom layer snow temperature	T_{sn3}	K
Temperature gradient index	TGI	-
Outputs	**Symbol**	**Unit**
Brightness Temperature at 10.65 GHz, H-polarization	10H	K
Brightness Temperature at 10.65 GHz, V-polarization	10V	K
Brightness Temperature at 18.7 GHz, H-polarization	18H	K
Brightness Temperature at 18.7 GHz, V-polarization	18V	K
Brightness Temperature at 36.5 GHz, H-polarization	36H	K
Brightness Temperature at 36.5 GHz, V-polarization	36V	K

* denotes column-integrated quantities

Meteorological fields (e.g., precipitation, humidity and wind speed/direction) used to force the Catchment model are derived from the Modern-Era Retrospective Analysis for Research and Applications (MERRA) product [Rienecker et al. 2011].

The MERRA data record spans 1979 through the present. MERRA outputs are produced at 1-hour intervals with a 1/2 degrees latitude × 2/3 degrees longitude × 72 vertical levels model configuration extending through the stratosphere [Rienecker et al. 2011].

In this study, the daily-averaged Catchment outputs were remapped on the Equal Area Scalable Earth Grid (EASE-Grid). These grids have a nominal cell size of 25km × 25km and are provided by the National Snow and Ice Data Center (NSIDC). The EASE-Grid features an equal-area projection, and thus there is no shape distortion at the poles while the greatest shape distortion occurs at the equator.

3.2. STUDY DOMAIN

The study domain used here includes all of North America poleward of 32°N, which allows for both regional and continental scale investigations. It includes the period 1 September 2002 through 1 September 2011, which is the coincident time period for all of the data sources to be used in this study. For simplicity, since glaciers are not the focus of this paper, locations such as south-central Alaska, which extends from the Alaska Peninsula to the border of the Yukon Territory in Canada, are of secondary interest for the SWE estimation in this study.

The continent is surrounded by the Arctic Ocean to the north, the Atlantic Ocean to the east, the Pacific Ocean to the west and south, and the Caribbean Sea to the southeast, which made it possible for the domain to embrace all types of climatic zones and vegetation cover. Because of the highly dynamic variation of spatial

climatology, the domain embraces all types of major snow classes --- tundra, taiga, maritime, prairie, alpine, and ephemeral shown in **Figure 1.1-2**.

The study utilizes a percent tree cover product by Hansen et al. [2011] based on the dataset from Moderate Resolution Imaging Spectroradiometer (MODIS). The tree cover product has a resolution of 500m × 500m. It is generated using a supervised regression tree algorithm (**Figure 3.2-1**). For purposes of this study, the original product was re-mapped as forest cover fraction onto the 25km EASE-Grid. About one third of North America is forested [Aaron et al. 2013], which will greatly impact SWE estimation via PMW emission [Langlois et al. 2011]. Without considering the effects of changes in biotic disturbances and other climatic aspects, this study assumes that the forest cover percentage is relatively constant across the time period of investigation.

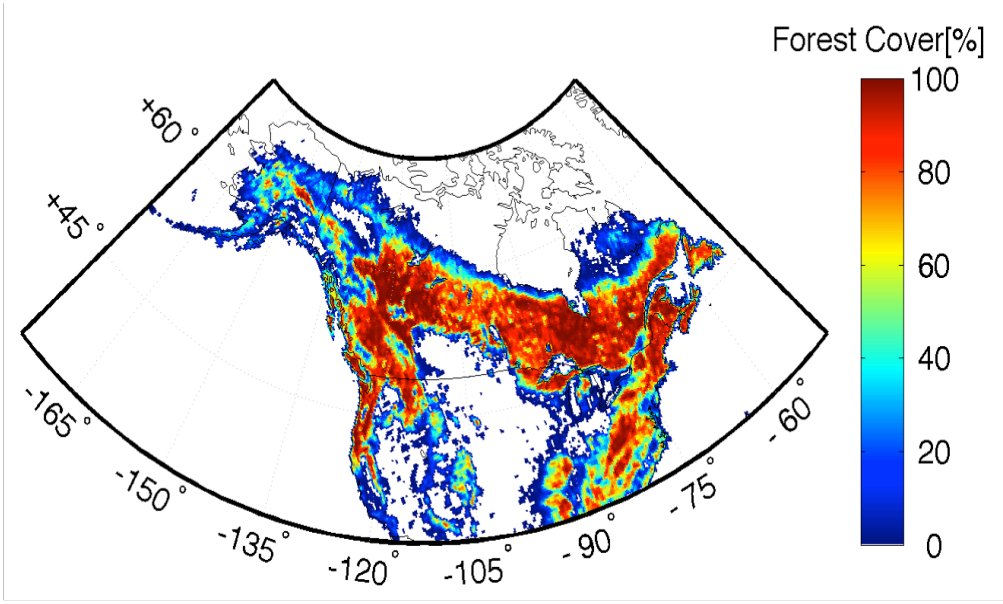

Figure 3.2-1 Forest cover across the North America.

3.3. MACHINE LEARNING IN LARGE-SCALE SWE ESTIMATION

In this study, relevant snow, land surface, and atmospheric states derived from the Catchment model are used as inputs to both the ANN and SVM frameworks. The goal of using machine learning is to model the complex relationships between these model inputs (including snow-related state variables) and the measured Tb outputs.

3.3.1. ANN Framework

Model input space may have different units as well as a wide range of magnitudes. For example, in this study, Tb's are in a reasonable range of [150K, 300K], whereas the SWE input is varying between 0m and 2m. Hence, except for each neuron Ω_m(m^{th} output neuron) in the output layer, most neurons in the ANN are required to transform their net inputs using a scalar-to-scalar function, which is called the activation function [Bishop, 1995].

Activation functions are bounded and can take on various forms, such as a binary step function, sigmoid function, threshold function, and hyperbolic function. The selection of the activation function form is dependent on the problem itself. In this study, activation functions for the hidden units are utilized to introduce more non-linearity into the network associated with nonlinear hydrologic and electromagnetic processes related to SWE estimation. The activation function f(x) employed in this study is the tangent (non-linear) sigmoid function, which can be expressed as:

$$f(x) = \frac{2}{1+e^{-2x}} - 1. \qquad (3.3\text{-}1)$$

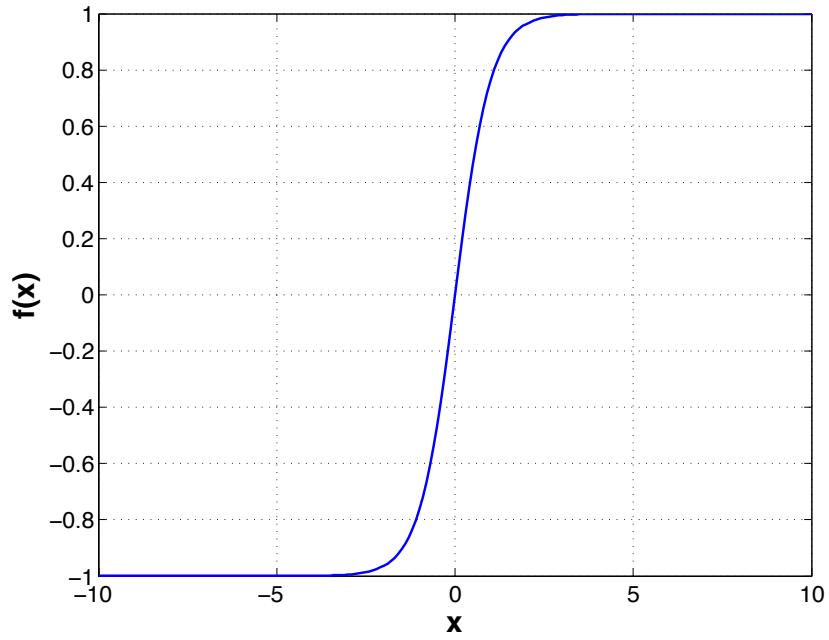

Figure 3.3-1 Tangent sigmoid function.

The activation function used at the output of each neuron (except for the ones in the output layer) has a range of [-1,+1], which are dimensionless. The output units of the mapping network are supposed to have appropriate (e.g., Tb $\in [150\text{K}, 300\text{K}]$) target values instead of arbitrary values between -1 and 1. Hence, the activation function g(x) for estimating the state of output neurons has to be a positive, linear transfer function. The mapped space of Tb will be produced after being rescaled to the proper target Tb with units of K.

The selection of the number of hidden layers and the number of neurons in each hidden layer is critical in constructing an ANN. The number of neurons in the hidden layer must be large enough to form a decision region that is as complex as required by the problem, but not so large that the weights cannot be reliably estimated from the training data [Lippmann. 1987]. For the one hidden-layer-based ANN used in this

study, the number of hidden neurons is determined by the following equation [Cao et al. 2008; Forman et al. 2013]:

$$N_h=\lceil\sqrt{N_i+N_o}+5\rceil \quad (3.3\text{-}2)$$

where N_i is the number of inputs; N_o is the number of model outputs; N_h is the unknown number of hidden neurons; and $\lceil\cdot\rceil$ is the integer ceiling of the expression. This study has 11 model inputs derived from the Catchment model output, and thus N_i is 11. The generated network output from a trained-ANN (see Section 3.3.2.1) based on AMSR-E measurements includes Tb at 10.65 GHz, 18.7 GHz, and 36.5 GHz at both horizontal and vertical polarizations, as shown in **Table 3.1-1** (additional details provided Section 3.3.2.1.). Accordingly, there are six (6) model outputs of multi-frequency and multi-polarized Tb (i.e., N_o= 6). Therefore, the number of hidden neurons is N_h=10.

3.3.2. ANN Training

3.3.2.1. ANN Training Targets

The AMSR-E instrument on the NASA's Earth Observing System (EOS) Aqua satellite provides global PMW measurements of the Earth from 19 June 2002 to 27 September 2011 with a swath width of 1445 km. Tb's (in tenths of kelvins) at 6.9 GHz, 10.65 GHz, 18.7 GHz, 23.9 GHz 36.5 GHz, and 89.0 GHz at both horizontal and vertical polarization are measured. The spatial resolution of the raw data varies with frequency since the sensors requires a minimum number of photons in order to record a single signal at a single frequency. Hence, measuring Tb at 6.9 GHz (lowest

energy photons) has the coarsest resolution at 56 km and the 89 GHz (highest energy photons) possesses the finest resolution at 5.4 km.

In this paper, Tb measurements from the gridded Level-3 land surface product (AE_Land3) were utilized as the training targets. The data were available twice a day from descending (night) and ascending (day) overpasses and are made available by the NSIDC [Knowles et al. 2006]. However, only measurements from nighttime (approximately 01:00 to 01:30 hours local time) AMSR-E overpasses were employed to minimize the effects of liquid water present in the snow [Forman et al. 2013]. Data are stored in the Hierarchical Data Format - Earth Observing System (HDF-EOS format) and resampled into global cylindrical EASE-Grid cell spacing at a 25km × 25km horizontal resolution [Knowles et al. 2006], the same grid used for the Catchment output.

Not all of the channels (frequencies) are used in this study. The 6.9 GHz channel was not used in the study because it has a spatial resolution of 75km × 43km at 3-dB footprint size, which is much coarser than the remapped EASE-Grid. However, higher frequency channels with finer spatial resolution, such as the 89 GHz, is often designed for atmospheric observation [Chang and Tsang, 1992] and largely affected by water vapor and clouds [Mätzler, 1994]. In addition, it is more sensitive to surface properties of snow (e.g., surface grain size) than to the snow depth [Durand et al., 2008; Durand and Margulis, 2007]. Thus, the 89 GHz channel is not optimal for SWE estimation. The 23.9 GHz channel is also avoided being used in this study since it is strongly impacted by atmospheric water vapor [Pampaloni, 2000].

As suggested by Kelly [2009], moderate depth snow can be derived from the spectral difference between 10.65 GHz and 36.5 GHz and the calculation of deeper snow depth/SWE is based on vertically polarized Tb at 10.65 GHz and 18.7 GHz. Therefore, the ANN is trained with satellite observations from AMSR-E in the study domain from 1 September 2002 to 1 September 2011 (total time period of nine years) for both vertically polarized and horizontally polarized Tb at 10.65 GHz, 18.7 GHz and 36.5 GHz.

3.3.2.2. ANN Training Approach

This study utilizes the Neural Network Toolbox provided by Matlab© to independently generate a neural network system for each location in space. Due to its high efficiency in performing matrix calculations, Matlab© is an ideal tool for working with ANNs. Details about the working principle of the Toolbox are discussed below.

First of all, it is necessary to sort out what are the appropriate locations with enough valid and relatively accurate information related to network input (e.g., snow-related information) for the ANN to "learn". In order to minimize erroneous inputs to the ANN framework, the model utilizes the National Oceanic and Atmospheric Administration (NOAA) Interactive Multisensor Snow and Ice Mapping System (IMS) product [Helfrich et al. 2007] to verify the model inputs derived from the Catchment model [Forman et al. 2013]. As a result, the ANN-based framework is able to ensure the presence of snow as simulated by Catchment for each cell/grid in the study domain. Snow cover as predicted by the Catchment model for a given pixel is deemed reasonable if the IMS product at the same time indicates the existence of

snow. After remapping the IMS product from its native 24 km resolution onto the EASE-Grid (25km × 25km), this study utilizes the post-processing IMS map to act as the "truth" in snow cover detections and to compare with the occurrence of snow as modeled by Catchment. In response, Forman et al. [2013] pointed out that the agreement between the Catchment model output and IMS snow cover extent is good with the hit ratio of 0.88 across the NA for the nine (9) years investigated.

The ANN training was conducted based on the back-propagation learning cycle to minimize the MSE (see Section 2.4.1) between the ANN-estimated Tb and the AMSR-E Tb training target value. For example, in terms of a single location for a given time period, we are given a training set $\{(\mathbf{I_1}, \mathbf{\Omega_1}),\ldots(\mathbf{I_p}, \mathbf{\Omega_p})\}$ consisting of p pairs of input space **I** and output training space **Ω** using the same time period from all of the available years except for the pre-defined validation year. During training, the MSE for a single output neuron can be computed using the following equation:

$$\mathrm{MSE} = \frac{1}{2}\sum_{i=1}^{p} \|\mathbf{\Lambda}_i - \mathbf{\Omega}_i\|^2 \qquad (3.3\text{-}3)$$

where Λ_i is the i^{th} ANN-estimated value of Tb [K]; Ω_i is the i^{th} value of the AMSR-E training target Tb [K]; p is the total number of evaluated time steps; and $\|\cdot\|$ represents the Euclidean norm operator between the estimated (ANN-derived) Tb and the measured (AMSR-E collected) Tb.

Since the output of a neuron depends on the weighted sum of all its inputs, the back propagation method is employed and aims to find a set of weights that could minimize the errors [Rojas, 1996]. To start the minimization algorithm, the initial weights applied in between the input and output neurons are randomly selected. After that, the Levenberg-Marquardt optimization algorithm [Levenberg, 1944; Marquardt,

1963] is applied iteratively to update the weights until the MSE achieves its minimum for each output neuron. In other words, the back propagation method aims to calculate the gradient of the error of the network with respect to the network's modifiable weights to quickly converge on its satisfactory local minima [Baboo and Shereef, 2010].

Based on a suitable training algorithm and a well-constructed neural network, the accuracy of the training result will be improved as more training dataset are made available. However, due to the enormous computational expenses, we divide these 9-year-span AMSR-E measurements into several parts with sufficient model inputs information for faster processing speed, as well as for the purpose of capturing the seasonality of the snow properties.

The ANN is trained separately for each fortnight (two-week period) of each year. Further, each location (cell) in the NA domain has its own unique ANN for a particular fortnight. Reasons for selecting a fortnight, rather than a week or a month as the basic training period, are discussed in Forman et al. [2013]. It was shown that a one-month training period cannot adequately capture the temporal variability of AMSR-E targets whereas a one-week-period size of AMSR-E measurements did not provide a sufficiently large enough training dataset. Therefore, a fortnightly training period was eventually selected to address the strong seasonality in the snow process while also providing a sufficient training data size for use during training activities.

In order to assess the accuracy of the trained-ANN outputs, a validation approach called "Jackknifing" [McCuen, 2005] (a.k.a. leave-one-out) was used in the study of Forman et al. [2013]. Each time the study withholds one-year of Tb from AMSR-E

to be used later as the independent validation dataset, with the remaining eight-year dataset of Tb measurements is used as training data. The validation results based on different model accuracy assessment statistics (e.g., bias, root mean squared error and anomaly correlation coefficient), can be seen in Forman et al. [2013], which demonstrated that the Tb estimations based on the ANN agree well with the AMSR-E measurements in the NA domain across the nine-year time period.

3.3.3. SVM Framework

In the context of this study, the input space of **x** incorporates 11 variables that characterize the snow properties and near-surface conditions governing the energy exchange in between the atmosphere and the snow pack. The inputs used in the SVM are identical to those used by the ANN. The training targets of **z** are the multi-polarized Tb at 10.65 GHz, 18.7 GHz and 36.5 GHz based on the satellite measurements.

It is assumed that $\phi(\mathbf{x})$ is a nonlinear function that maps the geophysical inputs from the land surface model, **x,** into Tb space [Forman and Reichle, 2014]. This study defines the C, the penalty parameter (see Chapter 1) as the range of the training targets, which can be written as [Mattera and Haykin, 1999]:

$$C = \max\{\mathbf{z}\} - \min\{\mathbf{z}\} \tag{3.3-4}$$

An alternate formulation was tested in Forman and Reichle [2014] using $C = 6\,\sigma_z$, where σ_z is the standard deviation of the training targets. It was suggested that there are not significant differences between using different C parameter ranges.

Hsu et al. [2003] found that employing exponentially growing sequences of γ, the adjustable parameter (see Chapter 1), is a practical method for identifying reasonable values for the parameter. Initially, this study defines γ as:

$$\gamma=2^{-7}, 2^{-6}, 2^{-5}, \ldots, 2^{5}, 2^{6}, 2^{7} \quad (3.3\text{-}5)$$

Parameter selection is an important technique in training SVM problems since model users are supposed to construct the SVM framework by first defining a set of parameters. The SVM utilized in this study adopts a "grid-search" technique in order to locate the "best" penalty parameter C and RBF parameter γ. In the context of this study, a 6×15 grid was pre-defined to test various pairs of (C, γ) values. The one with the best cross-validation accuracy (see Section 3.3.4.2.) was selected. This type of exhaustive parameter search can be parallelized since each (C, γ) is independent from one another, and therefore computational time can be reduced [Hsu et al. 2003].

3.3.4. SVM Training

3.3.4.1. SVM Training Targets

The SVM is trained with the satellite-based observations obtained from AMSR-E for both horizontally and vertically polarized Tb measurements at 10.65GHz, 18.7GHz and 36.5GHz assessed from 1 September 2002 to 1 September 2011, which are exactly the same training targets as the ANN used.

3.3.4.2. SVM Training Approach

The LIBSVM library, a library for Support Vector Machines (SVMs), provided by the National Taiwan University, was employed for SVM training in this study. LIBSVM is currently one of the most widely used software in classification,

regression, and learning tasks [Chang and Lin, 2011]. The LIBSVM provides users with various types of SVM formulations, QP solutions with different constraints, performance measurement metrics, and possible solutions to unbalanced data classification and regression.

Before further discussion on SVM training, it is worthwhile to first highlight several steps that are essential to efficiently improving the SVM-based model performance:

- *Step I: Quality Control*

  Similar to the ANN-based model, the SVM-based framework used the same IMS product to validate the accuracy of model inputs in both space and time before allowing the SVM to "learn" from the information collected in the model inputs.

- *Step II: Input Scaling (a.k.a. normalization or standardization)*

  Scaling before applying the SVM learning algorithm is important [Ben-Hur and Weston, 2010] since large margin regression algorithms are sensitive to the way features are scaled. In this study, there are a total of 11 geophysical variables and each of them is measured in a different scale with a different unit and has a different range of possible values. It is often beneficial to scale all features to a common range [Ben-Hur and Weston, 2010] such that attributes in greater numeric ranges will not dominate those in smaller ranges [Hsu et al. 2003]. Another advantage of scaling is to avoid numerical difficulties in calculating inner products of feature vectors [Hsu et al. 2003].

The scaling method used in this study can be described by illustrating the following example in terms of scaling SWE data for a location within a fortnight training period time, and the standardization algorithm used in this study can be written as:

$$\hat{x}_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \times (b-a) + a \qquad (3.3\text{-}6)$$

where $\hat{x}_i$ is nominal SWE [dimensionless] after scaling; x_i is the original input value of SWE [m] at time i; min (x) is the minimum SWE [m] input value across this fortnight training period; max (x) is the maximum SWE [m] derived from the Catchment model for the specified fortnight; a is the specified lower bound of the scaling range; and b is the upper bound of the defined range of scaling. Alternatively, the scaling can be performed onto the model input space as the example shown above, and also the projected higher dimensional feature space (or at the level of the kernel function itself).

In defining the scaling intervals, Sarle [1997] concluded the two most useful ways to standardize inputs. One of them is to scale the data with the mean of zero and the standard deviation of one, and the other method is to have a scaled dataset with the midrange of 0 and the range of 2 (i.e., [-1,1]). However, Hsu et al. [2003] recommended SVM users to linearly scale each attribute of the model input to the range of [-1, +1] or [0, 1].

This study randomly selected five (5) places spread out across the study domain and then trained the SVM by using the scaling intervals of [-1, 1], [0, 1], [0.1, 1.1], [0, 2], [0.5, 1], [1,2], [1, 3], and [0.5, 1.5], respectively. The

results demonstrate that there are no significant differences between the SVM-based models and these scaling intervals in terms of Tb predictions.

However, there are significant differences in terms of the computation of the Normalized Sensitivity Coefficients (NSCs) of different model states (See Chapter 5). SVM models using scaling intervals of [0.5, 1], [1,2], [1,3] and [0.5, 1.5] produce almost the same numeric value of NSCs. Since NSCs are computed in the post-scaled space, the Tb nominal value, which should also be in the range of the defined interval $[a, b]$ (3.3-), functions as the denominator based on the NSC calculation (Equation (4.2-1). When the Tb nominal value approaches zero, the NSC is close to infinity (or a very large number), which is not desirable. This explains why the tested scaling intervals with either midrange of zero or including zero (e.g., [-1, 1], [0, 1], [0.1, 1.1], [0, 2]) are not able to produce similar results of NSCs. Therefore, the SVM utilized in this paper defines the scaling interval with a lower bound of 1 and a higher bound of 2.

- *Step III: Cross-validation*

  As a standard technique for adjusting hyper-parameters (the parameters can not be automatically tuned by the learning algorithm and thus have to be tuned manually) of predictive models [Chan et al. 2013], v-fold cross-validation (**Figure 3.3-**) method is made available in LIBSVM. The v-fold cross-validation divides the training set into v subsets of equal size. Sequentially one subset is tested using the SVM model trained on the remaining (v-1) subsets [Hsu et al. 2003]. Afterwards, the cross-validation

accuracy is computed as the percentage of data that is correctly classified. In the context of SVM regression, the parameters with the minimum cross validation error are selected.

The study also compares the performance between different SVM models with various numbers of partitions, and the results suggest that there are negligible differences when the number of subsets (v) varies between 2 to 10. Hastie et al. [2009] suggested using five (5) or ten (10) as the number of partitions. In this study, v is set to 5 for cross-validation during the selection of model parameters C and γ.

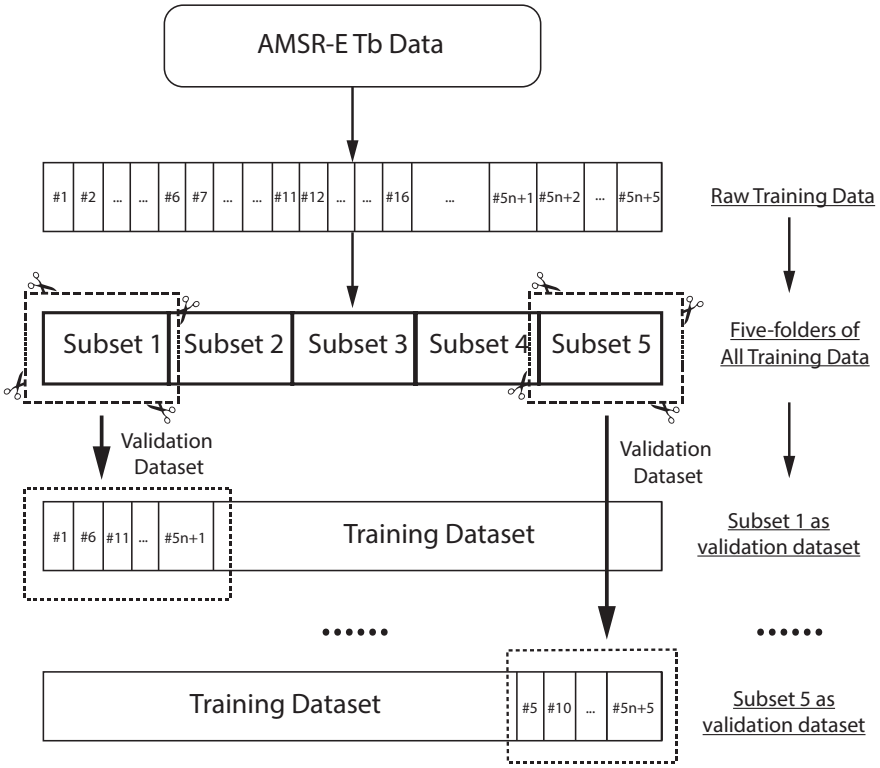

Figure 3.3-2 Cross-validation with five subsets.

Finally, the study trained the SVM model using all data points from the Catchment model output for each fortnight for each year (as discussed in Section 3.3.2.2.), and defined the optimal parameters pair (C, γ). It is also worth noting that a

rescaling metric is also needed before conducting Tb predictions in order to transform the normalized value of SVM output into the measurement space of Tb.

Goodness-of-fit statistics for assessing SVM-based model performance are provided in Forman and Reichle [2014]. It is concluded that the SVM possesses the capability to serve as a model operator within a DA framework for Tb predictions across large spatial scales. However, it is still unknown which parameter(s) in the model inputs is (are) relatively important compared to the others. The sensitivity analysis of the SVM-based model outputs with respect to different model inputs will be introduced in Chapter 4 for this purpose.

3.3.5. Similarities and differences between machine learning techniques

In summary, there are some similarities between SVMs and ANNs in that (1) they are data-driven models used when the underlying physical relationships are not fully understood [He et al. 2014], (2) they can be used to reproduce nonlinear processes [Baughman and Liu 1995; Suykens et al., 2001] as well as to solve noisy, black-box problems [Sjoberg et al. 1995] via iterations without prior knowledge about the relationships between the parameters [Živković *et al.* 2008], and (3) a SVM-based model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network [Souza, 2010], and thereby have similar performance in solving certain types of regression problems.

However, there are still some differences between these two types of machine learning. The existence of local minima [Smola and Schölkopf, 2004] would prevent an ANN from finding the unique global minimum solution to a constrained optimization problem, which is not the case for a SVM that possesses a more simple

geometric interpretation that ultimately yields a sparse solution [Burges 1998]. Further, the efficiency of a neural network largely lies in the hidden layer of nodes [Tu 1996]. The selection of the number of neurons in the hidden layer and the number of hidden layers is a significant issue related to ANN performance. That is, a neural network with too many nodes will "overfit" data while too few hidden neurons will "underfit" the data [Fletcher et al. 1998]. For SVMs, support vectors serve as the function centers, which are calculated as the result of a QP procedure based on a RBF kernel [Valyon and Horváth, 2003]. Furthermore, when the model is associated with a large number of model states, the SVM would outperform the ANN [Byvatov et al. 2003] since the SVM approach does not attempt to control model complexity by keeping the number of features small [Rychetsky 2001]. Finally, if the size of the training examples are not large enough, the SVM is still expected to perform well based on a properly-selected mechanism of model parameters because the number of support vectors in the decision space is far less than the number of training points [Tsang et al. 2005] whereas ANN are always in need of a relatively large amount of training points.

CHAPTER 4: SENSITIVITY ANALYSIS FORMULATION

The following chapter discusses the importance of sensitivity analysis used in machine learning. It also analyzes the effects of different perturbation sizes to both models (either ANN- or SVM-based) based on their sensitivity results to model inputs. Further, an important metric - Normalized Sensitivity Coefficients (NSCs) is introduced to quantify the relative importance of model inputs.

4.1. SENSITIVITY ANALYSIS

Modeling is the process of simulating the real world. A typical modeling process consists of four elements, including: (1) model conceptualization, (2) model formulation, (3) model calibration, and (4) model verification [McCuen, 2002]. Sensitivity analysis, defined as the rate of change of one factor with respect to change in another factor [McCuen, 2002], is important in each of the modeling steps.

4.1.1. Importance of sensitivity analysis

Sensitivity analysis is important in model formulation. It is used to understand the behavior of the model, to validate the reasonable performance of the model with the physical response of the real system, to evaluate the applicability of the model, and to determine the stability and rationality of the model [Yao, 2003].

Sensitivity analysis is important in model calibration. A complex model system is always dependent upon numerous model parameters. To look into the objective function (e.g., minimization of the root mean squared error) of the model with respect

to each calibrated coefficient in the response surface is important in order to validate which parameter(s) has/have converged to the optimum. Hence, insensitive parameters can be removed to simplify the model and save computational expenses. Insensitive parameters have large standard errors, so their lack of accuracy can contribute to the overall error of the model. In addition, obtaining an understanding of the sensitivity of the model output to the calibrated coefficients is essential to model optimization.

The sensitivity of model outputs is important in model verification. It can be used to determine which model component causes greater change in the model output. It can also show the effects of uncertainties in the fitted model output with respect to model input errors.

4.1.2. Sensitivity analysis in machine learning

Sensitivity analysis is an important tool in machine learning in terms of assessing the relative importance of causative factors in the model. This is especially significant in the ANN-based model, which is often referred to as a "black box" [Tzeng and Ma, 2005]. ANN is a powerful learning tool; however, most of the time, users are not able to tell how the ANN "learns" from the input data and how the hidden layer establishes connections in between the input neurons and output neurons. Hence, the performance of the ANN cannot be consistently ensured [Tzeng and Ma, 2005]. Similarly, the SVM is constructed on the basis of the statistical learning theories and often performs well in solving various regression problems. However, it

is still unknown if the performance of the SVM-based model can be explained by the physical response of the real system.

Previous studies conducted by Forman et al. [2013] and Forman and Reichle [2014] concluded that both ANN and SVM could serve as computationally efficient measurement operators for data assimilation at the continental scale. As a follow-up to these previous studies, this study conducts the sensitivity analysis in the model verification phase to validate the response of either an ANN- or a SVM-based model with respect to small perturbations in model inputs and whether or not such small perturbations result in a physically-consistent response. The sensitivity analysis is conducted here to address the following questions: What is the physical rationale for the relatively accurate predictions based on machine learning techniques [Forman et al. 2013; Forman and Reichle, 2014]? What is/are the most significant parameter(s) among all of the 11 geophysical variables in the model inputs derived from the Catchment model using either SVM or ANN? Is it SWE? Or is it due to non-snow-related quantities?

Recall that ANN and SVM have the same model inputs of 11 snow related and near-surface-related conditions and six multi-frequency, multi-polarized Tb's as model outputs; hence, the study conducted here is able to compare and contrast the sensitivity of Tb to each model input, respectively, between these two different machine learning techniques.

4.2. NORMALIZED SENSITIVITY COEFFICIENT

In accordance with different goals that a sensitivity analysis will achieve in each modeling phase, Isukapalli [1999] generally categorized sensitivity analysis methods into three categories: (1) variation of parameters, (2) domain-wide sensitivity analysis, and (3) local sensitivity analysis. The local sensitivity analysis method, whose focus is on estimates of model sensitivity to input variation in the vicinity of a sample point [Isukapalli, 1999], is utilized in this study. It is often dependent on the computation of gradient or partial derivatives at the nominal value [Yao, 2003].

There are three types of local sensitivity indicators: (1) absolute sensitivity, (2) deviation sensitivity, and (3) relative sensitivity [McCuen, 2002]. In this study, relative sensitivity is mainly used to quantify the relative importance of each model input parameter. The main advantage of the relative sensitivity analysis is its dimensionlessness, which makes it available to compare the response within a model between different model inputs as well as between different models.

The Normalized Sensitivity Coefficients (NSCs) [Willis and Yeh, 1987], of each model input (state) parameter is calculated as:

$$NSC_{i,j}=\left(\frac{\partial M_j}{\partial p_i}\right)\cdot\left(\frac{p_i^0}{M_j^0}\right)\approx\left(\frac{M_j^i-M_j^0}{\Delta p_i}\right)\cdot\left(\frac{p_i^0}{M_j^0}\right) \qquad (4.2\text{-}1)$$

where, p_i^0 is the initial parameter value; M_j^0 is the initial metric value; M_j^i is the perturbed metric value; Δp_i is the amount of perturbation; i = 1, 2, …, n (n is the number of parameters); and j = 1, 2, …, m (m is the number of metrics). In this study, the NSC is computed with respect to each individual Tb frequency. For instance, if p_1^0 is the input top layer snow density derived from Catchment at a given location on 01

Jan 2004, M_2^0 is the ANN- or SVM-based model output of the vertically-polarized Tb at 10.65 GHz given the same location and time, Δp_1 is defined by the model user, which is the perturbed amount of snow density (e.g., 5% or 10% of the nominal snow density), and M_2^1 is the re-computed model output of the Tb at 10.65 GHz with the perturbed snow density (while the other ten model inputs remain unchanged) as the model input, then $NSC_{1,2}$ is interpreted as the expected relative change in the estimated vertically-polarized Tb at 10.65GHz based on that model (either ANN- or SVM-based) given a 5% change in snow density.

The study perturbs only one input parameter at a time in order to calculate NSC for each model state parameter. As discussed above, the level of perturbation is pre-defined based on the feature of each model. The perturbation cannot be too small; otherwise, the model noise will be amplified, which leads to an overestimation of the NSC. In addition, the perturbation cannot be too large, otherwise, the model will fall into the nonlinear region where the marginal function (i.e., slope of a line tangent to the curve) evaluated at the given point is no longer the representation of the rate of change in the model output with respect to the change in the input. The too-large-perturbation effects will be even worse when it falls into a strongly non-linear region where the difference between the marginal function and the "truth" on the curve is relatively large. Therefore, the model requires a linear response in the metric over a "small perturbation range". A perturbation size of +/-5% has been shown to be appropriate to obtain the linear response.

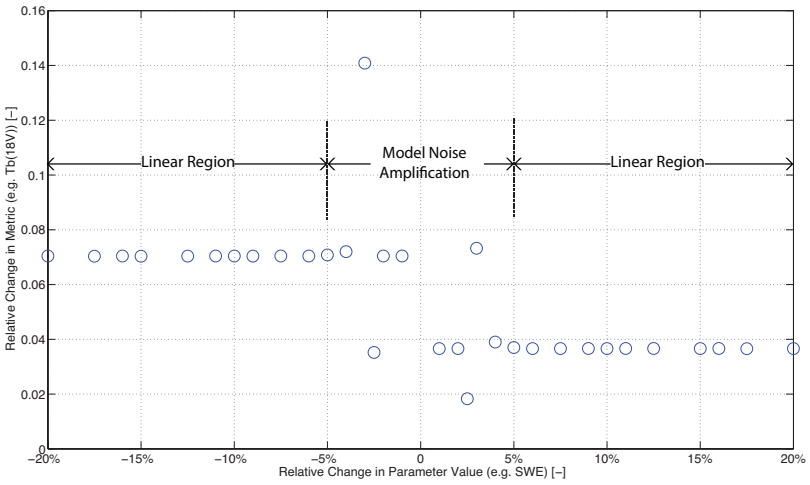

Figure 4.2-1 Perturbation effects in the sensitivity analysis of the ANN model.

Figure 4.2-1 demonstrates that when the relative change in daily SWE varies from -20% to +20%, the relative change in the ANN-based model output of the vertically polarized Tb at 18.7 GHz will be in a range from 0 to 15%. If the perturbation size of the model state (e.g., SWE) is too small, varying from -4% to 5%, the relative change in Tb is very large, which amplifies the noise instead of representing the real system response. It is also worth noting that, in the linear region, when the relative SWE value changes from -20% to -5%, there is almost no response in the relative Tb at 18 GHz for the ANN-based model at this given location. A preliminary assumption can be made that SWE might not be a sensitive parameter in the ANN-based model. Or perhaps the SWE has not shown its sensitivity at this selected location for assessing perturbation effects on this particular day. Hence, more details are still needed to be investigated about SWE in Chapter 5.

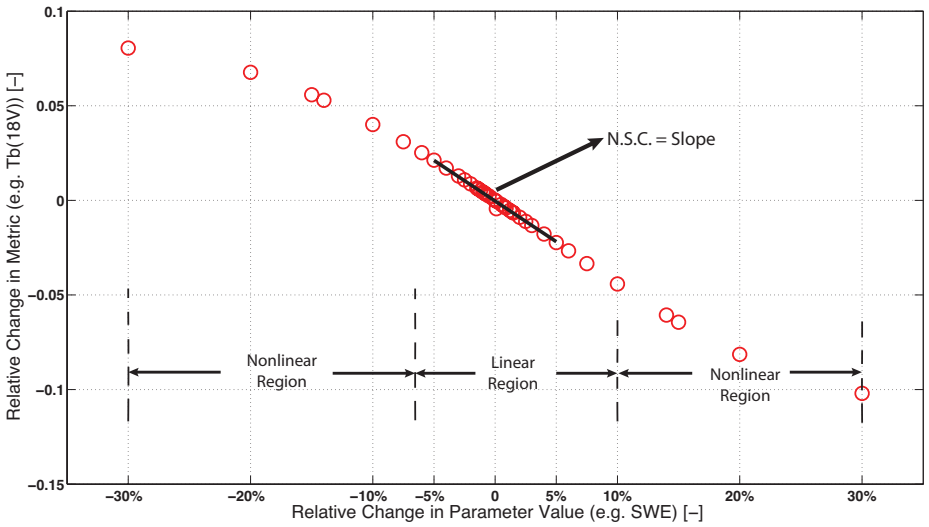

Figure 4.2-2 Perturbation effects in the sensitivity analysis of the SVM model.

Figure 4.2-2 demonstrates that when the relative change in daily SWE varies from -30% to +30%, the relative change in the SVM-based model output of the vertically polarized Tb at 18.7 GHz will be in a range from -10% to 8%. If the perturbation size of SWE is approximately in between -7% and 10%, the model falls into the linear region. Any perturbation size falling beyond the linear region would be invalid to reflect the real system response. There is almost no model amplification region in the SVM-based model of SWE state, with only a point "falling off" the NSC slope line. In the linear region, the ratio between the relative change in the Tb estimation at 18 GHz and the corresponding relative change in the metric (e.g., SWE), as interpreted as the gradient at the nominal metric value, is the physical interpretation of a NSC.

In this study, a perturbation size of +/-5% of the nominal model state variable is selected for all model state variables one-at-a-time in the NSC computation for both ANN- and SVM-based models. The model outputs for both ANN- and SVM-based models are the Tb predictions at both horizontal and vertical polarization at 10.65 GHz, 18.7 GHz and 36.5 GHz. The study mainly investigates the response of the

model outputs of vertically-polarized Tb estimations at 18.7 GHz and 36.5 GHz since these two combinations of frequencies are commonly used in SWE retrieval algorithms [Chang et al. 1986; Goodingson and Walker 1994; Kelly et al. 2003; Chang et al. 1996]. Additional details on the sensitivity analysis results regarding other model states are provided in Chapter 5.

4.3. SENSITIVITY ANALYSIS FORMULATION

Seven of 11 model input parameters were selected as the most sensitive model states (except for TGI, which is only for the comparison purpose with the SVM model) based on numerous NSCs calculations from 2002 to 2008 for both ANN- and SVM-based models. These seven selected model states are: (1) top-layer snow density, (2) SWE, (3) near-surface air temperature, (4) near-surface soil temperature, (5) skin temperature, (6) top layer snow temperature, and (7) TGI.

Since vegetation is one of the biggest challenges in accurate measurement of SWE-related Tb. As discussed in Chapter 2, in the areas covered with vegetation, the Tb measured by the satellite is a mixed signal from both snow cover and vegetation. At the same time, the overlying vegetation will tend to mask the signal coming from the underlying snow cover. In Chapter 5 of sensitivity analysis results, four scenarios are categorized for both ANN- and SVM-based models with various amounts of forest cover and SWE for a given day of interest.

The forest cover [%] values are obtained via the Hansen et al. [2011] forest product, which was derived from MODIS. The SWE [m] values are obtained from the land surface model. In general, these representative locations are selected for this

study because: (1) there is no sea ice found in this pixel, and (2) there is no significant lake fraction within the region (25 km × 25 km) even though the area may still be surrounded by some open water. In such cases, locations with percentages of vegetal cover greater than 50% are defined as “High Veg” areas, and those with vegetal cover less than 10% are defined as “Low Veg” class. For the specified day of interest, locations with SWE magnitudes greater than 0.15 m (~0.45 m snow depth) are categorized into “High SWE” class, while those with SWE values less than 0.04 m (~0.12 m snow depth) are defined as “Low SWE” areas.

CHAPTER 5: SENSITIVITY ANALYSIS RESULTS

In this chapter, the sensitivity results of both ANN- and SVM-based Tb estimations in terms of its spatiotemporal variability in forested and non-forested regions are presented. The following section will discuss the NSC computations of Tb of vertically polarized Tb at 18.7 GHz and 36.5 GHz under these four scenarios, respectively. Further, this chapter explains the reason for their differences in sensitivity to different model states. Year 2004 will be used as an example for demonstrating the sensitivity analysis since the 2004-2005-snow season is a fairly representative set of conditions during the 9-year study period.

5.1. SPATIAL VARIABILITY OF NSCS OF ANN-BASED MODEL

The study categorizes the NA domain under four scenarios with various amounts of forest cover and SWE for a given day of interest. The study selected one location in the study domain for each scenario as an example shown in the **Figure 5.1-1**.

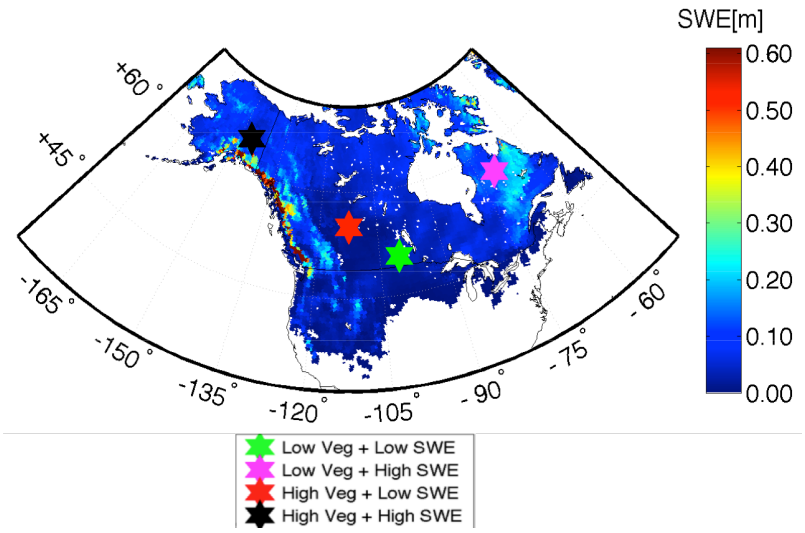

Figure 5.1-1 Examples of four locations (shown by markers with four different colors) with various amounts of SWE and vegetation on the SWE map in the NA domain on 14 Jan 2014.

Table 5.1-1 Canopy cover [%] and SWE [m] for the selected locations under different scenarios of various amounts of SWE (14 Jan 2004) and vegetation.

Scenarios	Canopy Cover [%]	SWE [m]
Low Veg + Low SWE	1.04	0.0323
Low Veg + High SWE	5.08	0.1625
High Veg + Low SWE	79.54	0.0152
High Veg + High SWE	67.74	0.1726

5.1.1. NSCs in the regions with low forest cover and low SWE

The representative location (latitude 50.4446° and longitude -100.7220°) of "Low SWE" and "Low Veg" class is in the southwestern corner of Manitoba, Canada, as indicated by the green marker in **Figure 5.1-1**. The NSCs of the top layer snow density, SWE, skin temperature, top layer snow temperature, and TGI are all zeros for both 18V and 36V because the snowpack is so shallow that most of the recorded signals are from the deep-layer snow or underlying soil. As indicated by **Table 5.1-2**, in such a case, the sensitivity of the near-surface air temperature and the skin

temperature are not exactly the same, the former one is zero and the latter one is 0.0085. This may arise from the effects of 5.04% forest cover in the 25 km × 25 km area.

Additionally, the signs of both NSCs at 18V and 36V are positive, which means that given an increase in the near-surface air temperature or soil temperature, there will be increase in Tb estimation at both microwave frequencies. This agrees well with the physical interpretation that Tb will increase as the physical temperature increases under the assumption that the emissivity remains the same for the object.

In terms of the magnitude (absolute value) of NSCs, the change in the near-surface soil temperature will result in a greater rate of change in Tb at 36V compared to that of 18V. More variation occurs in the temperature of the surface of the soil due to its frequent interactions with the overlying atmosphere, vegetation and snow, rather than with deeper layer of the soil. Hence, compared to the Tb at 18 GHz, the 36 GHz with a shorter wavelength can not penetrate as deeply into the snowpack, which will be more capable of capturing the variability of the near-surface soil temperature (or other soil-related properties, rather than snow). Therefore, the variation in the near-surface soil temperature will have more effects on vertically polarized Tb predictions at 36 GHz.

Table 5.1-2 NSCs computations on 14 Jan 2004 for seven model states in an area with low forest cover and low SWE.

Model states	NSCs of single Tb frequency	
	ANN (18V)*	ANN (36V)**
Top layer snow density	0	0
SWE	0	0
Near-surface air temperature	0.0823	0.2557
Near-surface soil temperature	0.0085	0.2107
Skin temperature	0	0
Top layer snow temperature	0	0
TGI	0	0

*: ANN (18V) denotes the vertically polarized ANN-based Tb at 18.7 GHz
**: ANN (36V) denotes the vertically polarized ANN-based Tb at 36.5 GHz
(same for other tables in this chapter)

5.1.2. NSCs in the regions with low forest cover and high SWE

The representative location (latitude 56.7349° and longitude -70.3197°) of "High SWE" and "Low Veg" class is in the northern part of Quebec, Canada, as indicated by the magenta marker in **Figure 5.1-1**. Except for TGI, the change in the other six model states exerts effects on the Tb estimation at both 18V and 36V as shown in **Table 5.1-3**.

In such cases, the SWE state plays a role in determining Tb. Based on the snow retrieval algorithm derived by Chang et al. [1996], if the vertically polarized Tb at 18V increases or the Tb at 36V decreases, the SWE will increase when the snow density is fixed. This could potentially explain the sign change between the NSC at 18V and 36V. However, it is still difficult to relate this sign-change issue of NSCs between different Tb frequencies with physical interpretations of snow. More

investigations are still needed in terms of fully understanding physical mechanism of radiation (i.e., microwave) interactions between snow, soil, air and vegetation.

Top layer snow density is as equally sensitive as the SWE state, which may be due to the physical relation between the snow density and the SWE. This is reasonable since Equation 1.1-2 demonstrates that snow density and SWE are connected via snow depth. According to Equation 4.2-1, since

$$NSC_{(SWE,\ Tb)}=\left(\frac{\Delta Tb}{\Delta SWE}\right)\cdot\left(\frac{SWE^{0}}{Tb^{0}}\right) \tag{5.1-1}$$

also,

$$SWE=\frac{D\times\rho_{snow}}{\rho_{water}} \tag{5.1-2}$$

thus,

$$\begin{aligned} NSC_{(SWE,\ Tb)} &= \left(\frac{\Delta Tb}{\frac{D\times\Delta\rho_{snow}}{\rho_{water}}}\right)\cdot\left(\frac{\frac{D\times\rho_{snow}^{0}}{\rho_{water}}}{Tb^{0}}\right) \\ &= \left(\frac{\Delta Tb}{\Delta\rho_{snow}}\right)\cdot\left(\frac{\rho_{snow}^{0}}{Tb^{0}}\right)=NSC_{(\rho_{snow},\ Tb)} \end{aligned} \tag{5.1-3}$$

where $NSC_{(SWE,\ Tb)}$ is the rate of change in Tb with respect to changes in SWE; ΔTb [K] is the increase or decrease in Tb estimation; ΔSWE [m] is the change in SWE magnitude, which is related to the defined perturbation size; SWE^{0} [m] is the nominal value of SWE before exerting any perturbations; Tb^{0} [K] is the nominal value of Tb; D is the snow depth [m], which remains the same during the calculation of NSCs (this is different in the SVM-based model, which will be discussed in the Section 5.2); ρ_{water} is the density of water $[\frac{kg}{m^3}]$, which is a constant; ρ_{snow}^{0} is the nominal value

of top-layer snow density [$\frac{kg}{m^3}$]; and $NSC_{(\rho_{snow},\ Tb)}$ is the relative change of Tb with respect to the perturbation in snow density. Hence, it seems that the top layer snow density should have the same performance with the SWE state, with the same ANN-based NSCs at both 18V and 36V. However, the Equation (5.1-3 can only be valid under the condition that the top-layer snow density has the same quantity as the column-integrated (three-layer-integrated) snow density. Hence, if the measured snow pack is uniform, the equivalent sensitivity derived from the ANN-based model between SWE and snow density is valid.

Table 5.1-3 NSCs computations on 14 Jan 2004 for seven model states in an area with low forest cover and high SWE.

Model states	NSCs of single Tb frequency	
	ANN (18V)	ANN (36V)
Top layer snow density	0.0491	-0.0069
SWE	0.0491	-0.0069
Near-surface air temperature	-0.0722	-0.0627
Near-surface soil temperature	0.4649	0.8492
Skin temperature	-0.0722	-0.0627
Top layer snow temperature	-0.0224	-0.0698
TGI	0	0

5.1.3. NSCs in the regions with high forest cover and low SWE

The representative location (latitude 55.0024° and longitude -112.7064°) of "High SWE" and "High Veg" class is in the middle of Alberta, Canada, as indicated by the red marker in **Figure 5.1-1.** The snow-related model states, such as top layer snow density, SWE and snow morphology proxy, TGI, are insensitive states in the ANN-based model under the scenario of high forest cover and low SWE. It is largely due to

the thick forest cover on the top of the shallow snow pack such that microwaves emitted from the underlying snow pack are significantly attenuated.

However, it is difficult to explain why the near-surface air temperature, skin temperature and the top-layer snow temperature have equal sensitivity in predicting Tb at both 18V and 36V. In the absence of vegetation, the skin temperature is expected to possess the same sensitivity as the top-layer snow temperature, whereas this location is covered with 79.54% forest. The disagreement with the physical fundamentals may come from: (1) model forcing error (e.g., precipitation and air temperature etc.), (2) measurement error associated with MODIS forest cover product, or (3) learning inability of the ANN in regions with high forest cover and relatively little snow. This learning inability may arise from ANN's learning algorithm in terms of converging to a local minima instead of the global minimum value of its objective function of mean squared errors.

Table 5.1-4 NSCs computations on 14 Jan 2004 for seven model states in an area with high forest cover and low SWE.

Model states	NSCs of single Tb frequency	
	ANN (18V)	ANN (36V)
Top layer snow density	0	0
SWE	0	0
Near-surface air temperature	0.1190	0.0702
Near-surface soil temperature	0.1444	0.2497
Skin temperature	0.1190	0.0702
Top layer snow temperature	0.1190	0.0702
TGI	0	0

5.1.4. NSCs in the regions with high forest cover and high SWE

The representative location (latitude 64.2750° and longitude -146.1695°) of "High SWE" and "High Veg" class is in the middle of Alaska, U.S. The model states of SWE, top layer snow density, and TGI do not exert their effects on Tb predictions. It might arise from the fact that high forest cover attenuates the emission of radiation from the snowpack prior to reaching the PMW sensor.

The positive signs of NSCs seem more reasonable under such a scenario that as the temperature of the near-surface air, or the soil, or the top-layer snow increases, the vertically-polarized Tb's at both 18 GHz and 36 GHz also increase.

Further, the sensitivity of the near-surface air temperature and the skin temperature are not exactly the same, which is largely due to the dense vegetation cover (67.74%) in this area.

Table 5.1-5 NSCs computations on 14 Jan 2004 for seven model states in an area with high forest cover and high SWE.

Model states	NSCs of single Tb frequency	
	ANN (18V)	SVM (36V)
Top layer snow density	0	0
SWE	0	0
Near-surface air temperature	0.4374	0.5200
Near-surface soil temperature	0.2277	0.2334
Skin temperature	0.3407	0.4099
Top layer snow temperature	0.1437	0.3853
TGI	0	0

5.2. TEMPORAL VARIABILITY OF NSCS OF ANN-BASED MODEL

A representative location (latitude 64.2750° and -146.1695°) in the middle of the Newfoundland and Labrador, Canada, was selected in the investigation of temporal variability of NSCs. It location is selected due to the sensitivity analysis results showed the NSC of SWE is non-zero when the investigated location is with relatively little vegetation cover and a relatively thick snowpack. It is also noticeable that the snowpack cannot be too thick (greater than 0.40m of SWE) - the deeper the snow depth is, the more scattering and attenuation will take place inside of the snowpack. Hence, the amount of energy emitted by a super thick (greater than 1.45m of snow depth) snowpack is still largely attenuated before reaching the sensor. Therefore, given that this location is covered by 6.06% forest with a maximum SWE of 0.22m during the snow accumulation phase and 0.24m during the snow ablation phase, it is suitable for a time series investigation of NSCs.

Model states of the SWE and the near-surface soil temperature are investigated in the temporal sensitivity analysis during both snow accumulation and ablation phases. The SWE state is selected for investigation since enhancing SWE estimation is the main objective in a future Tb assimilation. The examination of the NSCs time series is critical since the calculation of NSCs in Section 5.1 has suggested that: (1) SWE is not a relatively sensitive model parameter when using the ANN, and (2) ANN-based Tb prediction are most sensitive to soil temperature. In order to further verify this premise, a time series was needed to investigate whether the SWE is insensitive during either the snow accumulation or ablation phases for the ANN model.

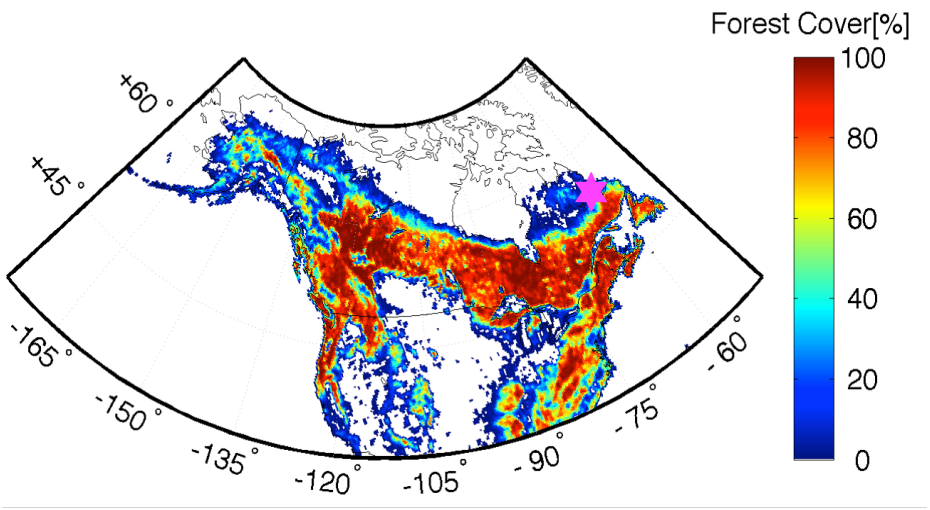

Figure 5.2-1 A selected location for the time series investigation of NSCs of different model states on the forest cover map in the NA domain.

5.2.1. Snow accumulation phase

During the snow accumulation phase (from 01 Jan 2004 to 10 Mar 2004), SWE increases from 0.12m to 0.22m and the near-surface soil temperature varies from 268.4K to 272.2K. As indicated by **Figure 5.2-2,** the soil temperature is not always decreasing or increasing. One the one hand, the overlying snowpack is behaving as a blanket covering on the top of the soil to keep the soil warm; on the other hand, the air temperature keeps decreasing and tends to cool the ground. Hence, the variation of the soil temperature contains the effects arising from both of the cooling and warming mechanisms.

The temporal NSCs results in the SWE and near-surface soil temperature states during the snow accumulation phase are also shown in **Figure 5.2-2**. The ANN-based Tb estimations at both 18 GHz and 36 GHz are sensitive to SWE on some days (five out of 72 days) during the snow accumulation phase in 2004. On the contrary, the near-surface soil temperature is a more sensitive parameter during this period.

Further, there are five days when the NSCs of the near-surface soil temperature are greater than one at 36 GHz, which means if there is a small change in the soil temperature, the Tb predictions will be altered significantly. These greater-than-one absolute values of NSCs might be explained by the physics that the near-surface soil temperature, whose depth is roughly equivalent to the penetration depth of 36 GHz microwave at 0.2 cm of the soil to the surface/ground. Hence, the ANN-based estimation of Tb contains more information about soil, rather than snow.

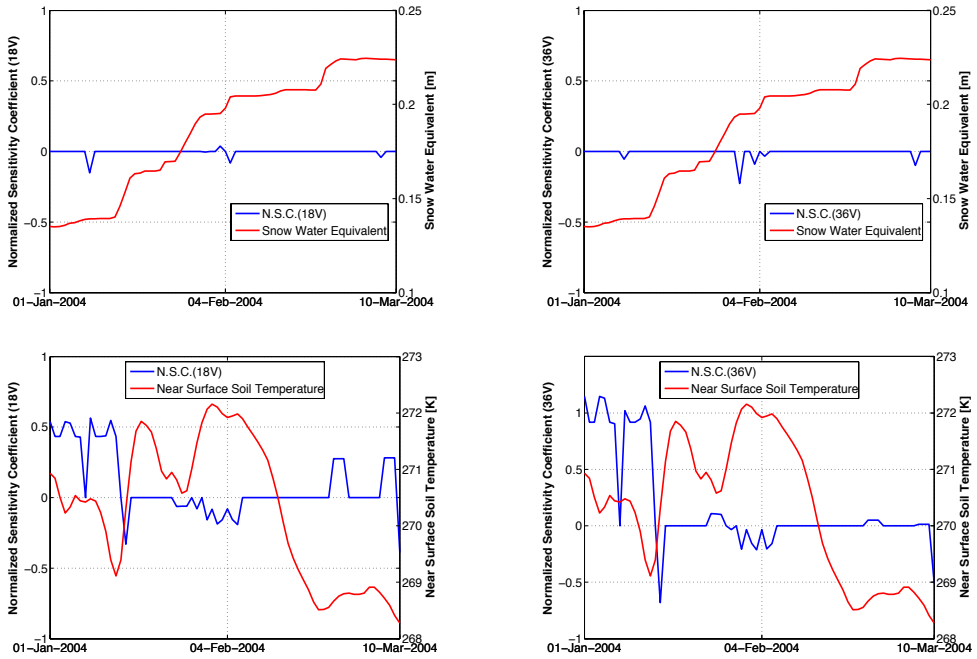

Figure 5.2-2 Time series investigation of ANN-based NSCs at both 18 GHz and 36 GHz Tb predictions of SWE and near-surface soil temperature from 01 Jan 2004 to 10 Mar 2004.

5.2.2. Snow ablation phase

During the snow ablation phase (from 25 Mar 2004 to 02 Jun 2004), SWE is still not sensitive for most of the time during the ANN-based predictions. Only two out of 72 days result in Tb estimations that are affected by a change in SWE. It is worth noting that on those two days, the NSC of SWE during the snow ablation state is roughly

eight times greater than that during the accumulation phase. This is possibly due to the presence of the liquid water within the snowpack, which significantly increases the absorption and emission of the microwave energy that results from the increase in the dielectric constant of the snow (see Chapter 1). However, melting snow may also increase the size of the snow grains relative to the microwave wavelengths used by the passive sensors due to a larger vapor pressure gradient during the ablation phase. More energy emitted by the snow may be scattered prior to reaching the sensor. Therefore, the greater sensitivity of SWE during snow melt is more likely to be a trade-off between an increase in the snow pack radiation absorptivity and a simultaneous increase in the snow grain size. In such a case, the effects induced by the presence of moisture within the snow pack likely takes a more dominant role than those from the snow grain size.

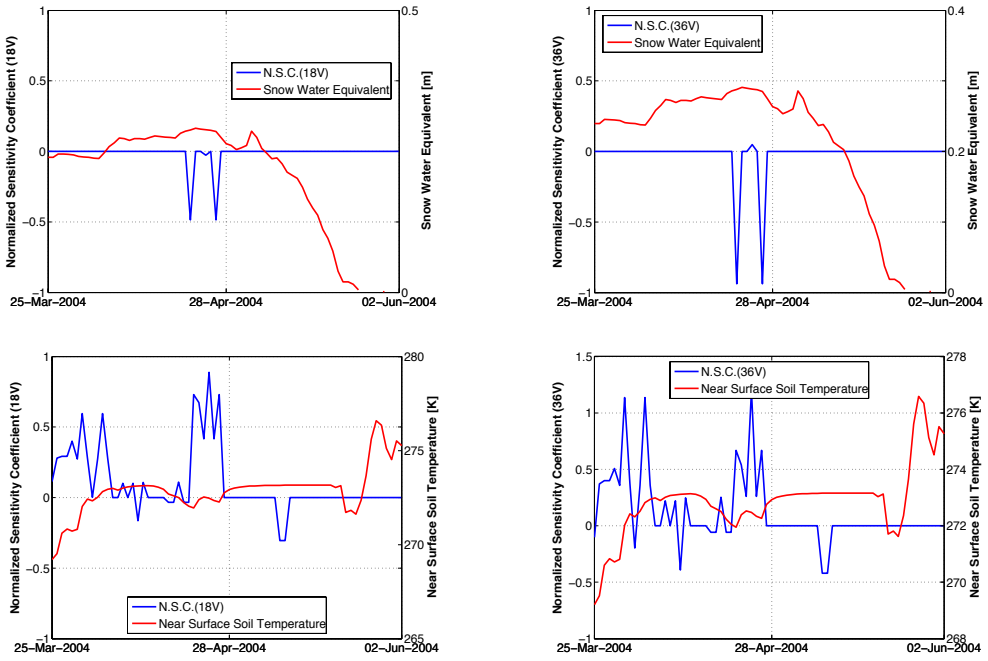

Figure 5.2-3 Time series investigation of ANN-based NSCs at both 18 GHz and 36 GHz Tb predictions of SWE and near-surface soil temperature from 25 Mar 2004 to 02 Jun 2004.

5.3. SPACIAL VARIABILITY OF NSCS OF SVM-BASED MODEL

The sensitivity analysis results of the ANN-based model are presented in Sections 5.1 and 5.2, where the preliminary finding is that SWE might not be the reason for accurate Tb predictions based on the ANN model, which is discouraging to some extent given the original intent of using ANN-derived Tb to update modeled SWE. This section will continue to explore the reason for the relatively accurate prediction of Tb based on the SVM model.

This section presents the differences between the ANN- and the SVM-based models for a given day on 11 Jan 2004 across the seven most sensitive states of the 11 model states. As discussed in Section 4.3, the study also divides the whole NA domain into four categories: (1) low vegetation with low SWE; (2) low vegetation with high SWE; (3) high vegetation with low SWE; and (4) high vegetation with high SWE. These specific locations within these four categories in the following section are selected differently from those in Section 5.1 of the ANN-based model analysis since the study is going to further verify if the insensitivity of SWE is highly dependent on location.

5.3.1. NSCs in the regions with low forest cover and low SWE

The first test location (latitude 50.4885° and longitude -100.3943°) of "Low SWE" and "Low Veg" class is in the southwest corner of Manitoba, Canada (see **Figure 5.3-1**). The forest cover percentage at that location is 5.04%, and the SWE value on 11 Jan 2004 was 0.03m, and therefore relatively little snow existed on that day.

Some similarities in the model performance were evident. For example, for both ANN- and SVM-based model, the NSCs of the skin temperature and the top layer snow temperature are the same since the area is only covered with 5.04% vegetation and 0.03m of SWE, hence the skin temperature is most representative of the top layer snow temperature. In addition, the near-surface soil temperature plays a role in both of the models based on the absolute value of the NSCs, whereas soil temperature is more sensitive in the ANN-based Tb predictions at 36 GHz, compared to that at 18 GHz. This is because a higher passive microwave frequency possesses a smaller emission depth, hence it captures more of the surface variability of the model state variables.

Some differences are still evident in the model behavior. The ANN-based model is not as sensitive to several snow-related states, such as SWE, top layer snow density, and top layer snow temperature in the presence of a shallow snowpack. However, Tb predictions at both 18 GHz and 36 GHz based on the SVM model are still sensitive to small perturbations in the snow states in the model inputs.

TGI, the snow grain size proxy, is the most sensitive state with the NSC value of 0.0781 in the SVM-based Tb estimation at 18 GHz. This is likely because there are some relatively large-size snow grains within the snowpack (or the presence of internal ice layers and/or ice crust), which behave as effective radiation scatters. Most of the scattered signals from the snowpack can still be recorded by the passive sensors due to less attenuation in the presence of low vegetation cover.

In Section 5.1, the study derived the Equation 5.1-3 of the NSC relationship between the top layer snow density and SWE, which does not hold true in the SVM-

based model. One of the interpretations might be that the change in SWE will possibly induce the change in snow depth as well (snow depth is not a constant after the perturbation of SWE) such that there is no guarantee that the sensitivity of the snow density and the SWE will always be the same. The other explanation is that the snow density in the Equation 5.1-2 is the column-integrated density, which is not necessarily the same as the top layer snow density in the model input when the uniform snowfield assumption is violated. In such case, the SVM-based NSC of the top layer snow density is more reasonable than that derived from the ANN-based model.

Table 5.3-1 NSCs computations on 11 Jan 2004 for seven model states in an area with low forest cover and low SWE.

Model states	NSCs of single Tb frequency			
	ANN (18V)*	ANN (36V)*	SVM (18V)**	SVM (36V)**
Top layer snow density	0	0	0.0377	0.1017
SWE	0	0	-0.0076	-0.0069
Near-surface air temperature	0.0662	0.2652	0.0036	0.006
Near-surface soil temperature	-0.0349	-0.5668	-0.0256	-0.1459
Skin temperature	0	0	0.0375	0.0939
Top layer snow temperature	0	0	0.0375	0.0940
TGI	0	0	0.0781	-0.0275

**: SVM (18V) denotes the vertically polarized Tb at 18.7 GHz based on the SVM model
**: SVM (36V) denotes the vertically polarized Tb at 36.5 GHz based on the SVM model
(same for other tables in this chapter)

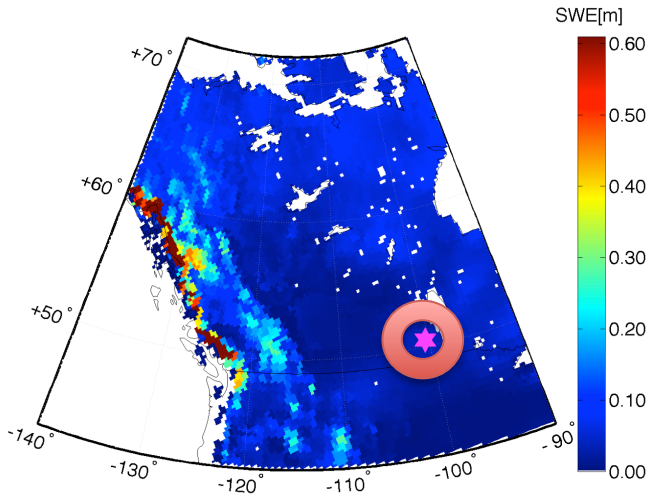

Figure 5.3-1 An example of a location (shown by the red circle) with low forest cover and low SWE value on the SWE map in the NA domain on 11 Jan 2004.

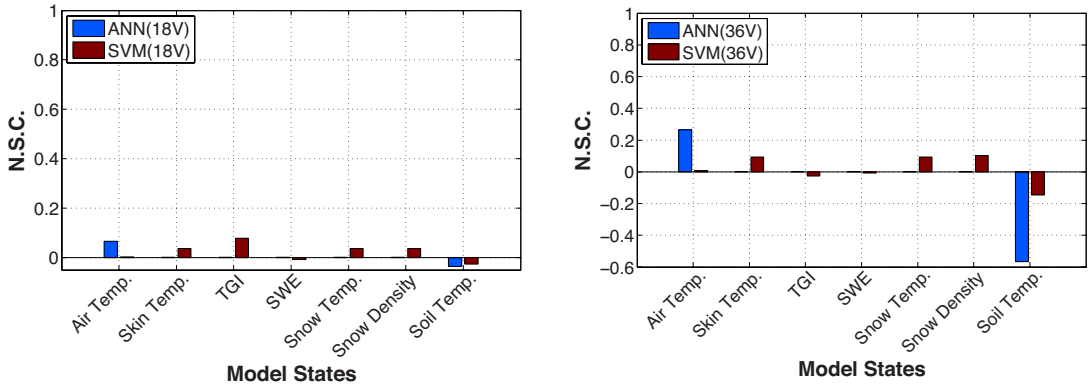

Figure 5.3-2 NSCs of seven model states for the location with low forest cover and low SWE in the NA domain on 11 Jan 2004 between ANN- and SVM-based vertically polarized Tb estimations at both 18 GHz and 36 GHz.

5.3.2. NSCs in the regions with low forest cover and high SWE

The representative location (latitude 54.6459° and longitude -61.7747°) of "High SWE" and "Low Veg" class is in the middle of Newfoundland and Labrador, Canada **(**see **Figure 5.3-3)**. The forest cover percentage within that region is 6.02% and the SWE value on 11 Jan 2004 is 0.14m, and therefore there is a moderate amount of snow on that day. Since this area is covered by relatively little vegetation, both the

ANN and SVM have the same performance in terms of the sensitivity of skin temperature and top layer snow temperature.

The scenario with low forest cover and high SWE possesses the highest NSC of SWE with the value of 0.3225 by comparing all the NSCs computations of both models. In such a case, forest effects are not significant because the emitted radiation from the underlying snowpack will not be strongly diminished by the forest cover. The SVM-based model captures the greatest amount of SWE information at 36 GHz among other model inputs related to the ANN-based model.

TGI also plays a role in the SVM-based Tb estimation model with the NSC value of 0.1342 for estimated Tb at 36 GHz. It is known that the snow temperature profile is not uniform due to heat flux exchanges between the snow, air, and underlying soil. The temperature of the snow surface responds to all types of weather conditions as well as daytime heating and nighttime cooling mechanism. Meanwhile, there is likely to be heat exchange in between the basal-layer snow and top-layer soil. In such case, the temperature gradient on the surface might be greater than that in the deeper layer. Hence, the sensitivity of TGI at 36 GHz is higher than that at 18 GHz for the SVM model.

Table 5.3-2 NSCs computations on 11 Jan 2004 for seven model states in an area with low forest cover and high SWE.

Model states	NSCs of single Tb frequency			
	ANN (18V)	ANN (36V)	SVM (18V)	SVM (36V)
Top layer snow density	0	0	-0.0272	0.119
SWE	0	0	0.0946	0.3225
Near-surface air temperature	0.0423	0.3035	-0.1347	0.0364
Near-surface soil temperature	0.4432	0.9189	-0.0385	0.0427
Skin temperature	0.0423	0.3035	0.1385	0.1542
Top layer snow temperature	0.0423	0.3035	0.1386	0.1542
TGI	0	0	0.0339	0.1342

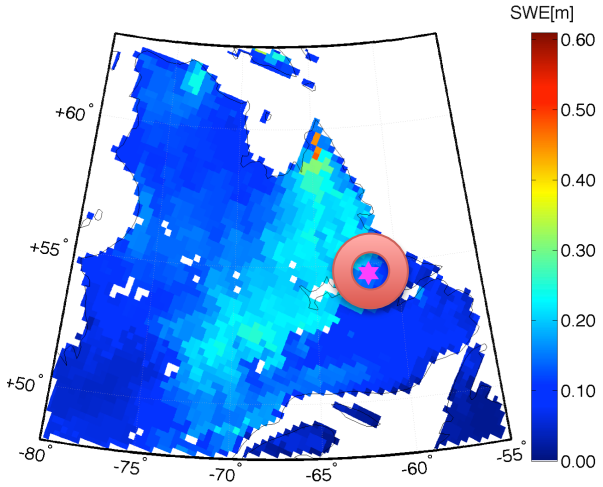

Figure 5.3-3 An example of a location (shown by the red circle) with low forest cover and high SWE value on the SWE map in the NA domain on 11 Jan 2004.

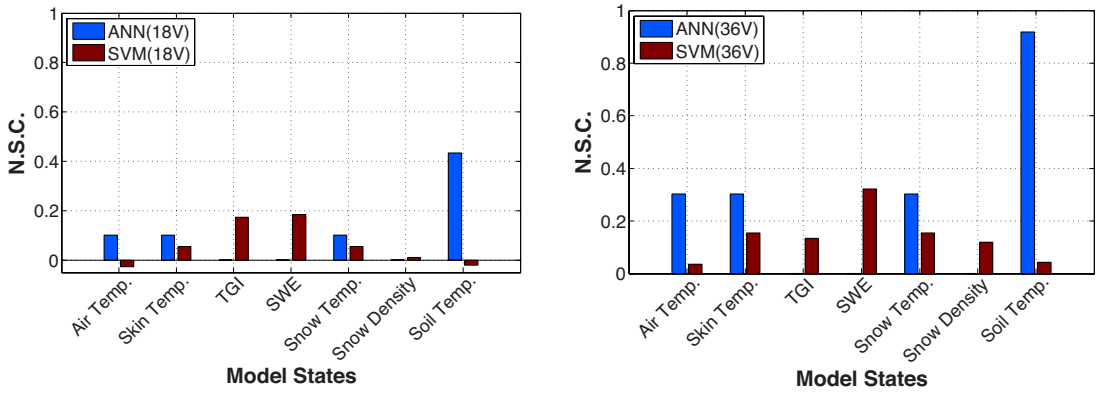

Figure 5.3-4 NSCs of seven model states for the specified location in the NA domain on 11 Jan 2004 between ANN- and SVM-based vertically polarized Tb estimations at both 18 GHz and 36 GHz.

5.3.3. NSCs in the regions with high forest cover and low SWE

This study location (latitude 60.7030° and longitude -113.3742°) of "Low SWE" and "High Veg" class is in the southeast part of Northwest Territories, Canada (see **Figure 5.3-5**). 88.02% of the area is covered with forest and with 0.03 m of SWE on 11 Jan 2004 and therefore there is relatively little snow on that day.

The ANN-based Tb predictions are still not sensitive to the snow-related states, except for the top layer snow temperature with the NSC value of 0.0899 for estimated Tb at 18 GHz. It is more likely that the accurate prediction of the ANN-based model does not depend on the model input of SWE. On the contrary, even during conditions with high forest cover and limited snow, the SVM-based model is still sensitive to all seven model states. Further, model states of SWE, skin temperature, and top layer snow temperature are the three most sensitive model inputs. It is encouraging to see that the SVM-based model is able to capture the variability of SWE in estimating Tb at both 18 GHz and 36 GHz, which suggests a larger sensitivity to SWE during the prediction of Tb by the SVM model.

Table 5.3-3 NSCs computation on 11 Jan 2004 for seven model states in an area with high forest cover and low SWE.

Model states	NSCs of single Tb frequency			
	ANN (18V)	ANN (36V)	SVM (18V)	SVM (36V)
Top layer snow density	0	0	-0.0061	-0.0272
SWE	0	0	0.1003	0.0946
Near-surface air temperature	0.0899	0.0423	0.0334	-0.1347
Near-surface soil temperature	0.272	0.4432	0.0593	-0.0385
Skin temperature	0.0899	0.0423	0.1004	0.1385
Top layer snow temperature	0.0899	0.0423	0.1005	0.1386
TGI	0	0	0.0134	0.0339

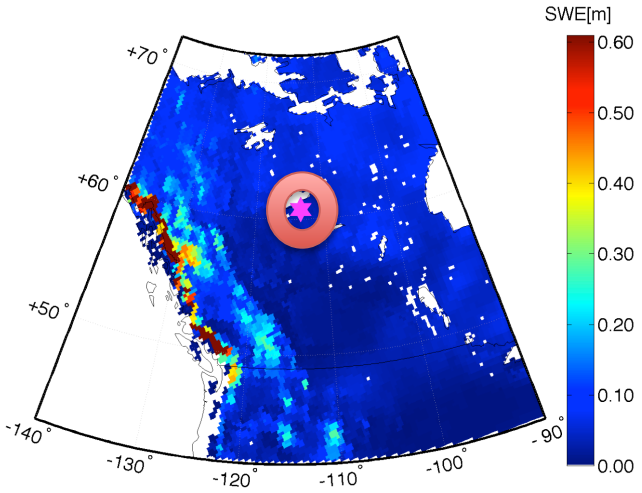

Figure 5.3-5 An example of a location (shown by the red circle) with high forest cover and low SWE value on the SWE map in the NA domain on 11 Jan 2004.

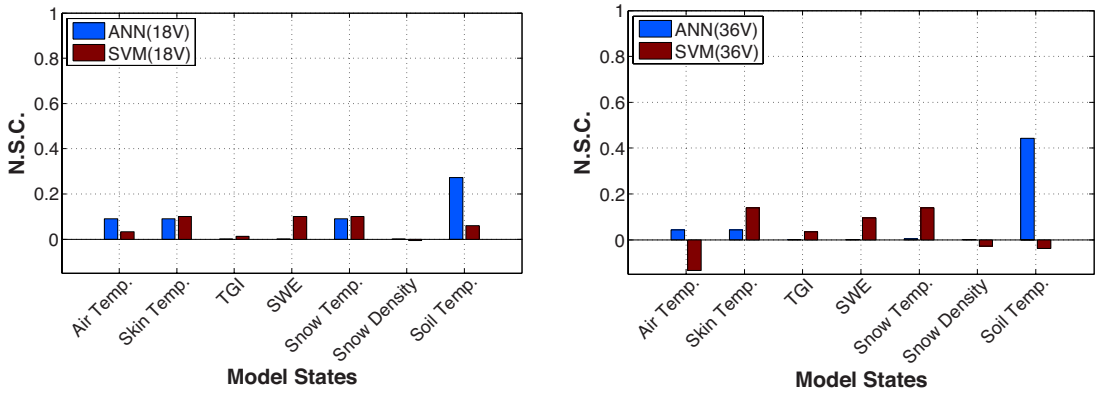

Figure 5.3-6 NSCs of seven model states for the specified location in the NA domain on 11 Jan 2004 between ANN- and SVM-based vertically polarized Tb estimations at both 18 GHz and 36 GHz.

5.3.4. NSCs in the regions with high forest cover and high SWE

The representative location (latitude 52.7082° and longitude –75.0232°) of "High SWE" and "High Veg" class is in the middle of Quebec, Canada (see **Figure 5.3-7**). The area is covered with 56.92% of forests with 0.13 m of SWE on 11 Jan 2004. Similar to other scenarios from **Table 5.3-1** to **Table 5.3-4**, the ANN-based model is most sensitive to the model input change in the soil temperature and has no response with respect to the relative change in SWE. It is also worth noting that the ANN-based NSC of SWE is also highly dependent on the location since only one **(Table 5.3-3)** out of nine selected regions in Chapter 5 contains SWE information that can partially influence Tb estimation.

Unlike the ANN-based model, the SVM-based model is sensitive to all seven model states in the area with high SWE coupled with high forest cover. SWE is still the most important model parameter in the model with the NSC value of 0.1553 for estimated Tb at 36 GHz, which will provide future study with more opportunity of

exploring the possibility of enhancing SWE estimation in the densely forested regions via Tb assimilation.

Table 5.3-4 NSCs computations on 11 Jan 2004 for seven model states in an area with high forest cover and high SWE.

Model states	NSCs single Tb frequency			
	ANN (18V)	ANN (36V)	SVM (18V)	SVM (36V)
Top layer snow density	0	0	-0.012	-0.0136
SWE	0	0	0.0939	0.1543
Near-surface air temperature	0	0	-0.012	-0.0136
Near-surface soil temperature	-0.093	-0.1864	-0.1053	-0.0609
Skin temperature	0	0	-0.0354	-0.0494
Top layer snow temperature	0	0	-0.0354	-0.0494
TGI	0	0	0.0014	0.0566

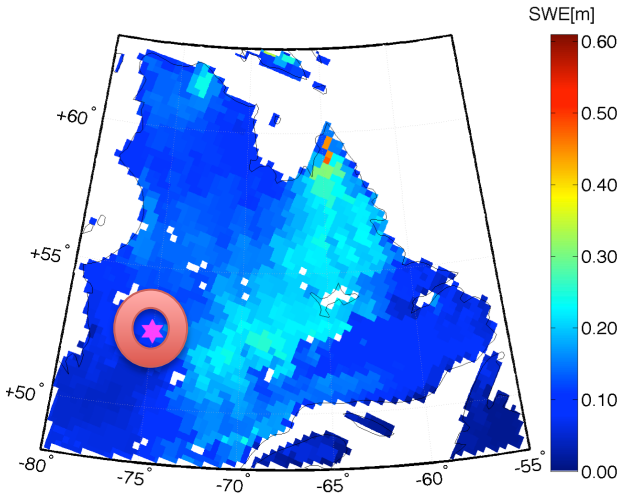

Figure 5.3-7 An example of a location (shown by the red circle) with high forest cover and high SWE value on the SWE map in the NA domain on 11 Jan 2004.

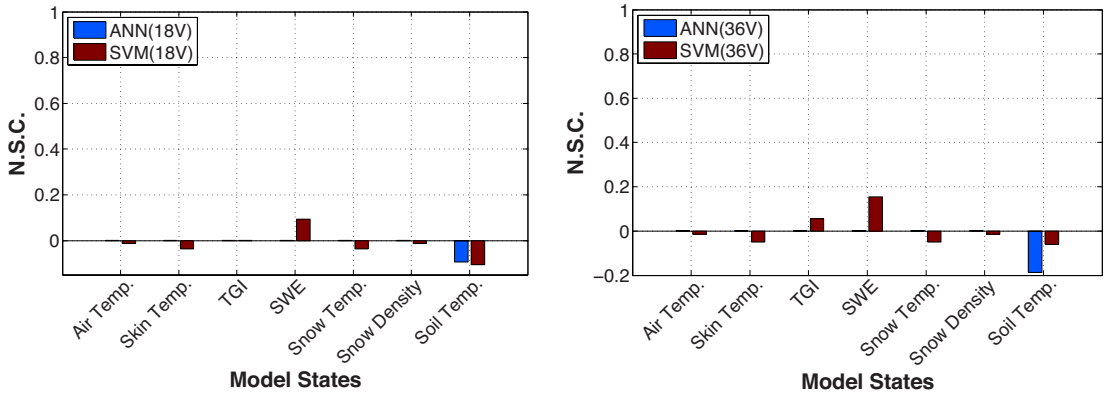

Figure 5.3-8 NSCs of seven model states for the specified location in the NA domain on 11 Jan 2004 between ANN- and SVM-based vertically polarized Tb estimations at both 18 GHz and 36 GHz.

5.4. TEMPORAL VARIABILITY OF NSCS OF SVM-BASED MODEL

In order to better compare the model behavior, the NSC analysis of the SVM-based model in the following section selects the same location and the same model states as stated in Section 5.2 for the ANN-based time series investigation. The temporal variability of the NSC is investigated under the snow accumulation phase and the ablation phase, respectively.

5.4.1. Snow accumulation phase

During the snow accumulation phase, compared with the ANN-based model, the SVM-based Tb estimation is more sensitive to the change in SWE, as the spatially-variable sensitivity analysis results suggested in Section 5.3. When the daily SWE values change abruptly (indicated by the slope of the green line in **Figure 5.4-1)**, which may result from a snowstorm that occurred on that day, the NSC for the SVM model has a strong response with respect to the daily-change in SWE. However, when there is no change in the SWE for a period of time, such as the time period from

06 Feb 2004 to 16 Feb 2004, both ANN- and SVM-based Tb estimations remain unchanged. This agrees well with the snow retrieval algorithm (Equation 2.3-4) derived by Chang et al. [1996]. If there is no change in the measured spectral difference (e.g., Tb at 37 GHz and Tb at 19 GHz), the SWE value is not expected to change. Therefore, the SVM-based model seems to be more reasonable with a more solid physical foundation.

In addition, Tb estimations from both models are highly sensitive to the near-surface soil temperature during the accumulation phase. This is because all points within the soil layer emit thermal radiation, and in the microwave region the intensity of the radiation is proportional to the thermal dynamic temperature [Choudhury et al. 1982] based on the Rayleigh-Jeans approximation (see Chapter 1). The equivalent temperature of the soil mainly depends on the soil moisture conditions and the inner soil temperature profile. Hence, it is reasonable that soil temperature is another important model parameter in the Tb prediction via machine learning.

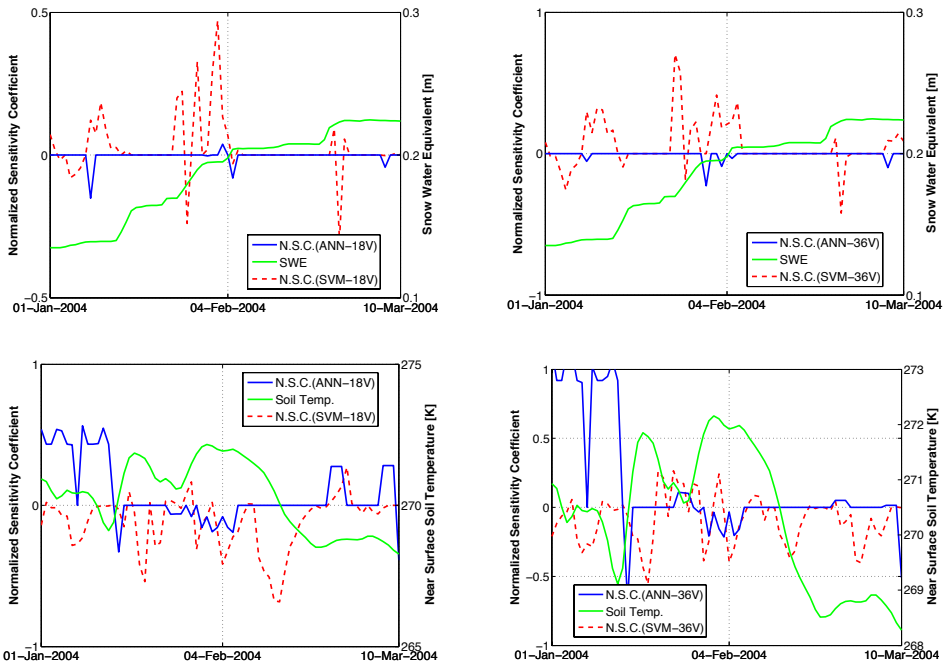

Figure 5.4-1 Time series investigation of NSCs at both 18 GHz and 36 GHz Tb predictions of SWE and near-surface soil temperature from 01 Jan 2004 to 10 Mar 2004.

5.4.2. Snow ablation phase

During the snow ablation phase, when the amount of snow drops dramatically from 01 May 2004 to the end of the May in 2004, the NSC of SWE for both models is zero. It may suggest that both machine-learning techniques can only be used during the onset of snow melting period in extracting SWE based on measured Tb.

The ANN-based model is highly sensitive to the soil temperature state. One of the hypotheses is that during the snow ablation season, melting snow will penetrate into the soil. Hence, the presence of more soil moisture will take the dominant role in significantly increasing the radiation emission ability of the soil, which will result in a higher estimation of Tb. The other preliminary conclusion is SWE is not a sensitive model parameter in the ANN-based Tb prediction. In other words, the good performance of the ANN model operator in "learning" measured Tb across the NA

domain does not have direct linkage with SWE information. In addition, Forman and Reichle [2014] pointed out that the ANN is less capable of capturing much of the temporal variability found in the original AMSR-E Tb measurements. Compared with soil temperature, snow states (e.g., SWE, snow grain size and snow temperature) are more variable due to more interactions with the overlying air and canopy cover. The relatively high sensitivity of SWE in the SVM-based model possibly depends on its capability of capturing more of interannual variability of the Tb estimates across the entire NA domain. Therefore, the attempt to improve Tb prediction within a DA framework will not necessarily improve SWE estimation since there might not exist a large error covariance between these two variables.

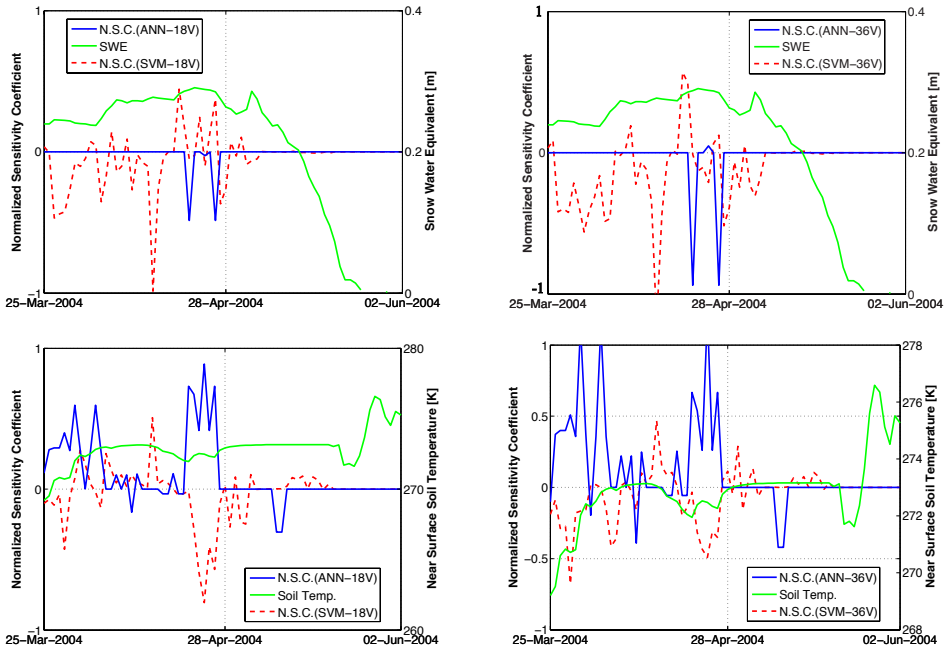

Figure 5.4-2 Time series investigation of NSCs at both 18 GHz and 36 GHz Tb predictions of SWE and near-surface soil temperature from 25 Mar 2004 to 02 Jun 2004.

5.5. SENSITIVITY ANALYSIS OF ANN- AND SVM-BASED SPECTRAL DIFFERENCE

All of the discussions above regard the relative change in the estimation of a single vertically polarized Tb frequency, either at 18.7 GHz or at 36.5 GHz, with respect to the relative change in SWE (or other model states). It can be concluded that shorter wavelengths (i.e., 36 GHz) do not have the capacity to penetrate as deeply into the snowpack, hence, some of the snow-related information or signal may be lost. However, less radiation is scattered at lower frequencies, which has the potential to provide more information about snow conditions, such as SWE. There is a trade-off between these two Tb frequencies in SWE estimation.

Based on the snow retrieval algorithm derived by Chang et al. [1996], SWE is proportional to the vertical spectral difference between 18.7 GHz and 36 GHz. Hence, the NSC of SWE to vertically-polarized spectral difference will be investigated in this section, and can be expressed as:

$$NSC_{(SWE,\ \Delta Tb)} = \left\{\frac{\Delta(Tb_{18V}\text{-}Tb_{36V})}{\Delta SWE}\right\} \cdot \left\{\frac{SWE^{0}}{\Delta\left(Tb^{0}_{18V}\text{-}Tb^{0}_{36V}\right)}\right\} \qquad (5.5\text{-}1)$$

where $NSC_{(SWE,\ \Delta Tb)}$ [dimensionless] is the rate of change in vertical spectral difference (ΔTb) with respect to changes in SWE; $\Delta\left(Tb_{18V}\text{-}Tb_{36V}\right)$ [K] is the difference between Tb estimation at 18.7 GHz and 36.5 GHz; ΔSWE [m] is the change in SWE magnitude; SWE^{0} [m] is the nominal value of SWE before exerting any perturbations; and $\Delta\left(Tb^{0}_{18V}\text{-}Tb^{0}_{36V}\right)$ [K] is the difference between the nominal value (before perturbation) of Tb estimations at 18.7 GHz and 36.5 GHz.

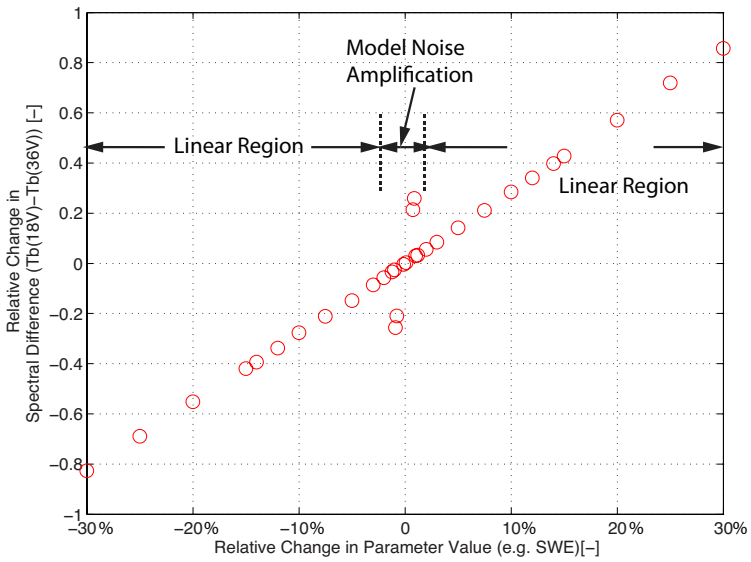

Figure 5.5-1 Perturbations effects in the sensitivity analysis of the SVM-based Tb predictions at the spectral difference between 18.7 GHz and 36.5 GHz with respect to SWE.

Figure 5.5-1 is an example of the NSC of SWE for a given location (latitude 54.8172° and longitude –66.6055°) at the Tb spectral difference between 18.7 GHz and 36.5 GHz based on different perturbation sizes. Only the SVM model is presented here since the NSC for spectral difference is zero for the ANN-based model for this given location. When the perturbation size of SWE varies from -2% to +2%, the model response falls into the model noise amplification region. During the linear region, the relative change in the spectral difference is proportional to the SWE magnitude variation with a high correlation coefficient, which can be indicated by the positive slope of the line.

The comparison results between the NSC of SWE for individual frequencies as well as the spectral difference between two frequencies on 11 Jan 2004 over the entire NA domain is shown in the **Figure 5.5-2.**

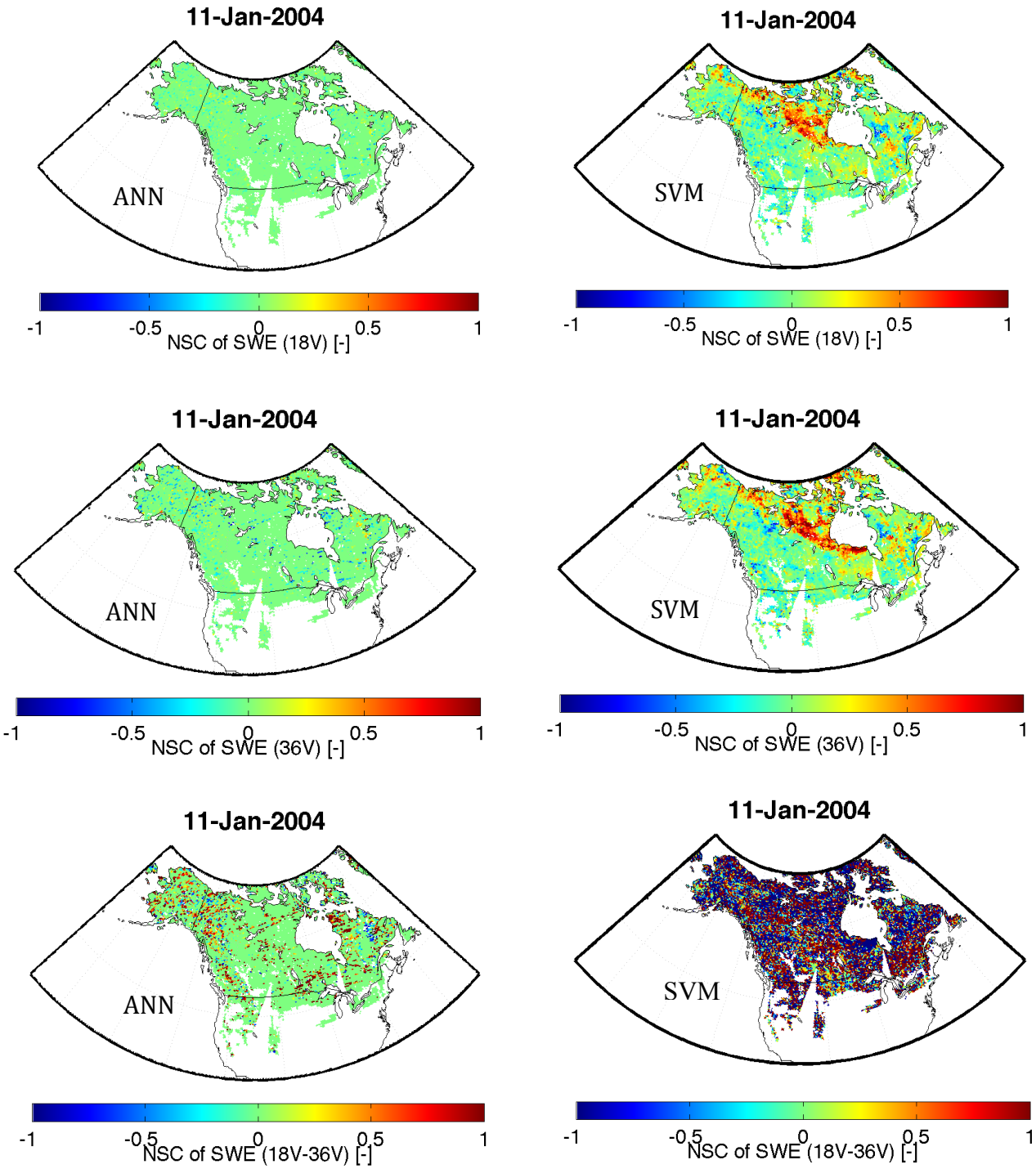

Figure 5.5-2 The NSC of SWE at single spectral frequency and the NSC of SWE at vertical spectral difference on 11 Jan 2004 in the NA domain.

*: NSC of SWE (18V): rate of change in vertically polarized Tb at 18 GHz with respect to SWE
*: NSC of SWE (36V): rate of change in vertically polarized Tb at 36 GHz with respect to SWE
*: NSC of SWE (18V-36V): rate of change in the difference of the vertically polarized Tb at 18 GHz and 36 GHz with respect to SWE

From **Figure 5.5-2**, it can be seen that both ANN- and SVM-based models are sensitive to the SWE state to some extent at some locations on 11 Jan 2004 during the snow accumulation phase. However, as discussed in Section 5.3, SWE plays a more dominant role in most of the regions in the NA domain for the SVM-based estimation of Tb compared to that of the ANN-based model. The NSC of SWE for the ANN-based Tb estimation is more dependent on a specific location. For instance, based on the NSCs computation of SWE at single Tb frequency for both 18 GHz and 36 GHz across the NA domain, only 11.48% of the snow-covered regions have non-zero values of NSC of SWE on 11 Jan 2004.

In terms of the SVM model, the NSC at a single Tb frequency suggests that regions such as the middle of the Canadian Shield (Laurentian Plateau) possess the largest sensitivity to SWE. The reason for the strong sensitivity is still unknown since it is affected by forest cover, snow formation, climate conditions, and topography. However, the NSC map of SWE at single Tb frequency provides the future study with great opportunity in terms of Tb estimation applied in forested areas, which to some extent solve the restriction posed by the traditional radiative transfer model as a model operator to invert Tb into model state variables.

In addition, as indicated by the NSC map of SWE at the vertical spectral difference between 18 GHz and 36 GHz in the NA domain, the relative change in SWE plays a more dominant role in determining the spectral difference for both models instead of a single Tb frequency. This phenomenon agrees well with the empirical snow retrieval equation [Change et al. 1996]. The original equation can be written as:

$$SWE = a\frac{(T_{18,V}-T_{37,V})}{1-ff} \quad (5.5\text{-}2)$$

where a ($a > 0$) is a constant determined by regression analysis, and ff ($0 \leq ff \leq 1$) is the forest cover percentage. Further, a can also be written as:

$$a = \frac{\Delta SWE(1-ff)}{\Delta(T_{18,V}-T_{37,V})} \quad (5.5\text{-}3)$$

After replacing the right hand side of the equation with the NSC, it can now be expressed as:

$$a = \frac{SWE_0(1-ff)}{NSC \times Tb_0} \quad (5.5\text{-}4)$$

Hence,

$$NSC = \frac{SWE_0(1-ff)}{a \times Tb_0} > 0 \quad (5.5\text{-}5)$$

where the NSC is the relative change in Tb with respect to small perturbations in SWE; SWE_0[m] is the nominal SWE value before perturbation; ff [%] is the forest cover percentage; Tb_0 [K] is the nominal Tb prediction; and a is a constant that empirically should be greater than zero. Hence, the NSC of SWE should be greater than zero based on the theory.

As the NSC map of SWE at vertical spectral difference has demonstrated, most of the NSCs are indeed positive when using the SVM-based model. In such a case, the SVM may be a superior model measurement operator to the ANN in terms of enhancing SWE estimation at regional- or continental-scale.

CHAPTER 6: COCLUSIONS AND RECOMMENDATIONS

Based on the previous sensitivity analysis (NSCs computations), some key findings are concluded in this chapter. Additionally, possible explanations of the insensitivity of the ANN-based model to the snow-related states will be briefly discussed. In order to further verify the applicability of the SVM-based model, this chapter also briefly describes several research objectives that need to be addressed in the future.

6.1. SUMMARY AND CONCLUSION

The sensitivity analysis of Tb estimations for both ANN and SVM models are performed with respect to different models states. Based on the NSCs computation in Chapter 5, the key findings are summarized as follows:

- Compared to the vertically polarized Tb at 18 GHz for both ANN- and SVM-based estimations, the Tb at 36 GHz tends to have a higher sensitivity with respect to small perturbations in the model inputs. This is partially explained by the fact that higher PMW frequencies possess a smaller emission depth. Hence, the 36 GHz channel can capture more variability on the surface of the snowpack, which has more interactions with the atmosphere, and overlying vegetation.
- Sensitivities are greatest for non-forested or sparsely-forested regions with relatively high amounts of snow for both of the machine learning techniques during the snow ablation phase, where the NSC of SWE for both models for Tb at

36 GHz can be closer to -1 (see **Figure 5.4-2**). Melting snow introduces the presence of liquid water into the snowpack, which behaves like a blackbody at the physical temperature of the snow layer. It significantly increases the emission of microwave energy. Additionally, in the absence of forest cover, Tb measurements are more directly related to PMW emission from the snowpack.

- The SVM-based model is more sensitive to snow-related variables, for example, SWE, TGI, and upper-layer snow temperature. However, in the ANN-based model, Tb predictions are relatively insensitive to TGI and snow density, whose NSCs are often zeros. Further, the ANN's sensitivity to SWE is more dependent on a specific location or a specific period of time. Alternatively, the ANN is more sensitive to the near-surface soil temperature across a range of locations and time periods, and sometimes the magnitude of NSCs can be greater than one. Hence, the SWE information cannot always be leveraged during the ANN-based Tb estimation.
- In highly-vegetated areas, the sensitivity of the system is more dominated by vegetative canopy and surface temperature and less so with snow-related variables. Forest cover can attenuate the emission of radiation from the snowpack prior to reaching the PMW sensor while simultaneously adding its own contribution to the measured radiation.
- Even in areas of dense vegetation and relatively low SWE, the SVM-based model still shows the highest sensitivity to snow-related model states (e.g., NSC of SWE is ~0.1 for both 18 GHz and 36 GHz vertically-polarized Tb), compared with those of the ANN-based model.

- The output of the model, either SVM- or ANN-based, of the spectral difference in Tb ($Tb_{,18V}$-$Tb_{,36V}$), is more sensitive to small perturbations in the model inputs, which agrees well with the empirical relationship established by previous studies [Chang et al. 1996]. However, the SVM-based model possesses a more significant performance in predicting the spectral difference in Tb than the ANN across the NA domain as indicated by the areal distribution of positive values of NSCs in **Figure 5.3-2**. This may suggest that relative to the ANN, the SVM can better retrieve SWE from Tb measurements.

Previous studies conducted by Forman and Reichle [2014] has demonstrated the inability of the ANN-based model to capture the inter-annual variability of the measured Tb across a time period in the NA domain. Some weird step functions are found in the ANN-based Tb predictions for both 18 GHz and 36 GHz. However, the snow-related properties (e.g., SWE and snow temperature) will fluctuate more vibrantly than other soil-related properties, since the overlying snow has more opportunities to interact with air, vegetation and ground/soil. Hence, it is postulated that the ANN-based model may have difficulty in capturing the fluctuations of highly variable model inputs, such as SWE. One of the possible explanations for the insensitivity of the ANN-based model with respect to snow-related states may be resulting from its learning algorithm. The ultimate goal of the ANN is to minimize the objective function of the mean squared error (Chapter 2), however, sometimes it may converge to a local minimum point instead of a global minimum. On the contrary, formulations of SVM-based models are convex optimization problems and

thereby unique global optima will be found and the algorithm will not be affected by the local minima problem.

In summary, compared with the ANN, the SVM could potentially serve as a more efficient measurement model operator at the regional- and continental-scale for forested and non-forested regions as part of a data assimilation framework in enhancing SWE estimation.

6.2. RECOMMENDATIONS FOR FUTURE RESEARCH

6.2.1. Physical interpretations of NSCs

As indicated by the computed NSCs in Chapter 5, there are often different signs associated with the NSC of various model states for both ANN- and SVM-based models. This behavior may be explained by the deficiency of both models in "learning" regularities or on the dependency of the specificity of the location of interest. It may also result from sub-grid scale lakes and depth hoar effects. Hence, in order to further validate the SVM-based model, the physical interpretations of NSCs, especially for those changing in signs, need to be better understood.

6.2.2. NSCs of SWE in forested regions

Even if the SVM-based model shows great potential in being implemented in the densely-vegetated areas (**Table 5.3-3**), it is still difficult to draw a sound conclusion that the SVM can be successfully applied anywhere with dense vegetation. One possible solution, such as introducing the Normalized Difference Vegetation Index

(NDVI) into the model inputs, could be an effective method to better illustrate the role of vegetation in Tb predictions. A sensitivity analysis is still needed to examine the Tb response with respect to small perturbations in the vegetation index state.

6.2.3. Investigation of polarization ratio

The polarization ratio, P_r, can serve as an indicator for the presence of ice layers or ice crusts across the study domain, which can be defined as [Cavalieri et al. 1984]:

$$P_r(f^*) = \frac{Tb_{f^*V} - Tb_{f^*H}}{Tb_{f^*V} + Tb_{f^*H}} \quad (6.2\text{-}1)$$

where $P_r(f^*)$ is the polarization ratio at frequency f^*; Tb_{f^*V} [K] represents the vertically polarized Tb at frequency f^* (e.g., 10.65 GHz, 18.7 GHz, and 36.5 GHz); Tb_{f^*H} [K] is the vertically polarized Tb at frequency f^*. The AMSR-E sensor has twelve passive sensors consisting of six dual-polarized frequency channels, which provides ample opportunity to investigate the existence of ice layers or ice crusts that can dramatically reduce the measured Tb via increased scattering effects.

6.2.4. Machine learning with other passive microwave products

The hypothesis has been proposed that machine learning can be applied to other remote sensing products measuring Tb such as SSM/I and SSMR. Therefore, after verifying the rationality of the SVM-based model with AMSR-E observations, it would be worthwhile to investigate the robustness of the SVM-based predictive capability on other sources of Tb observations. A sensitivity analysis would also be

required to investigate the model sensitivity to SWE and other snow-related input states.

6.2.5. SWE estimation within data assimilation framework

Enhancing SWE estimation at regional and continental scales is the eventual goal for this study. SWE can be determined by using a DA framework (**Figure 6.2-1**) in order to yield a merged estimate of SWE that is superior to either the measurement or the model estimation from Catchment alone. Unlike previous trials of assimilating SWE estimates directly (see Chapter 1), the study proposed here will assimilate Tb by combining space-borne measurements with SVM-based Tb predictions. By utilizing a DA technique, SWE estimation may be improved based on the physical connections between SWE and Tb estimation as suggested by the sensitivity analysis results based on SVM in the Chapter 5.

Several DA techniques are available nowadays in many fields of geosciences, among which ensemble-based variants and variants of the Kalman filter (KF) are the most promising tools in hydrologic studies [Reichle et al. 2002; Andreadis and Lettenmaier, 2006]. The traditional KF is not suitable for solving such a complex, highly nonlinear Tb assimilation problem. Hence, the Ensemble Kalman Filter (EnKF) and the Ensemble Kalman Smoother (EnKS) are the two main techniques recommended for future study.

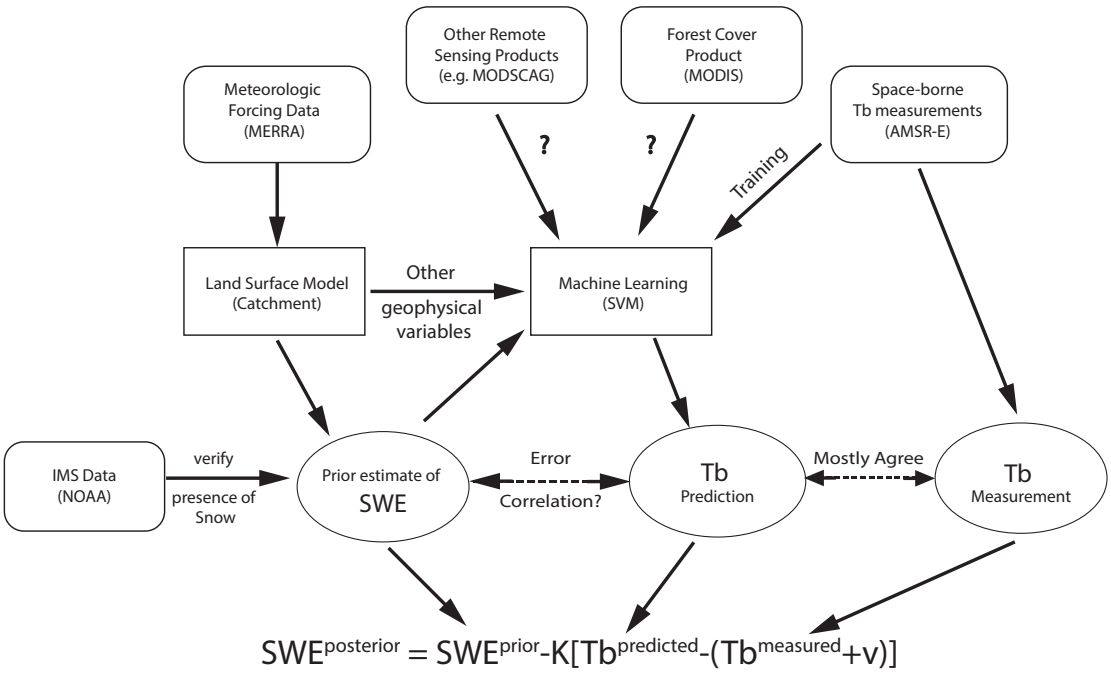

Figure 6.2-1 Expected SWE estimation within a DA framework. $SWE^{posterior}$ is the posterior estimate of SWE after implementing DA (i.e., filtering or smoothing); SWE^{prior} is the prior estimate of SWE prior to performing measurement assimilation; K is the Kalman gain used to weigh the difference sources of uncertainty; $Tb^{predicted}$ is the SVM-based Tb estimation; $Tb^{measured}$ is the measured Tb from AMSR-E; v is the AMSR-E measurement error matrix; and MODSCAG is short for MODIS Snow Covered Area and Grain Size.

REFERENCES

Aaron S. Weed, Matthew P. Ayres, and Jeffrey A. Hicke. (2013). Consequences of climate change for biotic disturbances in North American forests. *Ecological Monographs,* 83:441–470.

Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, *29*(6), 872-886.

Andreadis, K. M., Liang, D., Tsang, L., Lettenmaier, D. P., & Josberger, E. G. (2008). Characterization of Errors in a Coupled Snow Hydrology–Microwave Emission Model. *Journal of Hydrometeorology*, *9*(1).

Armstrong RL. (1985). Metamorphism in a subfreezing, seasonal snow cover: the role of thermal and vapor pressure conditions. Ph.D. dissertation, University of Colorado; 175 pp.

Armstrong, R. L., & Brun, E. (Eds.). (2008). Snow and climate: physical processes, surface energy exchange and modeling. Cambridge University Press.

Armstrong, R. L., Chang, A., Rango, A., & Josberger, E. (1993). Snow depths and grain-size relationships with relevance for passive microwave studies. *Annals of Glaciology*, *17*, 171-176.

Atkinson, P. M., & Tatnall, A. R. L. (1997). Introduction neural networks in remote sensing. *International Journal of remote sensing*, *18*(4), 699-709.

Baboo, S. S., & Shereef, I. K. (2010). An efficient weather forecasting system using artificial neural network. *International journal of environmental science and development*, *1*(4), 2010-0264.

Bader, H. (1970). The hyperbolic distribution of particle sizes. *Journal of Geophysical Research*, 75(15), 2822-2830.

Baughman, D. R., & Liu, Y. A. (1995). Neural networks in bioprocessing and chemical engineering.

Bechle, M. J., Millet, D. B., & Marshall, J. D. (2013). Remote sensing of exposure to NO2: Satellite versus ground-based measurement in a large urban area. *Atmospheric Environment*, 69, 345-353.

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. In *Data mining techniques for the life sciences* (pp. 223-239). Humana Press.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: a review. *Hydrological processes*, *9*(3-4), 251-290.

Bogataj, L. (2007). How will the Alps Respond to Climate Change. *Alpine space–man & environment,* 3, 43-51.

Bohr, G. S., and E. Aguado (2001). Use of April 1 SWE measurements as estimates of peak seasonal snowpack and total cold-season precipitation, *Water Resour. Res*., 37(1), 51–60, doi:10.1029/2000WR900256.

Brown, R. D., & Robinson, D. A. (2011). Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty. *The Cryosphere*, 5(1), 219-229.

Brucker, L., Royer, A., Picard, G., Langlois, A., & Fily, M. (2011). Hourly simulations of the microwave brightness temperature of seasonal snow in

Quebec, Canada, using a coupled snow evolution–emission model. *Remote Sensing of Environment*, *115*(8), 1966-1977.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, *2*(2), 121-167.

Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1882-1889.

C. Ma ̈tzler (1994). Passive microwave signatures of landscapes in winter. Meteorol. Atmos. Phys., 54:241–260.

Campbell, J. B. (2002). Introduction to remote sensing. CRC Press.

Cavalieri, D. J., Gloersen, P., & Campbell, W. J. (1984). Determination of sea ice parameters with the Nimbus 7 SMMR. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *89*(D4), 5355-5369.

Cawley, G. C., & Talbot, N. L. (2007). Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters. *The Journal of Machine Learning Research*, *8*, 841-861.

Cayan, D. R. (1996). Interannual climate variability and snowpack in the western United States. *Journal of Climate,* 9(5), 928-948.

Chan, S., Treleaven, P., & Capra, L. (2013, October). Continuous hyperparameter optimization for large-scale recommender systems. In *Big Data, 2013 IEEE International Conference on* (pp. 350-358). IEEE.

Chang, A. T. C. (1986). Nimbus-7 SMMR snow cover data. Available from the National Technical Information Service, Springfield VA 22161, as DE 86-011983. Price codes: A 12 in paper copy.

Chang, A. T. C., & Tsang, L. (1992). A neural network approach to inversion of snow water equivalent from passive microwave measurements. *Nordic Hydrology*, *23*(3), 173-182.

Chang, A. T. C., Foster, J. L., & Hall, D. K. (1987). Nimbus-7 SMMR derived global snow cover parameters. *Ann. Glaciol*, 9(9), 39-44.

Chang, A. T. C., Foster, J. L., & Hall, D. K. (1996). Effects of forest on the snow parameters derived from microwave measurements during the BOREAS winter field campaign. *Hydrological Processes*, *10*(12), 1565-1574.

Chang, A. T. C., Foster, J. L., Hall, D. K., Rango, A., & Hartline, B. K. (1982). Snow water equivalent estimation by microwave radiometry. *Cold Regions Science and Technology*, 5(3), 259-267.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

Chang, T. C., Gloersen, P., Schmugge, T., Wilheit, T. T., & Zwally, H. J. (1975). Microwave emission from snow and glacier ice. Goddard Space Flight Center.

Choudhury, B. J., Schmugge, T. J., & Mo, T. (1982). A parameterization of effective soil temperature for microwave emission. *Journal of Geophysical Research: Oceans (1978–2012)*, *87*(C2), 1301-1304.

Clifford, D. (2010). Global estimates of snow water equivalent from passive microwave instruments: history, challenges and future developments. *International Journal of Remote Sensing*, *31*(14), 3707-3726.

Colbeck, S. C. (1987). A review of the metamorphism and classification of seasonal snow cover crystals. *IAHS Publication*, *162*, 3-24.

Davis, D. T., Chen, Z., Tsang, L., Hwang, J. N., & Chang, A. T. (1993). Retrieval of snow parameters by iterative inversion of a neural network. *Geoscience and Remote Sensing, IEEE Transactions on*, *31*(4), 842-852.

Derksen, C., E. LeDrew, and B. Goodison (2000), Temporal and spatial variability of North American prairie snow cover (1988–1995) inferred from passive microwave- derived snow water equivalent imagery, *Water Resour. Res.,* 36(1), 255–266, doi:10.1029/1999WR900208.

Derksen, C., Toose, P., Rees, A., Wang, L., English, M., Walker, A., & Sturm, M. (2010). Development of a tundra-specific snow water equivalent retrieval algorithm for satellite passive microwave data. *Remote Sensing of Environment*, *114*(8), 1699-1709.

Derksen, C., Walker, A., & Goodison, B. (2005). Evaluation of passive microwave snow water equivalent retrievals across the boreal forest/tundra transition of western Canada. *Remote Sensing of Environment,* 96(3), 315-327.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78-87.

Dong, J., Walker, J. P., Houser, P. R., & Sun, C. (2007). Scanning multichannel microwave radiometer snow water equivalent assimilation. *Journal of Geophysical Research: Atmospheres* (1984–2012), 112(D7).

Duguay, R., & Pietroniro, A. (2005). Remote sensing in northern hydrology: Measuring environmental change (Vol. 163, pp. 1-160). American Geophysical Union.

Durand, M., & Margulis, S. A. (2006). Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent. *Journal of Hydrometeorology*, *7*(3).

Durand, M., & Margulis, S. A. (2007). Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *112*(D13).

Durand, M., Kim, E. J., & Margulis, S. A. (2008). Quantifying uncertainty in modeling snow microwave radiance for a mountain snowpack at the point-scale, including stratigraphic effects. *Geoscience and Remote Sensing, IEEE Transactions on*, *46*(6), 1753-1767.

Durand, M., Kim, E. J., Margulis, S. A., & Molotch, N. P. (2011). A first-order characterization of errors from neglecting stratigraphy in forward and inverse passive microwave modeling of snow. *Geoscience and Remote Sensing Letters, IEEE*, *8*(4), 730-734.

Dyer, J. L., & Mote, T. L. (2006). Spatial variability and trends in observed snow depth over North America. *Geophysical Research Letters,* 33(16).

Fletcher, L., Katkovnik, V., Steffens, F. E., & Engelbrecht, A. P. (1998, May). Optimizing the number of hidden nodes of a feedforward artificial neural network. In *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on* (Vol. 2, pp. 1608-1612). IEEE.

Foppa, N., Stoffel, A., & Meister, R. (2007). Synergy of in situ and space borne observation for snow depth mapping in the Swiss Alps. *International journal of applied earth observation and geoinformation,* 9(3), 294-310.

Forman, B. A. and R. H. Reichle (In review, 2014). Using a support vector machine and a land surface model to estimate large-scale passive microwave temperatures over snow-covered land in North America, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Forman, B. A., R. H. Reichle, and C. Derksen (2013). Estimating passive microwave brightness temperature over snow-covered land in North America using a land surface model and an artificial neural network, *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS. 2013. 2237913.

Foster, J. L., Chang, A. T. C., & Hall, D. K. (1997). Comparison of snow mass estimates from a prototype passive microwave snow algorithm, a revised algorithm and a snow depth climatology. *Remote sensing of environment*, 62(2), 132-142.

Foster, J. L., Hall, D. K., & Chang, A. T. C. (1987). Remote sensing of snow. *Eos, Transactions American Geophysical Union*, 68(32), 682-684.

Foster, J. L., Hall, D. K., Eylander, J. B., Riggs, G. A., Nghiem, S. V., Tedesco, M., ... & Choudhury, B. (2011). A blended global snow product using visible, passive microwave and scatterometer satellite data. *International journal of remote sensing,* 32(5), 1371-1395.

Foster, J. L., Sun, C., Walker, J. P., Kelly, R., Chang, A., Dong, J., & Powell, H. (2005). Quantifying the uncertainty in passive microwave snow water equivalent observations. *Remote Sensing of Environment*, *94*(2), 187-203.

Glickman, Todd S. (2000). "Glossary of Meteorology.".

Goodison, B. E., A. E. Walker (1994). Canadian development and use of snow cover information from passive microwave satellite data, Passive Microwave Remote Sensing of Land-Atmosphere InteractionsB. J. Choudhury, *et al.,* 245–262, *VSP,* Utrecht, The Netherlands.

Hall, D. (1987). Influence of depth hoar on microwave emission from snow in northern Alaska. Cold Regions Science and Technology, 13(3), 225–231.

Hall, D. K., & Martinec, J. (1985). Remote sensing of ice and snow.

Hall, D. K., Kelly, R. E., Foster, J. L., & Chang, A. T. (2005). Hydrological applications of remote sensing: surface states: snow. *Encyclopedia of Hydrological Sciences*, 2, 811-30.

Hall, D.K., A.T.C. Chang and J.L. Foster. (1986). Detection of the depth-hoar layer in the snow-pack of the Arctic Coastal Plain of Alaska, U.S.A., using satellite data, *Journal of Glaciology*, 32(110):87-94.

Hallikainen, M. T., Ulaby, F., & Abdelrazik, M. (1986). Dielectric properties of snow in the 3 to 37 GHz range. *Antennas and Propagation*, *IEEE Transactions* on, 34(11), 1329-1340.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., & Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, *7*(1).

Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, *27*(2), 83-85.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The elements of statistical learning* (Vol. 2, No. 1). New York: Springer.

He, Z., Wen, X., Liu, H., & Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, *509*, 379-386.

Helfrich, S. R., McNamara, D., Ramsay, B. H., Baldwin, T., & Kasheta, T. (2007). Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrological processes*, *21*(12), 1576-1586.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.

Isukapalli, S. S. (1999). *Uncertainty analysis of transport-transformation models* (Doctoral dissertation, Rutgers, The State University of New Jersey).

Josberger, E. G., & Mognard, N. M. (2002). A passive microwave snow depth algorithm with a proxy for snow metamorphism. *Hydrological Processes*, *16*(8), 1557-1568.

Kelly, R. (2009). The AMSR-E snow depth algorithm: Description and initial results. *Journal of the Remote Sensing Society of Japan*, *29*.

Kelly, R., Chang, A. T. C., Foster, J. L., & Tedesco, M. (2004). AMSR-E/Aqua daily L3 global snow water equivalent EASE-Grids V002. National Snow and Ice Data Center, Boulder, CO, digital media.[Available online at http://nsidc. org.].

Kelly, R.E.J., Chang, A.T.C, Tsang, L. and Foster, J.L. (2003). A prototype AMSR-E global snow area and snow depth algorithm, *IEEE Transactions Geoscience and Remote Sensing*, 4 1(2):230-242.

Knowles, K., M. Savoie, R. Armstrong, and M. Brodzik. (2006). *AMSR-E/Aqua Daily EASE-Grid Brightness Temperatures.* [indicate subset used]. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *105*(D20), 24809-24822.

Langlois, A., Royer, A., Dupont, F., Roy, A., Goïta, K., & Picard, G. (2011). Improved corrections of forest effects on passive microwave satellite remote

sensing of snow over boreal and subarctic regions. *Geoscience and Remote Sensing, IEEE Transactions on*, *49*(10), 3824-3837.

Lathi, B. P. (1990). Modern digital and analog communication systems. Oxford University Press, Inc.

Lea, J. and Lea, J. (1998). Snowpack Depth and Density Changes during Rain on Snow Events at Mt. Hood Oregon. *International Conference on Snow Hydrology: The Integration of Physical, Chemical and Biological Systems*, October 6-9, Brownsville, VT.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quarterly of applied mathematics*, *2*, 164-168.

Liang, D., Xu, X., Tsang, L., Andreadis, K. M., & Josberger, E. G. (2008). The effects of layers in dry snow on its passive microwave emissions using dense media radiative transfer theory based on the quasicrystalline approximation (QCA/DMRT). *Geoscience and Remote Sensing, IEEE Transactions on*, *46*(11), 3663-3671.

Libbrecht, K. G. (2005). The physics of snow crystals. *Reports on progress in physics,* 68(4), 855.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *ASSP Magazine, IEEE*, *4*(2), 4-22.

Lydolph, P. E. (1985). The climate of the earth. Government Institutes.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, *11*(2), 431-441.

Mathur, A., & Foody, G. M. (2004, September). Land Cover classification by support vector machine: Towards efficient training. In *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International* (Vol. 2, pp. 742-744). IEEE.

Mattera, D., & Haykin, S. (1999, February). Support vector machines for dynamic reconstruction of a chaotic system. In *Advances in kernel methods* (pp. 211-241). MIT Press.

McCuen, R. H. (2002). *Modeling hydrologic change: statistical methods*. CRC press.

McCuen, R. H. (2005). Accuracy assessment of peak discharge models. *Journal of Hydrologic Engineering*, *10*(1), 16-22.

McLaughlin, D. (2002). An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. *Advances in Water Resources*, *25*(8), 1275-1286.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning, neural and statistical classification.

Moradkhani, H. (2008). Hydrologic remote sensing and land surface data assimilation. *Sensors,* 8(5), 2986-3004.

Mott, H. (1986). Polarization in antennas and radar. New York, Wiley-Interscience, 1986. 312 p.1.

Mulders, M. A. (1987). Remote sensing in soil science (Vol. 15). Elsevier.

Pampaloni, P. (2000). Microwave radiometry and remote sensing of the earth's surface and atmosphere*: [a selection of refereed papers presented at the 6th Specialist Meeting on Microwave Radiometry and Remote Sensing of the*

Environment held in Florence, Italy on March 15 - 18, 1999]. Utrecht [u.a.]: VSP.

Petrenko, V. F., & Whitworth, R. W. (1999). *Physics of ice*. Oxford University Press.

Pomeroy, J. W., & Gray, D. M. (1995). Snowcover accumulation, relocation and management. *Bulletin of the International Society of Soil Science* no, 88(2).

Potschka, A., Kirches, C., Bock, H. G., & Schlöder, J. P. (2010). Reliable solution of convex quadratic programs with parametric active set methods. *Interdisciplinary Center for Scientific Computing, Heidelberg University, Im Neuenheimer Feld*, *368*, 69120.

Pulliainen, J. T., Grandell, J., & Hallikainen, M. T. (1999). HUT snow emission model and its applicability to snow water equivalent retrieval. *Geoscience and Remote Sensing, IEEE Transactions on*, *37*(3), 1378-1390.

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, *130*(1).

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, *31*(11), 1411-1418.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., ... & Molod, A. (2011). MERRA: NASA"s Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, *24*(14).

Robinson, D. A., Dewey, K. F., & Heim Jr, R. R. (1993). Global snow cover monitoring: An update. *Bulletin of the American Meteorological Society*, 74(9), 1689-1696.

Rojas, R. (1996). *Neutral Networks: A Systematic Introduction*. Springer.

Ryan, W. A., Doesken, N. J., & Fassnacht, S. R. (2008). Evaluation of ultrasonic snow depth sensors for US snow measurements. *Journal of Atmospheric & Oceanic Technology,* 25(5).

Rychetsky, M. (2001). Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches. Shaker.

Sabins, F. F. (2007). Remote sensing: principles and applications. Waveland Press.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *3*(3), 210-229.

Sarle, W. S. (1995). Stopped training and other remedies for overfitting. In *Proceedings of the 27th Symposium on the Interface of Computing Science and Statisfi. pp. 352-360. Interface Foundation of North America, Fairfax Station. VA, USA*.

Sarle, W. S. (1997). Neural Network FAQ, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp. ai. neural-nets.

Saunders, R., Matricardi, M., & Brunel, P. (1999). An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quarterly Journal of the Royal Meteorological Society*, 125(556), 1407-1425.

Scherer, D., Hall, D. K., Hochschild, V., König, M., Winther, J. G., Duguay, C. R., ... & Walker, A. E. (2005). Remote sensing of snow cover. *Geophysical Monograph Series*, *163*, 7-38.

Shuttleworth, W. J. (2012). Terrestrial hydrometeorology. John Wiley & Sons.

Siegel, R., & Howell, J. R. (1992). Thermal radiation heat transfer.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P. Y., ... & Juditsky, A. (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, *31*(12), 1691-1724.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222.

Sommerfeld, R. A., & LaChapelle, E. R. (1970). The classification of snow metamorphism. British Glaciological Society.

Souza, C. R. (2010). Kernel functions for machine learning applications. *Creative Commons Attribution-Noncommercial-Share Alike*, *3*.

Sturm, M., Holmgren, J., & Liston, G. E. (1995). A seasonal snow cover classification system for local to global applications. *Journal of Climate,* 8(5), 1261-1283.

Suykens, J. A., Vandewalle, J., & De Moor, B. (2001). Optimal control by least squares support vector machines. *Neural Networks*, *14*(1), 23-35.

Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, *39*(1), 43-62.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J. P., ... & Bojkov, B. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment,* 115(12), 3517-3529.

Tedesco, M., & Narvekar, P. S. (2010). Assessment of the NASA AMSR-E SWE Product. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 3(1), 141-159.

Tedesco, M., Kim, E. J., England, A. W., De Roo, R. D., & Hardy, J. P. (2006). Brightness temperatures of snow melting/refreezing cycles: Observations and modeling using a multilayer dense medium theory-based model. *Geoscience and Remote Sensing, IEEE Transactions on,* 44(12), 3563-3573.

Tedesco, M., Pulliainen, J., Takala, M., Hallikainen, M., & Pampaloni, P. (2004). Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data. *Remote sensing of Environment*, *90*(1), 76-85.

Tsang, I. W., Kwok, J. T., Cheung, P. M., & Cristianini, N. (2005). Core Vector Machines: Fast SVM Training on Very Large Data Sets. *Journal of Machine Learning Research*, *6*(4).

Tsang, L., Chen, Z., Oh, S., Marks, R. J., & Chang, A. T. C. (1992). Inversion of snow parameters from passive microwave remote sensing measurements by a neural network trained with a multiple scattering model. *Geoscience and Remote Sensing, IEEE Transactions* on, 30(5), 1015-1024.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, *49*(11), 1225-1231.

Tzeng, F. Y., & Ma, K. L. (2005, October). Opening the black box-data driven visualization of neural networks. In *Visualization, 2005. VIS 05. IEEE* (pp. 383-390). IEEE.

Ulaby, F. T., & Stiles, W. H. (1980). The active and passive microwave response to snow parameters: 2. Water equivalent of dry snow. *Journal of Geophysical Research:* Oceans (1978–2012), 85(C2), 1045-1049.

Valyon, J., & Horváth, G. (2003). A Weighted Generalized LS-SVM. *Periodica Polytechnica, Electrical Engineering*, *47*(3), 229-251.

Vapnik, V. (1998). Statistical learning theory. 1998.

Weston, J. (1998). Support vector machine (and statistical learning theory) tutorial. *NEC Labs America*.

Willis, R., & Yeh, W. W. (1987). Groundwater systems planning and management.

Wilson Bentley (1902). NOAA's National Weather Service (NWS) Collection.

Y. Cao, X. Yang, and X. Zhu (2008). Retrieval Snow Depth by Artificial Neural Network Methodology from Integrated AMSR-E and In-situ Data – A Case Study in Qinghai-Tibet Plateau. Chin. *Geogra. Sci.,* 18:356–360.

Yao, J. T. (2003, July). Sensitivity analysis for data mining. In *Fuzzy Information Processing Society, 2003. NAFIPS 2003. 22nd International Conference of the North American* (pp. 272-277). IEEE.

Živković, Ž., Mihajlović, I., & Nikolić, Đ. (2009). Artificial neural network method applied on the nonlinear multivariate problems. *Serbian Journal of Management*, *4*(2), 143-155.

ABBREVIATIONS AND ACRONYMS

AE_Land3	Level-3 land surface product
AMSR-E	Advanced Microwave Scanning Radiometer
ANN	Artificial neural network
Catchment	NASA Catchment land surface model
DA	Data assimilation
EASE-Grid	Equal Area Scalable Earth Grid
EnKF	Ensemble Kalman Filter
EnKS	Ensemble Kalman Smoother
EOS	NASA's Earth Observing System
GMAO-LDAS	NASA Global Modeling and Assimilation Office Land Data Assimilation System
HDF-EOS	Hierarchical Data Format – Earth Observing System
IMS	Interactive Multisensor Snow and Ice Mapping System
KF	Kalman Filter
MERRA	Modern-Era Retrospective analysis for Research and Applications
MODIS	Moderate Resolution Imaging Spectroradiometer
MODSCAG	MODIS Snow Covered Area and Grain Size
MSE	Mean squared error
NSC	Normalized Sensitivity Coefficient
NA	North America
NDVI	Normalized Difference Vegetation Index

NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center
NWS	National Weather Service
PMW	Passive microwave
QP	Quadratic program
RBF	Radial basis function
RTM	Radiative transfer model
SCE	Snow cover extent
SLWC	Snow liquid water content
SMMR	Scanning Multichannel Microwave Radiometer
SSM/I	Special Sensor Microwave/Imager
SVM	Support vector machine
SWE	Snow water equivalent
Tb	Brightness temperature
TGI	Temperature gradient index