ABSTRACT

Title of Thesis: SENSITIVITY ANALYSIS OF SUPPORT VECTOR MACHINE PREDICTIONS OF PASSIVE MICROWAVE BRIGHTNESS TEMPERATURES OVER SNOW-COVERED TERRAIN IN HIGH MOUNTAIN ASIA

Jawairia A. Ahmad, Master of Science, 2018

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

Spatial and temporal variation of snow in High Mountain Asia is very critical as it determines contribution of snowmelt to the freshwater supply of over 136 million people. Support vector machine (SVM) prediction of passive microwave brightness temperature spectral difference (ΔT_b) as a function of NASA Land Information System (LIS) modeled geophysical states is investigated through a sensitivity analysis. AMSR-E ΔT_b measurements over snow-covered areas in the Indus basin are used for training the SVMs. Sensitivity analysis results conform with the known first-order physics. LIS input states that are directly linked to physical temperature demonstrate relatively higher sensitivity. Accuracy of LIS modeled states is further assessed through a comparative analysis between LIS derived and Advanced Scatterometer based

Freeze/Melt/Thaw categorical datasets. Highest agreement of 22%, between the two datasets, is observed for freeze state. Analyses results provide insight into LIS's land surface modeling ability over the Indus Basin.

SENSITIVITY ANALYSIS OF SUPPORT VECTOR MACHINE PREDICTIONS OF PASSIVE MICROWAVE BRIGHTNESS TEMPERATURES OVER SNOW-COVERED TERRAIN IN HIGH MOUNTAIN ASIA

by

Jawairia Ashfaq Ahmad

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
2018

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

Acknowledgements

I am most obliged to my advisor, Dr. Forman, for all the help and support he has rendered me during the last two years. I further extend my gratitude to my thesis committee members, Dr. McCuen and Dr. Brubaker, for their invaluable feedback that has helped me grow professionally as well as personally.

I appreciate my UMD research group members; Gaohong, Jongmin, Meg, Jing, Yuan, Lizhao, and Yonghwan, for always coming to my aid.

To my mother and sisters; Maria, Sarah and Khudaija, I owe more than words can express. I am really grateful for the patience with which they listened to my constant string of woes and complaints. Thank you to my father and Uncle Siraj for believing in me and for always being in my team.

A big thank you to my roommates, Fatima and Wardah, for sharing my laughter and tears alike.

Lastly, I would like to acknowledge the support I received from the Fulbright Scholarship Program and hope many other students like me get to experience the phenomenon that is known as ‘Fulbright’!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMSR-E	Advanced microwave scanning radiometer for earth observing systems
ASCAT	Advanced Scatterometer
AT	Air temperature
BLST	Bottom-layer snow temperature
CETB	Calibrated enhanced resolution brightness temperature
F/M/T	Freeze/Melt/Thaw
HMA	High mountain Asia
LAI	Leaf area index
LIS	Land Information System
LSM	Land surface model
Noah MP	Noah multi-parametrization
SD	Snow density
SLWC	Snow liquid water content
SM	Soil moisture
SWE	Snow water equivalent
SVM	Support vector machine
T_b	Brightness temperature
TLST	Top-layer snow temperature
TLSO	Top-layer soil temperature
VT	Vegetation temperature
ΔT_b	Brightness temperature spectral difference (difference between brightness temperature observed at different frequencies)
ΔT_b (18.7V-36.5V)	Difference in brightness temperature measured at 18.7GHz (vertical polarization) and 36.5GHz (vertical polarization) frequencies

Chapter 1: Introduction

The following sections introduce important components of this study and the motivation that prompted this undertaking.

1.1. Importance of Snow in High Mountain Asia (HMA)

Snow is a critical component of the hydrologic cycle within the Earth's system. Under changing climatic conditions, the importance of snow quantification and monitoring has increased significantly. It is especially vital for areas that have experienced perennial snow up till now.

High Mountain Asia (HMA), consisting of Hindukush, Karakoram, and Himalayan mountain ranges, lies between 27.5^{o}N to 45.5^{o}N and 69.5^{o}E to 101.50^{o}E. Known as the 'third pole', it experiences both perennial and ephemeral snow within its bounds (depending upon the location and elevation). It serves as one of the primary sources of fresh water supply for over 800 million people, primarily in south Asia [1]. The principal source of this fresh water is the snow and glacier melt during the summer months.

Run-off flow in the Himalayan rivers (i.e. rivers originating in the Himalayan mountain ranges) is critically dependent on this snow and ice melt. Agrarian economies of the population residing in the Himalayan river basins are greatly influenced by the cryospheric conditions of HMA [2]. In Figure 1.1, this water dependency of the individual basin populations is outlined [1].

There has been increasing evidence regarding the loss of snow cover and glaciers under evolving climatic scenarios in this region [3] [4] [5]. Thus, for the well-being and sustenance of this area and its inhabitants, it is important to carry out a comprehensive hydrologic and cryospheric assessment of this area.

1.1.1. Hydrologic Modelling of HMA

Despite its importance in life sustenance in this area, there is still considerable uncertainty regarding the total amount of snow in HMA and its spatial and temporal

variation. HMA terrain has high spatial variation in elevation, with sizable differences between the high (highest peak >8500m) and low altitudes (<2500m). Simultaneous presence of mountains, valleys, and plateaus render the hydrologic modeling of this region quite complex. Current global land surface models have coarse spatial resolution that cannot resolve this complexity. Apart from that, there is little prior ground data in this region, which decreases the initializing accuracy of the land surface models. These factors create substantial discrepancies between the actual ground conditions and the land surface modeled parameter and state estimates.

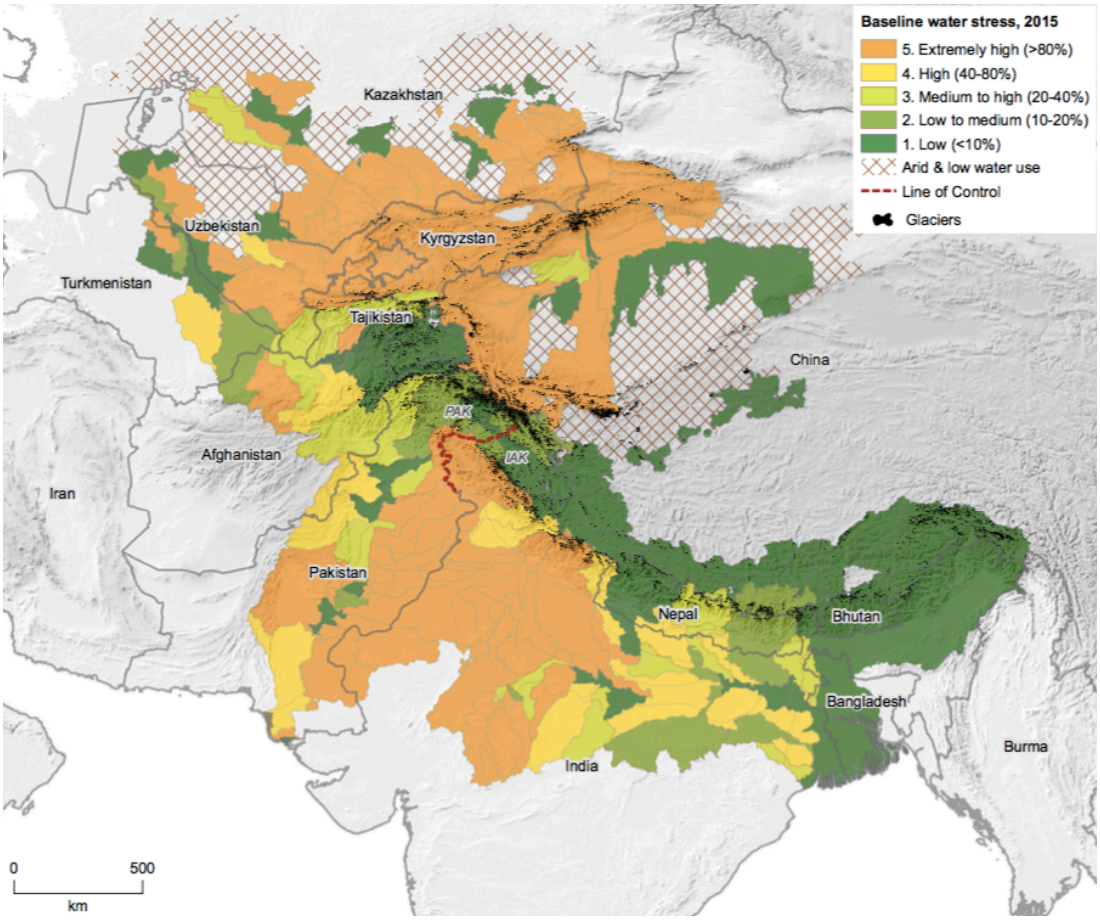

Figure 1.1. Baseline water stress (total annual water withdrawals as a percentage of the total annual available blue water) for HMA river basins in 2015. IAK refers to Indian administered Kashmir, PAK refers to Pakistan administered Kashmir [1].

Data assimilation is a technique used for increasing model accuracy through integration of observation data. Unfortunately, for HMA ground observations are few

and far. To counter this situation, satellite based remote sensing data can be used. Brightness temperature assimilation has been successfully applied for increasing SWE estimation accuracy in North America [6] [7] [8]. Land surface model estimated inputs were used to predict brightness temperatures using machine learning techniques. In the data assimilation framework, machine learning algorithms acted as the measurement operator and mapped the land surface model estimated states into the satellite observed brightness temperature space.

1.1.2. NASA Land Information System

In this study, NASA Land Information System (LIS) was used to model the hydrologic cycle over the Indus Basin. LIS is a software framework for high performance terrestrial hydrology modeling and data assimilation developed by NASA [9]. It integrates the use of:

- Land surface models
- Satellite and observed data
- Data assimilation techniques
- High performance computing tools

The land surface model used for this study is Noah-MP. Noah-MP provides layered modeling of the snow pack and is thus suitable for this study [10]. A model open-loop (stand-alone model estimation) was run within the LIS framework. Modern-Era Retrospective analysis for Research and Applications version-2 (MERRA2) forcings [11] were used as boundary conditions for the model. An important benefit of using LIS was the comparatively finer grid resolution (compared to other large scale land surface modeling frameworks) it provided with a grid cell size equal to 0.01^{O}x0.01^{O}. Further detail is provided in Chapters 2 and 3.

1.2. Passive Microwave Radiation

Electromagnetic radiation with wavelengths ranging from one meter to one millimeter and frequencies between 300MHz and 300GHz, is usually termed as microwave. However, there is some ambiguity regarding the exact frequency boundaries of the microwave radiation band. Some sources define the band frequency

limits to be 1 and 100 GHz, corresponding to wavelengths 300 mm and 3 mm respectively [12] [13] [14].

Microwave radiation inherently emitted by the Earth's surface and measured by a sensor is called passive microwave (PMW). Passive microwave radiation, measured by different Earth orbiting satellites, has been used to retrieve information regarding various geophysical properties of the Earth's surface such as precipitation, soil moisture [15], wind speed, and snow water equivalent [16].

1.2.1. Brightness Temperature

Brightness temperature is the temperature a black body in thermal equilibrium with its surroundings would have to possess to emit the same intensity of radiation (at a specific frequency) as a grey body object.

According to the Raleigh-Jean approximation for microwave radiation, brightness temperature is primarily dependent on emissivity (ε) and physical temperature (T_{phy}) of the emitting body [17].

$$T_b \approx \varepsilon * T_{phy} \quad (1.1)$$

Emissivity is a dimensionless inherent attribute. It is wavelength dependent and is affected by the dielectric constant of the material. It varies from 0 to 1 ('0' signifies no emission from the surface, while '1' represents total emission of radiation at a particular wavelength). A perfect black body has emissivity equal to 1.

Brightness temperature (T_b) is the fundamental parameter measured by passive microwave radiometers. Brightness temperatures, measured at different microwave frequencies, are utilized in different satellite retrieval algorithms.

1.2.2. Remote Sensing of Snow using PMW

Passive microwave remote sensing of snow utilizes the wavelength dependency of brightness temperature in the microwave spectrum. PMW remote sensing of snow is dependent on preferential scattering of microwave radiation at a higher frequency (18.7GHz or 36.5GHz) compared to a lower frequency (10.7GHz or 18.7GHz) by the snow pack. This preferential scattering at higher frequency decreases the emissivity

and hence lowers the corresponding measured brightness temperature [18]. In an idealized scenario, the snow layer depth is directly related to the scattering induced by the snow pack and inversely related to the ultimately measured brightness temperature at the corresponding frequency (Figure 1.2).

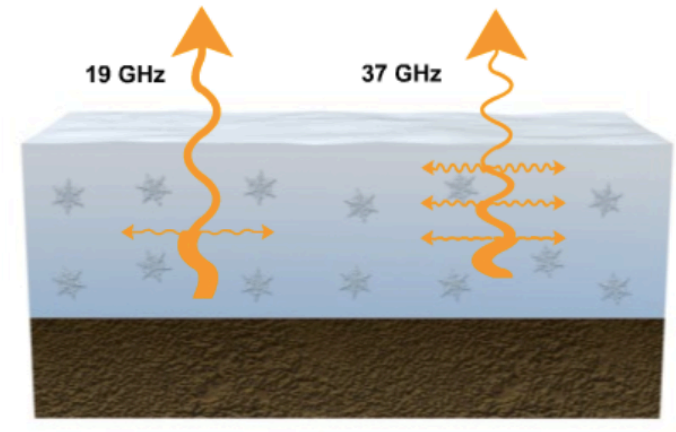

Figure 1.2. Preferential scattering of microwave radiation having frequency 37GHz compared to microwave radiation with frequency 19GHz by the snow pack [19].

Since this scattering is more prominent at higher frequencies, the brightness temperature measured at these frequencies will be lower. The difference between T_b at 18.7GHz and 36.5GHz is often used in snow water equivalent (SWE) retrievals [16]. This difference is representative of the amount of snow water equivalent present. The study detailed in this thesis is done for snow covered terrain in high mountain Asia and will contribute towards prediction and later assimilation of T_b spectral difference to improve SWE estimates in that region.

1.2.3. Advanced Microwave Scanning Radiometer – Earth Observing System

The Advanced Microwave Scanning Radiometer for Earth Observing Systems (AMSR-E) is a twelve-channel, six-frequency, passive-microwave radiometer, Figure 1.3. It is aboard the AQUA polar orbiting sun synchronous satellite. Some instrument specifications are presented in Table 1.1.

AMSR-E being a passive microwave radiometer is quite suitable for snow retrievals, especially considering the 10.65GHz, 18.7GHz and 36.5GHz frequency

bands [20] [16]. These bands are relatively unaffected by the presence of clouds. Microwave radiations at these frequencies have wavelengths that are large enough to pass through the atmosphere with minimal attenuation, irrespective of day or night. Higher frequency microwave radiations are relatively more prone to atmospheric attenuation.

Table 1.1. AMSR-E instrument specifications [21].

Platform	AQUA
Launch date	May 4, 2002
End date	October 4, 2011
Swath width	1450 km
Frequencies (GHz) Dual Polarization	6.9, 10.7, 18.7, 23.8, 36.5, 89.0
Sample footprint sizes (km)	74 x 43 (6.9 GHz); 14 x 8 (36.5 GHz); 6 x 4 (89.0 GHz)

AMSR-E observations are also used for retrieving various geophysical parameters such as precipitation rate, cloud water, water vapor, sea surface winds, sea surface temperature, ice, and soil moisture [15].

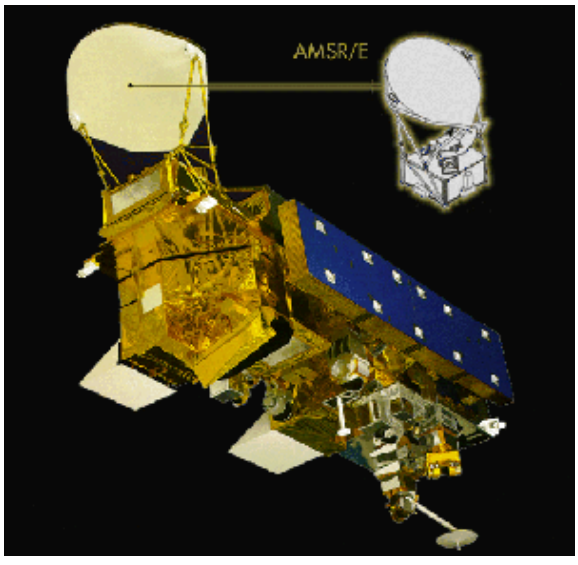

Figure 1.3. AMSR-E instrument aboard the AQUA satellite [22].

National snow and ice data center (NSIDC) provides processed and quality controlled brightness temperature measurements from AMSR-E that were used in this study [23]. End date in Table 1.1 marks the termination of operational data collection.

1.3. Machine Learning in Hydrology

Application of machine learning techniques in hydrology has witnessed an increase in the past years. These applications include usage of machine learning techniques to predict Earth's geophysical parameters such as rainfall [24], streamflow [25], and soil moisture [26]. More recently, these techniques have been utilized within data assimilation frameworks for improvement of land surface model estimates [6] [7] [8] [27]. In this study, we utilize a machine learning technique (Support Vector Machines) to predict brightness temperatures using geophysical states from a land surface model as training and prediction data, and then analyze the importance of each input state through a sensitivity analysis.

1.3.1. Support Vector Machine (SVM)

Support vector machine is a machine learning technique that typically involves two steps: 1) training of SVM utilizing two datasets (input and target) to learn prediction patterns from, and 2) use of the resulting well-trained SVM for prediction [28]. SVM 'training' consists of identifying and attuning the support vectors to predict a known target given known input [29]. This is a black-box method that is based on statistical learning theory and falls under the realm of 'supervised learning'.

In this scheme, our input states consist of the land surface model (LIS) estimated geophysical states, while the output is in the form of brightness temperature spectral difference. AMSR-E brightness temperature (T_b) measurements were utilized during the training process as training targets. The trained SVM is utilized to transport the model estimated geophysical states into the observation (T_b) space.

1.3.2. SVM Prediction Framework

SVM application has two stages: training and prediction. SVM training is performed for each pixel in the domain on a fortnightly interval. Training and prediction datasets are kept separate i.e., training is performed using data from all the years except the one that will ultimately be used for prediction. During the training stage, SVMs are setup using LIS (Noah-MP) modeled states as input and AMSR-E brightness temperature spectral difference (ΔT_b) as the training targets. LIS model resolution ($0.01^O x 0.01^O$) is much finer than the AMSR-E satellite data grid resolution (25km x 25km). To achieve consistency between the model estimated states and the satellite data, LIS states were scaled up to the coarser satellite grid.

The prediction mechanism thus comprised of three main components:

- Input: LIS model estimate (geophysical states)
- Predicting model: SVM
- Output: Brightness temperature spectral difference (ΔT_b)

A sensitivity analysis was carried out to assess the relative effect of each LIS state on the predicted output. Further detail is provided in Chapter 3.

1.4. Study Objectives

SVM is a statistical model based on statistical learning theory. It is a black box method with a plethora of nuances and complexity. The study goal was to infer physical meaning from this statistical model and to analyze whether the SVM predictions conform with the first-order physics. This goal was achieved through the following objectives:

1.4.1. Sensitivity Analysis of SVM Predictions

Normalized sensitivity coefficients (NSCs) were utilized in the sensitivity analysis. NSCs were calculated by perturbing each LIS input state (one at a time) while keeping the others at their default value and computing the normalized change in the predicted value. NSC has a magnitude and a sign. Importance of an input state is

represented by the NSC magnitude. NSC sign defines whether the relationship between the input state and the predicted ΔT_b is directly proportional or inversely proportional.

1.4.2. Relative Importance of LIS States for SVM Prediction

The second study objective concerns the identification of LIS states that are most important for maintaining the accuracy of SVM prediction. This objective is intended to contribute towards selection of the most important LIS input states that will be utilized for SVM prediction within a data assimilation framework, without compromising significantly on the prediction accuracy.

1.4.3. SVM Prediction Sensitivity to SWE

The overarching project's future work plan includes the improvement of SWE estimation in the HMA region using SVM within a data assimilation framework. According to the brightness temperature assimilation framework, any noticeable improvement can only be achieved if the SVM has significant relative sensitivity to SWE. The third objective includes analysis of the SVM prediction sensitivity to LIS estimated SWE using different number of predictors.

1.4.4. Assessment of LIS Modeled States

In order to refine the SVM prediction accuracy, the prediction mechanism can be improved in two ways:

a) Analyzing the predicting mechanism through a sensitivity analysis
b) Improving the accuracy of input states

Accuracy of input states can be examined using different approaches. One such approach is to compare the land surface model geophysical state estimation with other measurement based data. Improving the geophysical input is expected to lead to better prediction. In this study, we compare Advanced Scatterometer (ASCAT) based Freeze/Melt/Thaw categorical dataset (satellite based 'measurement') with a LIS derived Freeze/Melt/Thaw product.

Chapter 2: Background and Literature Review

2.1. Evaluation of Snow and Ice Melt in High Mountain Asia

Snow in high mountain Asia is still a relatively unexplored avenue. The spatial and temporal snowfall patterns, snow on glaciers, and the presence and evolution of massive glaciers are complex phenomenon that are very important to unravel. Cryosphere in HMA is mainly composed of perennial glaciers and seasonal snow. The seasonal snow cover controls in part the water supply to the Himalayan rivers, especially on the short-term scale e.g. seasonal cycle. The long-term water supply is more dependent on the prevalent glacier mass [1]. Fluctuations in seasonal snow are thus critical for the short-term water supply management in that area.

The three major Himalayan rivers are the Indus, Ganges and Brahmaputra, Figure 2.1. As discussed in Chapter 1, dependency of the population residing in the Himalayan river basins renders snow assessment vital for this region. In this study, we have restricted our domain to the Indus basin.

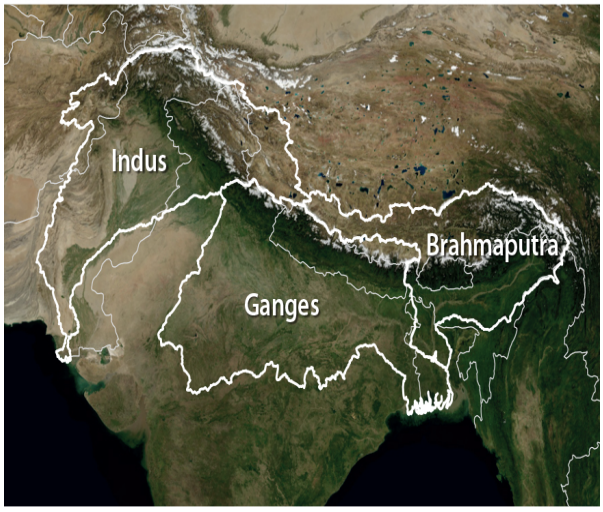

Figure 2.1. Three main Himalayan river basins in South-Asia [30].

Snowmelt valuation is very important for seasonal streamflow prediction [31]. Considering the water supply system of HMA, where bulk of the population is reliant

on river flow for domestic, commercial and industrial use, that importance is heightened considerably. Considerable population resides in these river basins. In case of river overflow, due to increased snow melt, the basins are flooded, resulting in loss of life and livelihood.

Another importance of accurate estimation of snow in this region is the fact that several dams are built on the Himalayan rivers. For the smooth and safe functioning of these dams (population resides downstream), the amount of water that would be generated by the snowmelt must be known beforehand.

2.2. Remote Sensing of Snow

Remote sensing of snow has been performed using various wavelengths of the electromagnetic spectra, depending on the snow property or attribute being studied. Moderate Resolution Imaging Spectroradiometer (MODIS) data has been used to derive snow cover products [32]. MODIS collects data within the infrared and visible bands (0.4 to 14.0 μm). NASA's Airborne Snow Observatory was launched to study snow using an imaging spectrometer (380 to 1050 nm) and a scanning LIDAR (1064 nm) [33]. Backscattering effects in active remote sensing have also been used to study snow [34]. Passive microwave remote sensing principles have been utilized in the generation of various snow retrieval products [35] [36].

Different remote sensing techniques have different strengths. The selection criterion is based on the snow quantity under consideration and the availability of pertinent data and resources. In this study, we focus on passive remote sensing of snow.

2.2.1. Passive Microwave Remote Sensing of Snow

Passive remote sensing of snow is based on the measurement of brightness temperature at a lower (e.g., 10.65 or 18.7 GHz) and a higher frequency (e.g., 36.5 GHz), within the microwave spectrum. Brightness temperature, in the microwave band, is primarily dependent on the emissivity and physical temperature of the emitting surface.

In remote sensing using satellite data, the attribute measured at the top of the atmosphere is usually different from the value at the surface of the earth. Similarly, in

our case the brightness temperature measured by the AMSR-E instrument, at an elevation of ~700 km above the Earth's surface, is not necessarily equal to the T_b value near the Earth's surface.

2.2.1.1. Ancillary Effects on Brightness Temperature

Ancillary influence on the AMSR-E measured brightness temperature is due to several simulated and parameterized processes. Some of these processes are:

- Effect of snow depth

Assuming the snow pack is dry, a higher snow pack depth will result in higher scattering of the upward radiation. T_b and snow depth are generally inversely proportional. Greater snow depth renders a lower brightness temperature, Figure 2.2. This effect is more prominent on T_b measured at higher frequencies.

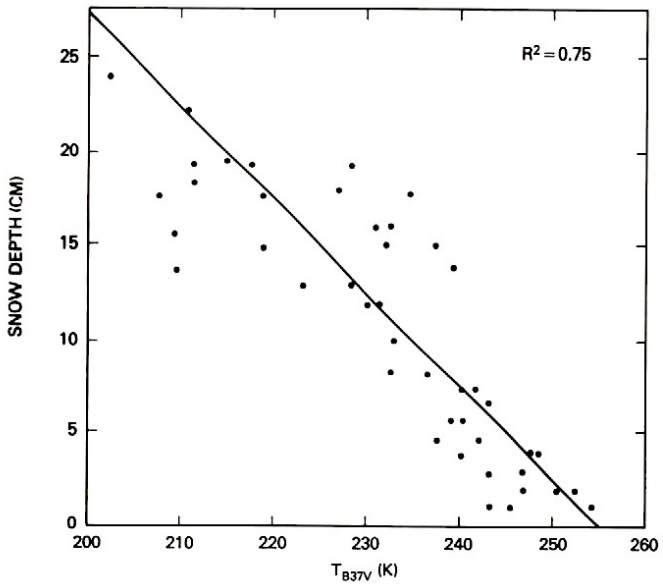

Figure 2.2. Brightness temperature, as measured at 37GHz (vertical pol.) by the Scanning Multi-channel Microwave Radiometer, versus snow depth for the Russian steppes, 15 February 1979. R^2 represents the coefficient of determination [18].

- Effect of snow grain size

Armstrong et al. discuss the relevance of snow grain size to passive microwave study of snow [37]. Scattering of the upward radiation is proportional to the snow grain

size. For a snow-covered surface, T_b in the microwave band usually decreases with increasing grain size. Figure 2.3 [20] presents a plot of T_b versus SWE for snow-packs of different mean grain size. The plots corroborate the inverse relationship between T_b and snow grain size i.e. for a given amount of SWE, the T_b observed for a snow pack having larger mean grain size is less than the T_b observed for a snow pack with lower mean grain size.

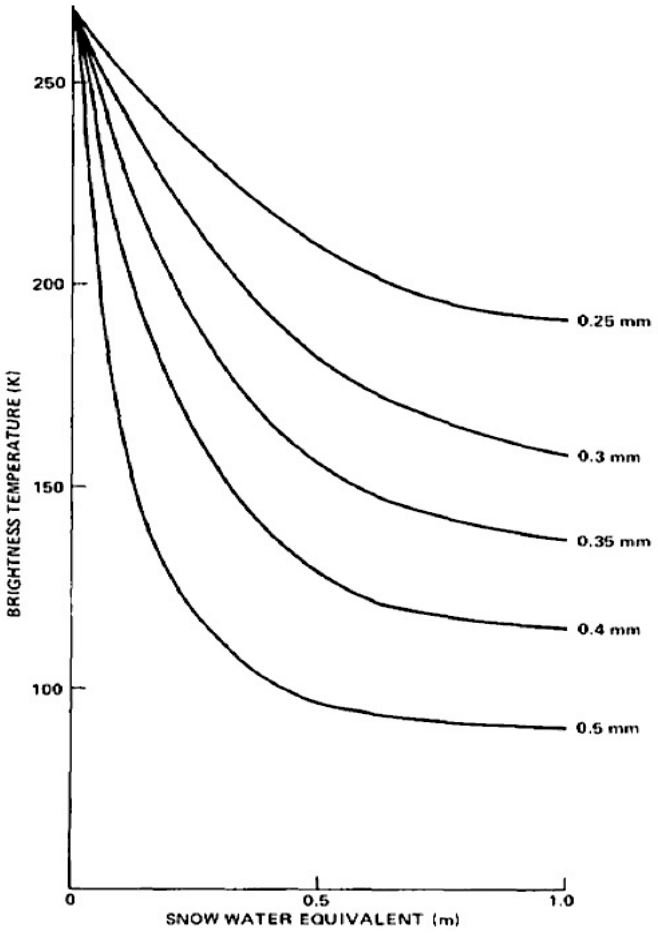

Figure 2.3. Relationship between 37GHz brightness temperature (vertical polarization) and snow water equivalent (m) as a function of snow grain diameter (mm) [20].

- Effect of ice layer

Ice layer within the ground (frozen soil) or within the snow pack will introduce greater scattering and hence lower the T_b. This can lead to over-estimation of snow depth and SWE.

- Wet snow effect

Wet snow complicates the snow-pack versus T_b dynamics. Wet snow emits significant radiation itself and thus increases the ultimately measured T_b [38]. This is due to the higher dielectric constant of water as compared to dry snow. Presence of liquid water in the snow-pack increases the dielectric of that snow mass. Contribution to the brightness temperature from liquid water as well as snow can result in erroneous estimation of snow retrievals.

- Effect of vegetation

Vegetation effect includes addition as well as attenuation of the emitted radiation. Presence of vegetation cover hinders the upwards transport of radiation, but can also increase the total upward microwave long-welling by the addition of microwave radiation emitted by the vegetation itself. This behavior is depicted in Figure 2.4. Accounting and removing the vegetation effect from the total T_b measured at the top of atmosphere is a complex process that can introduce considerable uncertainty in the ultimate T_b value recorded [39].

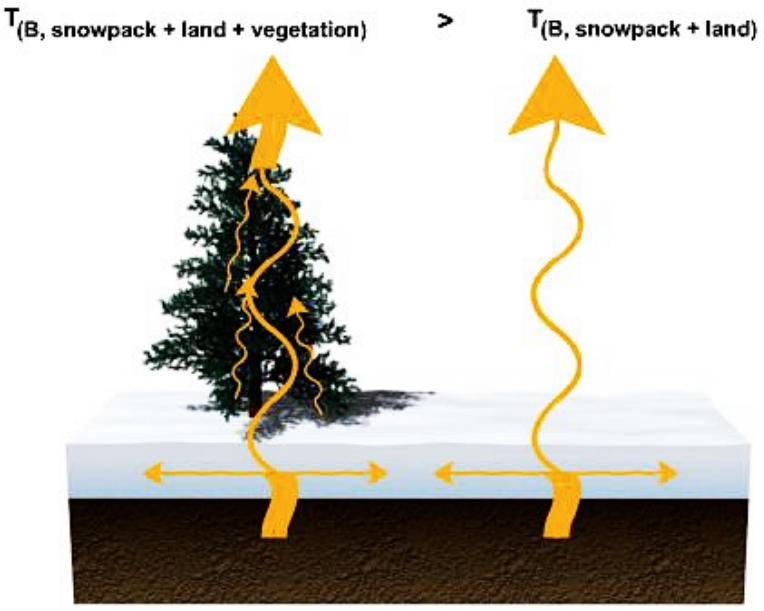

Figure 2.4. Effect of vegetation and snow pack on microwave radiation emitted by the land surface [19].

- Atmospheric attenuation

Clouds and aerosols present in the atmosphere not only attenuate the upward microwave radiation but also emit radiation themselves. The microwave bands generally utilized in remote sensing of snow (i.e., 10.67GHz, 18.7GHz, and 36.5GHz) lie within the atmospheric window and are not significantly affected by the presence of clouds and aerosols [40].

2.2.2. Why Passive Microwave Remote Sensing of Snow?

Although PMW remote sensing of snow faces numerous challenges, yet, there are a number of benefits of using this technique, some of which are:

- Microwave bands generally utilized in snow retrievals (i.e. 10.67GHz, 18.7GHz, and 36.5GHz) are not significantly influenced by the presence of clouds (atmospheric window).
- It can be carried out at night-time as well as day-time.
- It renders greater continuity of data and does not suffer from diurnal data gaps.
- Global datasets can be developed from remotely sensed global satellite data.
- Analyses can be performed on regional as well as continental scales.

2.2.3. AMSR- E Level-3 Brightness Temperature Dataset

The data set used in this study is developed through the NASA-sponsored 'Calibrated Passive Microwave Daily Equal-Area Scalable Earth Grid 2.0 Brightness Temperature Earth System Data Record' project.

The Calibrated Enhanced Resolution Brightness Temperature (CETB) dataset is an enhanced-resolution, gridded passive microwave record [41]. Multi-sensor, multi-decadal time series of high-resolution radiometer products are generated. The newest available Level-2 satellite passive microwave records from 1978 to the present are used in developing these time series. Since our study period lies within years 2002-2011, only AMSR-E data is used. AMSR-E was selected due to its comparative higher resolution relative to other global passive microwave instruments available.

CETB data processing comprises of two general steps. First, dataset pre-processing is carried out for spatial and temporal selection, followed by gridding and

reconstruction of the data. Reconstruction algorithms are employed to resolve spatial and temporal resolution issues. 12-hour averaged values are given on a global scale for 3.125km to 25km pixel spatial resolution.

In this study, the CETB data are used as provided by the developing team i.e. no quality control checks have been applied. No correction has been applied for ancillary effects on T_b by various sources (discussed in Section 2.2.1.1). Therefore, the existence of discrepancies in the T_b data used for SVM training is known and acknowledged beforehand.

2.3. Land Information System

NASA's Land Information System (LIS) was used to model the hydrologic cycle over the Indus basin. LIS is a software framework that assimilates satellite and ground-based observational data with advanced land surface models and computing tools to estimate land surface states and fluxes [9]. LIS can deal with the challenges introduced by the large scale and high resolution of the model outputs through scalable, high performance computing. LIS comprises of three main components: Land surface data toolkit, LIS core, and Land validation toolkit [42].

2.3.1. Main Components of LIS

2.3.1.1. Land Surface Data Toolkit - LDT

LDT functions as the front-end processor for the LIS core. It processes data inputs for land surface models on to a common grid domain [9].

2.3.1.2. LIS Core

LIS Core handles the land surface modeling part [9]. LIS supports various land surface models that predict water, energy, and biogeochemical processes through equations linking the soil, vegetation, and snowpack medium, Figure 2.5. Model results are averaged to the required temporal and spatial scales according to the study demands [42]. Three types of inputs are required to run a land surface model:

1. Initial conditions that describe the state of the land surface at the start of the simulation.

2. Boundary conditions that define the upper (atmospheric) fluxes and the lower (soil) fluxes.
3. Parameters that characterize the soil, vegetation, topography, and other surface properties.

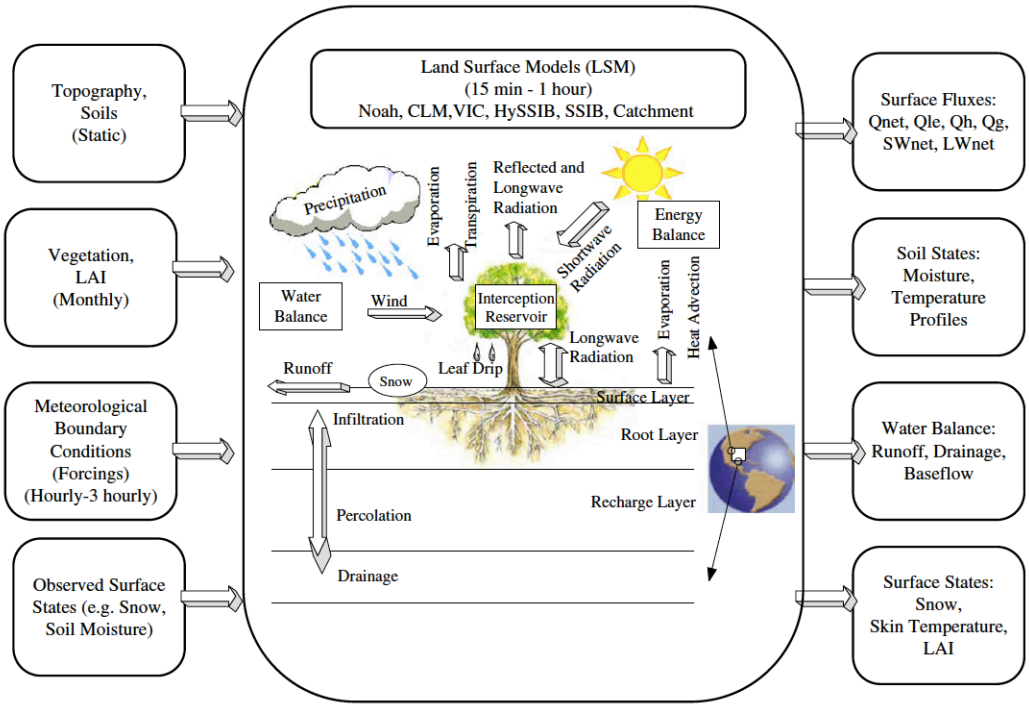

Figure 2.5. Depiction of land surface modeling as carried out in LIS [9].

2.3.1.3. Land surface Verification Toolkit - LVT

Land surface Verification Toolkit (LVT) is devised for the evaluation and comparison of outputs generated by the Land Information System (LIS) with respect to other measurements or datasets [9].

2.3.2. Noah MP

The land surface model used within the LIS framework for this study was Noah multi-parameterization (Noah–MP). It was particularly relevant for this study since it incorporates snow pack development and snow layering features. Noah-MP has been used in various simulation studies and has rendered positive validation results with respect to in-situ data [43].

Cold season processes such as snow cover development, snow albedo effect on shortwave, heat and moisture exchange between land surface and atmosphere through snow pack are characterized in detail within the model [10]. Patchy snow cover and relevant atmospheric processes (e.g. surface sensible heat flux, increase in surface albedo) are represented at sub-grid level. Frozen soil dynamics such as the effect of latent heat release during soil water freezing in winter and supplementary precise estimation of temperatures during periods of thaw are highlighted in this model. Some of the other important processes incorporated in the Noah–MP land surface model are:

- Separate soil layer dynamics (temperature, thermal conduction, soil moisture)
- Range of spatial and temporal scales
- Canopy conductance formulation
- Bare soil evaporation and vegetation phenology
- Surface runoff and infiltration
- Continuous self-cycling of soil moisture and temperature
- Coupled simulation mode i.e. three-dimensional operational mesoscale analysis
- Snow accumulation and ablation

2.4. Support Vector Machines

Machine learning is a technique in which systems acquire the ability to learn automatically, without being explicitly programmed. Systems are programmed to optimize a performance criterion using test data [44]. A common form of machine learning is the supervised learning. It is based on attaining a generalization ability, which refers to the capability of estimating an appropriate answer for unlearned questions [28].

Support Vector Machines (SVM) provide a supervised learning method that has proved to be quite successful in numerous applications [45]. The theoretical foundation of SVM is based on statistical learning theory, or Vapnik-Chervonenkis theory, which was developed by Vladimir Vapnik and Alexey Chervonenkis [46] [47] [48]. SVM has been used as a data-classifier as well as a regression tool.

2.4.1. Theoretical Basis of SVM

The SVM learning problem is based on the assumption that there is some unknown and non-linear dependency between an input vector '$\mathbf{x_i}$' and scaler output 'y_i' [49]. Our only source of information is the training data set $\{(x_1,y_1) , (x_2,y_2) , (x_3,y_3) , \dots , (x_l,y_l)\} \subset \mathcal{X} \times \mathbb{R}$, where $\mathcal{X}$ denotes the input pattern space, $\mathbb{R}$ specifies the real space to which y_i belongs and 'l' is equal to the number of training data pairs [46]. Using this training data set, the relationship between the input vector and the output scaler is estimated.

2.4.2.1. Linear objective function

In $\varepsilon -$ SV regression [48], the goal is to formulate an objective function **f(x)** that is as flat as possible, with at most $\boldsymbol{\varepsilon}$ deviation from the training targets y_i. A linear form of such a function can be represented as eq. 2.1 [29]:

$$\boldsymbol{f(x)} = \langle \boldsymbol{w} , \boldsymbol{x} \rangle + \boldsymbol{b} \text{ with } w \in \mathcal{X}, b \in \mathbb{R} \tag{2.1}$$

where $\langle .,. \rangle$ is the dot product in $\boldsymbol{\mathcal{X}}$, $\boldsymbol{w}$ is a vector of weights and $\boldsymbol{b}$ is analogous to a regression constant (known as intercept in a linear equation). In this case, $\boldsymbol{w}$ is reduced to achieve a flat function by minimizing the norm (i.e. $\|w\|^2 = \langle w , w \rangle$). This can be represented as a convex optimization problem:

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\|w\|^2 \\ &\text{subject to} && \begin{cases} y_i - \langle w , x_i \rangle - b \le \varepsilon \\ \langle w , x_i \rangle + b - y_i \le \varepsilon \end{cases} \end{aligned} \tag{2.2}$$

Slack variables, ξ_i and ξ^* [50], are introduced to handle the often unfeasible optimization problem constraints. Slack variables are integrated in the convex optimization problem as seen in eq. 2.3 [48]:

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\|w\|^2 + C \sum_{i=1}^{l} \left(\xi_i + \xi_i^*\right) \\ &\text{subject to} && \begin{cases} y_i - \langle w , x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w , x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i , \xi_i^* \ge 0 \end{cases} \end{aligned} \tag{2.3}$$

C is a constant that determines the trade-off between the allowance of deviations larger than $\boldsymbol{\varepsilon}$ and the weight vector $\boldsymbol{w}$. It is a user-defined parameter and is similar to a penalty parameter. A large **C** corresponds to reduced slack variable values. The required precision is represented by $\boldsymbol{\varepsilon}$ and is chosen by the user.

2.4.2.2. Formulation of the dual problem

The optimization problem is usually solved in its dual form. Dual formulation is further utilized to handle non-linear functions for SVM. Lagrange multipliers are used in the standard dualization method [51]. The dual or Lagrange function is formulated from the objective (primal) function and the corresponding constraints.

$$\begin{aligned} L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{l}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right) \\ - \sum_{i=1}^{l}\alpha_i\left(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle + b\right) \\ - \sum_{i=1}^{l}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b\right) \end{aligned} \tag{2.4}$$

where $\boldsymbol{L}$ is the lagrangian function with primal ($\mathbf{w}, \boldsymbol{\varepsilon}, \boldsymbol{\xi_i}, \boldsymbol{\xi_i}^*$) and dual variables ($\boldsymbol{\eta_i}, \boldsymbol{\eta_i}^*, \boldsymbol{\alpha_i}, \boldsymbol{\alpha_i}^*$). Dual variables are lagrange multipliers and have to fulfill the positivity constraints i.e. $\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0$.

According to its formulation, the Lagrangian has to be minimized w.r.t the primal variables and maximized w.r.t the dual variables [52]. At the optimal solution, the partial derivatives of the Lagrangian, $\boldsymbol{L}$, w.r.t the primal variables become zero. Resolving and substituting variables for the optimal solution, the dual optimization problem is reached [29]:

$$\text{maximize} \quad \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\varepsilon \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i + \alpha_i^*) \end{cases} \tag{2.5}$$

$$\text{subject to} \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C]$$

The values of $\boldsymbol{\alpha_i}, \boldsymbol{\alpha_i}^*$and the slack variables $\boldsymbol{\xi_i}$, $\boldsymbol{\xi_i}^*$are obtained from the solution of this optimization problem. Weight vector, $\boldsymbol{w}$, can be written in terms of $\boldsymbol{\alpha_i}$ and $\boldsymbol{\alpha_i}^*$ as follows:

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\, x_i \tag{2.6}$$

Substituting the value of $\boldsymbol{w}$ in the objective function, the following form is achieved:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\,\langle x_i\,, x\rangle + b \tag{2.7}$$

'b' can be computed using the Karush–Kuhn–Tucker (KKT) conditions [53]. The process of SVM-training is basically determining weight vector, **w**, and coefficient, b, so that the eq. 2.7 is achieved, which is then utilized during prediction by inputting vector, **x**.

2.4.2.3. Dealing with non-linear functions

The algorithm described up till now used a linear objective function. When dealing with non-linear objective functions, some adjustments and substitutions have to be made. Especially, the dot product, $\langle x_i\,, x\rangle$, in eq. 2.7 becomes unfeasible for a higher dimension input space. A popular technique of dealing with this complexity is based on mapping (Φ) the input space $\boldsymbol{\mathcal{X}}$ to some feature space $\mathcal{F}$, utilizing an appropriate kernel, represented as $\Phi : \mathcal{X} \rightarrow \mathcal{F}$.

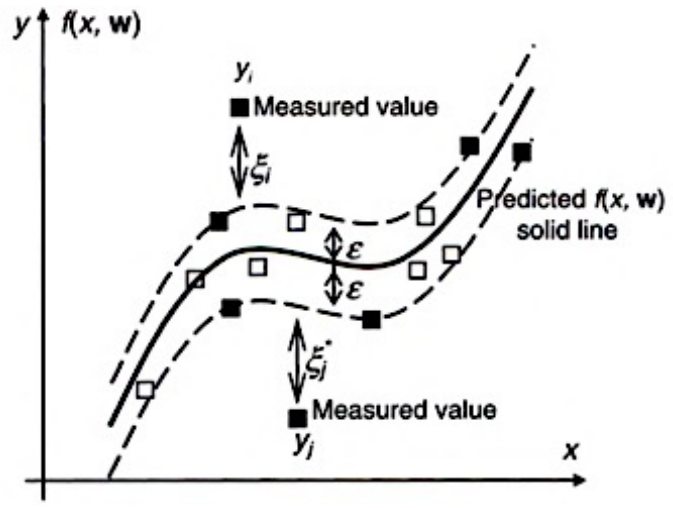

Figure 2.6. Diagramatic description of a non-linear (1-dimensional) support vector regression and the corresponding relevant variables ($y_i,\ x_i,\ w,\ \varepsilon,\ \xi_i$). Filled squares represent data points selected as support vectors; empty squares represent data points not categorized as support vectors, hence SVs only appear outside or on the tube boundary [54].

2.4.2.4. Application of the kernel method

In machine learning, kernel methods are used to apply linear model principles to non-linear functions. Kernel functions map the input (non-linearly connected) data to some feature space of higher dimension where they exhibit linear patterns. The dot product, $\langle \boldsymbol{x_i}, \boldsymbol{x} \rangle$, in eq. 2.7 is replaced by a kernel function, $\boldsymbol{k(x_i, x)}$ that implicitly computes the dot product in the feature space, $\mathcal{F}$.

Replacing the dot product with the kernel function in the optimization problem gives the following equations:

$$\text{maximize} \quad \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ -\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i + \alpha_i^*) \end{cases} \tag{2.8}$$

$$\text{subject to} \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C]$$

Similarly, the weight vector, $\boldsymbol{w}$, can be written as:

$$w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\, \Phi(x_i) \tag{2.9}$$

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\, k(x_i, x_j) + b \tag{2.10}$$

Mercer's condition constrains the kernel function selection [55]. Mercer's theorem deals with the dot product in a Hilbert space (a vector space with a dot product defined on it). Any kernel function, $\boldsymbol{k(x_i, x_j)}$, that satisfies the following condition can (theoretically) be appropriately used.

$$\int k(x_i, x_j)\, f(x_i) f(x_j) dx_i dx_j \geq 0 \tag{2.11}$$

for all functions $f(x_i)$, $f(x_j)$ satisfying:

$$\int f^2(x_i) dx_i \leq \infty \tag{2.12}$$

2.4.2.4.1. Radial Basis Function (RBF) kernel

In this study, the radial basis kernel function was used. It is described by the expression:

$$\begin{aligned} k(x_i, x_j) &= \langle \Phi(x_i) \,.\, \Phi(x_j) \rangle \\ &= \exp\left\{ -\gamma \left\| x_i - x_j \right\|^2 \right\} \end{aligned} \tag{2.13}$$

Here, $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are single instances of $\boldsymbol{x}$, $\| \,.\, \|$ represents the Euclidean norm, and γ is a positive parameter also known as the smoothing factor or bandwidth [52]. If γ is small, more weight will be allocated to points near $\boldsymbol{x_i}$, whereas a large γ may impart greater importance to far-off points. RBF is appropriate to use here as it nonlinearly maps samples into a higher dimensional space and can therefore handle the nonlinear relation between our output target and input attributes.

2.4.2. Diagrammatic Summary of SVM

Figure 2.7 adequately summarizes the SVM training and application process.

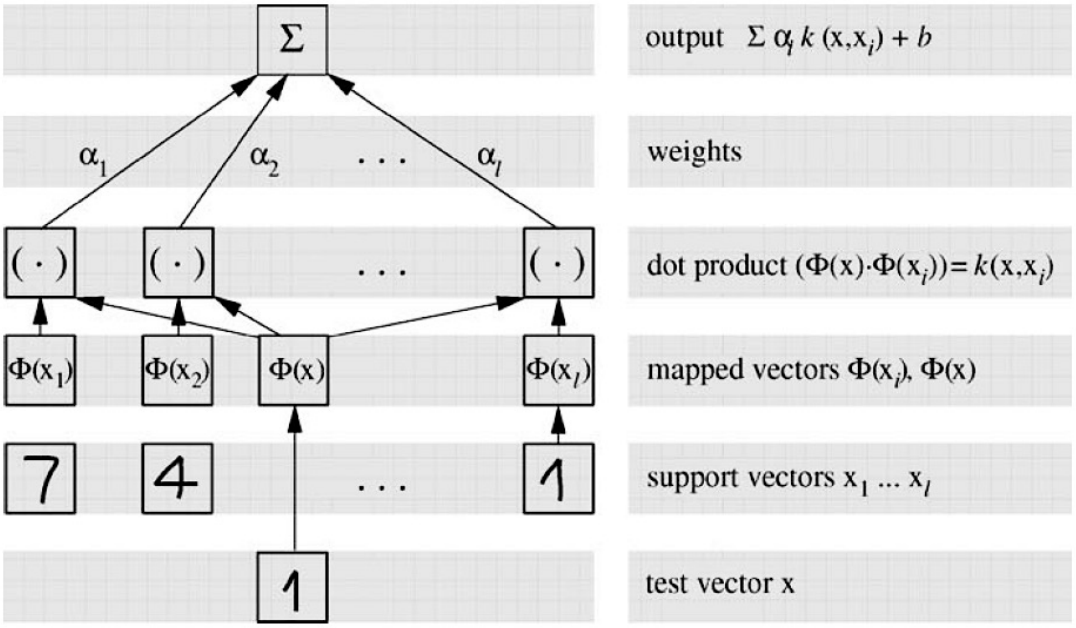

Figure 2.7. Description of a regression machine constructed by the support vector algorithm [29].

2.4.3. LIBSVM

LIBSVM [56] is an open source machine learning library. It was developed at the National Taiwan University and is written in C and C++. LIBSVM uses the sequential minimal optimization (SMO) algorithm [57] for solving the quadratic programming problem that arises during the training of support vector machines. This library package was utilized for the implemention of the SVM algorithm for this study.

Chapter 3: SVM Prediction Framework

3.1. Introduction

The first three study objectives of this thesis, as presented in Section 1.4, are:

- Analyzing sensitivity of SVM prediction to LIS states.
- Comparison of relative importance of LIS states for SVM prediction.
- Estimating SVM prediction sensitivity to SWE.

In this chapter, the methodology followed to attain these objectives is discussed. The SVM prediction and sensitivity analysis set-up is defined. It mainly consisted of three main phases:

- Phase-1: Noah MP land surface model simulation is accomplished within the LIS framework.
- Phase-2: SVM training and prediction using Noah MP output and AMSR-E brightness temperature data.
- Phase-3: Normalized sensitivity coefficient computation to analyze the sensitivity of SVM to LIS modeled states.

3.2. Phase-1: LIS Model Formulation and State Estimation

Noah-MP was used in the LIS framework to simulate the geophysical states of the study domain. These geophysical states served as the training input for SVM generation, and later as the prediction input for trained SVMs. **Henceforth, these states are referred to as LIS input states.** LIS 'open-loop' is the uncoupled stand-alone model (Noah-MP) simulation. The open-loop accuracy is dependent on model physics and supplementary data such as boundary conditions, initial conditions, and parameters that illustrate domain characteristics.

Boundary conditions used for the LIS open-loop simulation were obtained from the Modern-Era Retrospective analysis for Research and Applications - Version 2 (MERRA2) forcings. MERRA2 is the recent atmospheric reanalysis produced by NASA's Global Modeling and Assimilation Office (GMAO) [11]. The reanalysis

period starts from January 1980 and continues as a near real-time analysis. It has a native resolution of 0.5° lat x 0.625° lon x 72 model layers. It is based on the assimilation of ground and satellite data into the Goddard Earth Observing System – version 5 (GEOS-5) model to produce gridded datasets. MERRA2 is an improved version of the previously widely used MERRA datasets and incorporates updated methods in the GEOS model, new atmospheric variables, and the latest satellite data. MERRA2 boundary condition parameters included:

- Near surface air temperature
- Near surface specific humidity
- Incident shortwave radiation
- Incident longwave radiation
- Eastward wind
- Northward wind
- Surface pressure
- Rainfall rate
- Convective rainfall rate

Some other datasets used to characterize the model domain features included land cover maps from Moderate Resolution Imaging Spectroradiometer (MODIS), soil texture characterization from International Soil Reference and Information Centre (ISRIC), and topographic information such as slope, aspect, and elevation from Shuttle Radar Topography Mission (SRTM). The initial conditions were adjusted using a spin-up time of 22 years, starting in January 1980 and ending in December 2001. Model simulation period spanned from year 2002 to 2016. The study period extends from Sep-2002 to Sep-2011 (9 years). This coincides with the AMSR-E available CETB data period.

3.2.1. Description of the Study Domain

The study domain comprises the Indus Basin. The Indus Basin spans over parts of four countries: Pakistan, India, China, and Afghanistan. This includes the mountain ranges of Hindu Kush, Karakorum, and Himalaya. The total area is estimated to be 1.1 million km^2. Most of this area lies within Pakistan and India. The Indus river originates in the Tibetan Plateau and has 15 tributaries that contribute to it downstream.

The Indus basin has highly varying topography, with very high elevations (>7000 m) in the north, medium height mountains in the west (>2000m), and flat plateaus and plains comprising the rest of the basin area, as seen in Figure 3.1. This topographic setting renders the modeling of the entire basin a relatively complex task. The high mountains in the north provide greatest difficulty as little prior information is available regarding the prevalent glaciers and ephemeral snow patterns. Also, since the grid resolution of current land surface models is on the order of kilometers, the highly varying topography (and relevant land surface interaction) is not always adequately represented.

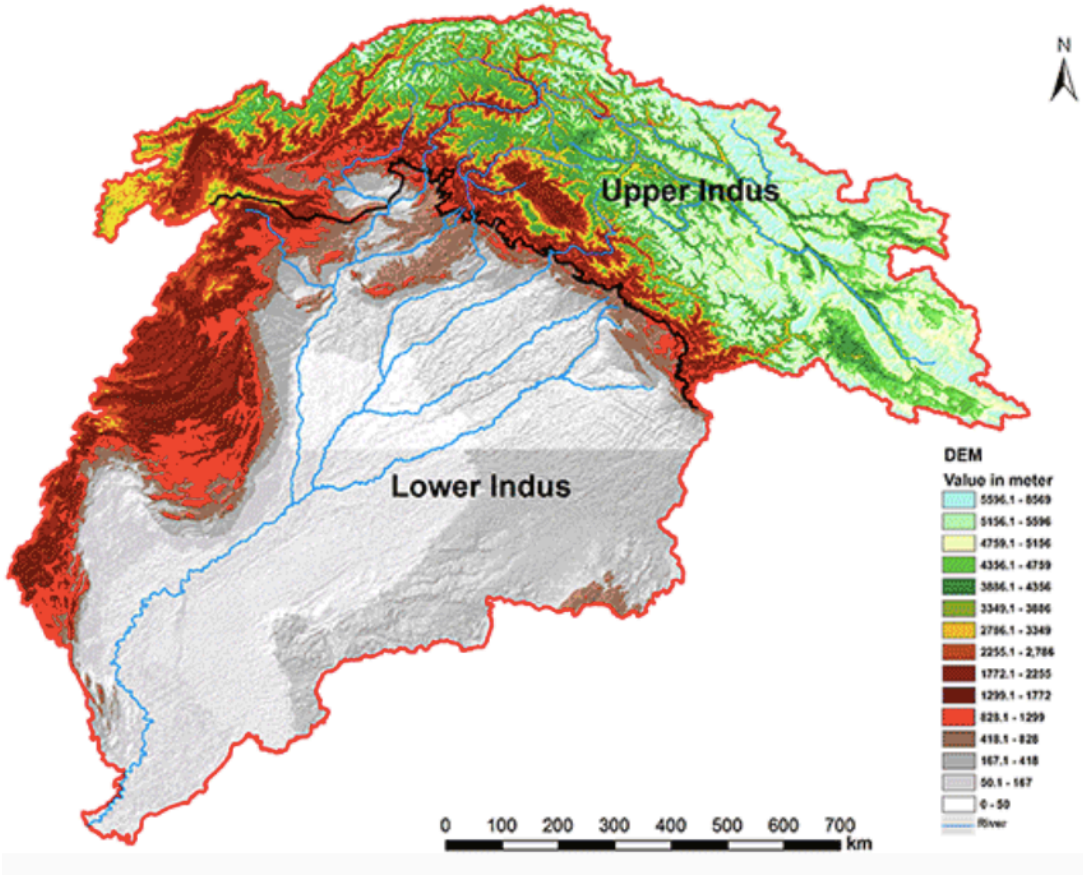

Figure 3.1. DEM of the Indus Basin showing variation in topography [58].

The largest lake in the area (Lake Saiful-Muluk) has a surface area of approximately 2.75 km^2. Considering the grid resolution of the AMSR-E brightness temperature data (625 km^2), the lake effect is neglected from our analysis. Heavy

snowfall primarily occurs in the northern Karakoram and Himalayan ranges and in the western Hindukush range. Moderate seasonal snow is also observed in the vicinities of Kashmir (Margalla and Murree Hills).

Figure 3.2 depicts the land cover variation within the Indus Basin. From permanent ice in the northern areas to cultivated plains in the south, the Indus basin has significant vegetation heterogeneity. Only snow-covered (seasonally and permanently) areas are considered for the purposes of this study. Figure 3.2 is presented here only to convey a general idea of the vegetation types in the study domain.

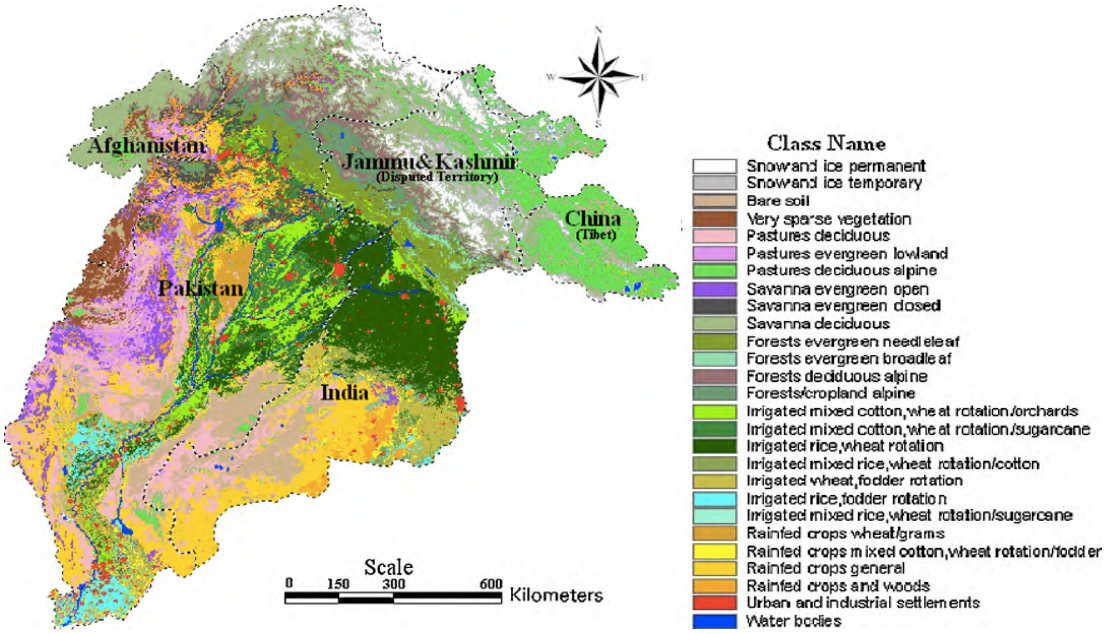

Figure 3.2. Land cover map of the Indus Basin developed from 36 SPOT-Vegetation based NDVI values for 2007 [59].

3.2.2. LIS Input States used in SVM Training and Prediction

LIS modeled states were inherently on an equidistant cylindrical grid (0.01^{O} x 0.01^{O}). To maintain consistency between LIS and AMSR-E data, these states were re-gridded to the EASE grid (pixel size = 25km x 25km) for further utilization.

Ten LIS modeled states were used for SVM training and prediction (Table. 3.1). Selection criteria for the LIS input states included previous similar studies as well as

their individual physical effect on brightness temperature. Input states that were physically expected to significantly influence brightness temperature were chosen.

Table 3.1. LIS modeled geophysical states used as input for SVM training and prediction

State	Unit
Snow water equivalent	m
Snow liquid water content	m
Top-layer snow temperature	K
Top-layer soil temperature	K
Air temperature (near surface)	K
Bottom-layer snow temperature	K
Snow density	Kg/m^3
Soil moisture (near surface)	m^3/m^3
Vegetation temperature	K
Leaf area index	dimensionless

3.2.2.1 Unit conversion of LIS input states

The LIS input states vary considerably in their units. Some quantities have dimensions of length, others temperature, while some are dimensionless. To introduce some consistency among the LIS input state magnitudes, all quantities were converted to such respective units that lie within the same order of magnitude. For example, temperatures varied between 170 to 350 Kelvin throughout, whereas SWE fluctuated between 0 and 7 meters. To place them in the same order of magnitude, temperature values were converted to 10^{-2} K.

Since SVM does not make any provision for the different units of the input states, thus to decrease chances of dominance of some states over others due to magnitude only, this unit conversion is advisable [60]

Table 3.2. List of unit conversion factors used for scaling LIS input states to constrict them on the same order of magnitude.

State	Unit Conversion Factor
Snow water equivalent	10
Snow liquid water content	1
Top-layer snow temperature	0.01
Top-layer soil temperature	0.01
Air temperature (near surface)	0.01
Bottom-layer snow temperature	0.01
Snow density	0.01
Soil moisture (near surface)	10
Vegetation temperature	0.01
Leaf area index	1

3.3. Phase-2: SVM Framework

SVM framework is divided into training and prediction sub-phases, described in Section 3.2.1 and Section 3.2.2.

3.3.1. SVM Training

SVM training was accomplished through the utilization of the LIBSVM library, provided by the National Taiwan University. During the SVM training process, the support vectors and their respective weights were selected through the process described in Chapter 2. AMSR-E T_b data from Sep-2002 to Sep-2011 (total 9 years) is used as the training target. The SVM training process described by [7], [61], and [6] was followed for this study.

A separate and independent SVM was generated for each fortnight in the study period. AMSR-E T_b target data used for each fortnight SVM training included data from 2-weeks before and 2-weeks after the pertinent 2-week data. Thus, each SVM training data consisted of a 6-week period. This 2-week overlap for each fortnightly

SVM was intended to maintain continuity and to include seasonal effects in snow dynamics. To limit the possibility of errors introduced by wet snow, measurements gathered during the nighttime AMSR-E overpass were used during training.

Parameter selection (described in section 3.3.1.1) was completed before the actual SVM training. While training for a certain fortnight, f, in a certain year, yr, the AMSR-E data from all the years except year 'yr' was used for training. Thus, each fortnight was trained using the relevant fortnight data from the remaining 8 years. The total number of AMSR-E data points for each fortnightly training was 336 (8 years x 6 weeks x 7 days = 336 daily values). AMSR-E measurements, for the relevant year, that were excluded from training were then later used for validation purposes.

3.3.1.1. SVM parameter selection (C, ε, γ):

Parameter selection is an important step in model formulation. This is especially so in SVM, considering the non-linearity and complex model dynamics. As discussed in Chapter 2, three parameters need to be set manually for each SVM; C, ε, and γ. 'C' is the penalty parameter and is defined in this study as the range of the training targets (z). This selection is based on the rationale provided by [62].

$$C = max\{\boldsymbol{z}\} - min\{\boldsymbol{z}\}$$

Selection of 'ε' and 'γ' was done using a two-phase SVM training method. This involved formation of two subsets, a and b, of the total 9-year training data. Subset-a data was used to train a test SVM. Subset-a data trained SVM was then used to predict the subset-b data and the corresponding mean squared error (MSE) was computed. This process was repeated for a range of 'ε' and 'γ' values. The same procedure was employed for subset-b and MSE values (for various combinations of 'ε' and 'γ') calculated by predicting subset-a using the subset-b trained SVM were obtained.

All of the MSE values were compared and the 'ε' and 'γ' pair that yielded the least MSE magnitude was selected for use during the second phase of SVM training. The second phase used the selected parameter values and training was completed using the entire 9-year AMSR-E data as described above.

3.3.2. ΔTb Prediction using SVM

SVM input consisted of the 10 geophysical LIS input states and the output consisted of four T_b frequency differences and polarization combinations (Table 3.3). Since our analysis is primarily focused on snow covered areas, predictions were carried out for those locations only where snow water equivalent (SWE) was greater than 1 cm.

Table 3.3. List of variables predicted by SVM using LIS input states

Training target / SVM output	Symbol	Unit
Difference in brightness temperature measured at 10.65GHz (vertical polarization) and 36.5GHz (vertical polarization) frequencies	ΔT_b (10.65V – 36.5V)	Kelvin
Difference in brightness temperature measured at 10.65GHz (horizontal polarization) and 36.5GHz (horizontal polarization) frequencies	ΔT_b (10.65H – 36.5H)	Kelvin
Difference in brightness temperature measured at 18.7GHz (vertical polarization) and 36.5GHz (vertical polarization) frequencies	ΔT_b (18.7V – 36.5V)	Kelvin
Difference in brightness temperature measured at 18.7GHz (horizontal polarization) and 36.5GHz (horizontal polarization) frequencies	ΔT_b (18.7H – 36.5H)	Kelvin

3.3.2.1 SVM prediction validation

SVM output validation was accomplished using the one year data that was omitted during training for each fortnightly trained SVM. Figures 3.3 and 3.4 summarize the over-all bias and RMSE of SVM predictions with respect to the AMSR-E satellite data not used during training from year 2002 to 2011. Results for ΔT_b(18.7V – 36.5V) are presented as this frequency difference and polarization is most relevant to SWE remote sensing [6].

Figure 3.3 shows the average bias observed at each location within the prediction area. Daily bias ($\Delta T_{b,\,SVM}$ - $\Delta T_{b,\,AMSR\text{-}E}$) values calculated for years 2002-2011, for each location, are averaged to obtain the values presented in Figure 3.3. Most of the values lie between -0.5K and 0.5K. Two prominent zones experienced high bias. One location lies near 33.5^{O}N and 68^{O}E. This cluster of pixels displays negative bias values and highlights the under-prediction of SVM at this location. In contrast to this, pixels around 37.5^{O}N, 74^{O}E display positive bias values and present the over-estimating tendencies of SVM in that cluster of pixels.

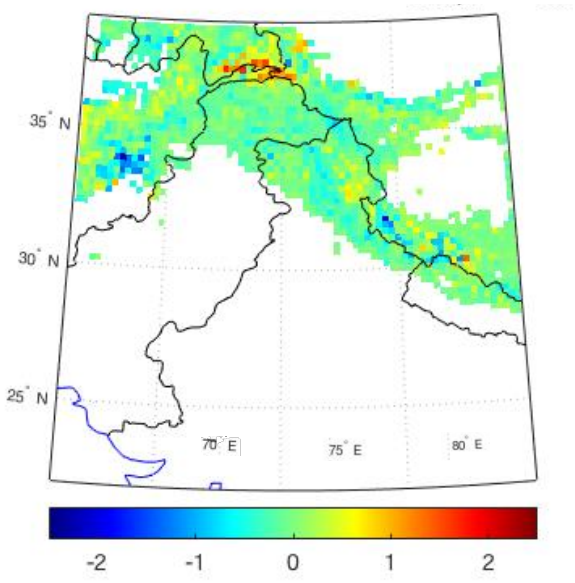

Figure 3.3. Average bias (SVM - AMSR-E) for ΔT_b (18.7V – 36.5V), in units of Kelvins, observed at each location where SWE > 1cm in the Indus Basin for years 2002-2011. The solid black line represents country boundaries and the solid blue line depicts the coastline.

Yearly RMSE values calculated for each year from 2002 to 2011, for each location, are averaged to obtain the values presented in Figure 3.4. Figure 3.4 demonstrates that the two primarily biased zones identified above show high RMSE values. These locations have low prediction accuracy and are thus expected to have irrational sensitivity analysis results. The diagonally elongated area stretching from 35^{O}N and 75^{O} E to 30^{O}N and 83^{O}E is collocated with glaciers. Brightness temperatures predicted over glaciers are expected to have errors since the list of LIS input variables used for SVM prediction does not include any ice sheet/glacier related variables.

Further details regarding SVM prediction validation and possible causes of bias and RMSE are provided by Forman et al. in [4].

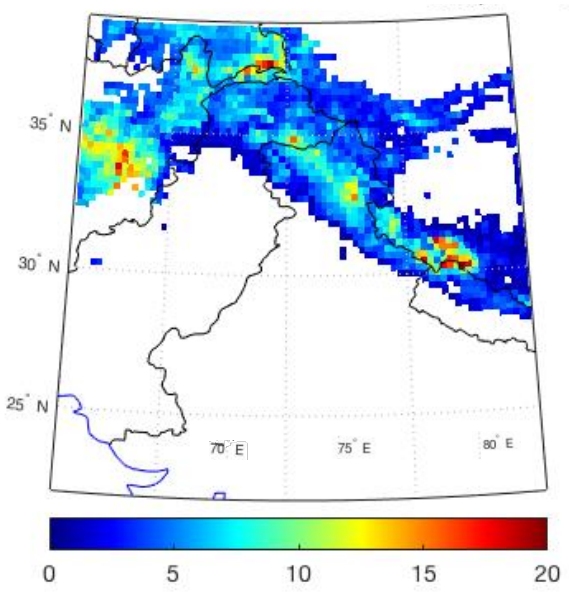

Figure 3.4. Average RMSE of SVM ΔT_b (18.7V – 36.5V) prediction, in units of Kelvin, observed at each location where SWE > 1cm in the Indus Basin for years 2002-2011. The solid black line represents country boundaries and the solid blue line depicts the coastline.

3.4. Phase -3: Sensitivity Analysis Metric Formulation

An analysis of comparative importance of predictors in a linear model is usually achieved using the standard value of t, correlation (R), and S (=t*R) value. These methods are not applicable in this case since SVM is a non-linear, highly complex model. Therefore, to analyze the comparative predictor importance a sensitivity analysis was performed. This was intended to contribute towards the selection of the most important (sensitive) predictors for the formulation of the final SVM that is intended for use in a SWE data assimilation framework as a measurement operator.

A sensitivity analysis is targeted towards examination of the sensitivity of model output to a change in each predictor. The three main types of sensitivity used for analyzing predictor importance are; absolute sensitivity, relative sensitivity, and deviation sensitivity [63].

- Absolute sensitivity, S_a: change in one factor, y, with respect to change in another factor, x. It has units of 'y/x' and thus cannot be used for comparing sensitivity of different 'x' values.

$$S_a = \frac{\partial y}{\partial x} \tag{3.1}$$

- Relative sensitivity, S_r: percent change in one factor, y, due to some prescribed percent change in another factor, x. It is unit-less and can, thus, be used for comparing the importance of various predictors.

$$S_r = \left(\frac{\partial y}{\partial x}\right) * \left(\frac{x}{y}\right) \tag{3.2}$$

- Deviation sensitivity, S_d: incremental change in one factor, y, due to incremental change in another factor, x. It has units of Δy and can thus be utilized for comparative study of various predictor sensitivities. It is usually employed in error analysis.

$$S_d = \frac{\partial y}{\partial x} * \Delta x \tag{3.3}$$

For our specific case, Normalized Sensitivity Coefficients (NSC) are employed to compare the relative sensitivity of each LIS input state. NSCs will be used as deciding factors for predictor importance in SVM model prediction.

3.4.1 Normalized Sensitivity Coefficient

Normalized sensitivity coefficients [64] are computed to assess the sensitivity of a well-trained SVM to each LIS modeled state variable.

$$NSC_{i,j} = \frac{\partial M_j}{\partial P_i} * \frac{{P_i}^o}{{M_j}^o} \approx \frac{{M_j}^i - {M_j}^o}{\Delta P_i} * \frac{{P_i}^o}{{M_j}^o} \tag{3.4}$$

where i = state index, j = output metric index, ${M_j}^i$ = perturbed metric value, ${M_j}^o$ = initial metric value, ${P_i}^o$ = initial state value, ΔP_i = amount of perturbation.

Each LIS input state is perturbed (one at a time), while maintaining the original value of all the other states. The observed change in output relative to the induced perturbation is a measure of the effect a change in that LIS input state will have on the

SVM predicted output. Since one LIS input state is perturbed at a time, while the other states remain constant, independence between the states is assumed; however, the sensitivity is dependent on the values of all states. The NSC value is representative of the effect of the perturbed state only on SVM output, while assuming the other states are not affected by that perturbation at all.

A normalized sensitivity coefficient is representative of local sensitivity as compared to global sensitivity. Local sensitivity analysis is useful in studying the role of parameters or input variables in the model [65]. This technique has been effectively used in various studies [66] [67].

3.4.1.1 Suitability of NSCs

NSCs are suitable for analyzing the relationship between SVM predicted output and LIS input states due to the following reasons:

- SVM is based on statistical learning theory and thus has a statistical rather than a physical model basis. Considering this statistical origin, it is uncertain whether the results obtained by SVM prediction are physically based or not. The NSCs help in exploring this avenue.
- NSCs are dimensionless and, thus, it is rational to utilize them in inter-predictor sensitivity comparisons.
- NSCs have a magnitude as well as a sign, and thus relate not only the importance of the predictor, but also the direction of its relationship (direct vs. inverse) to the model output.
- By perturbing each LIS input state (predictor variable) by the same amount and analyzing the resulting effect on the output, a measure of relative importance of the predictors is attained.

3.4.1.2 NSC Formulation

The amount of perturbation (ΔP_i in eq. 3.4) is selected manually by the users. The perturbation size should be large enough to detect a change in the output, yet small enough that the model behaves approximately linearly. For our case, a range of

perturbations was tested and the corresponding relative change was analyzed. Figure 3.5 presents an example of the variation in relative change of ΔT_b (18.7V − 36.5V) as perturbation in SWE is increased from -20% to + 20%. A relative change is formulated as:

$$RC = \frac{PO - NO}{NO} * 100 \tag{3.5}$$

where RC = Relative change, PO = Perturbed Output, NO = Nominal Output.

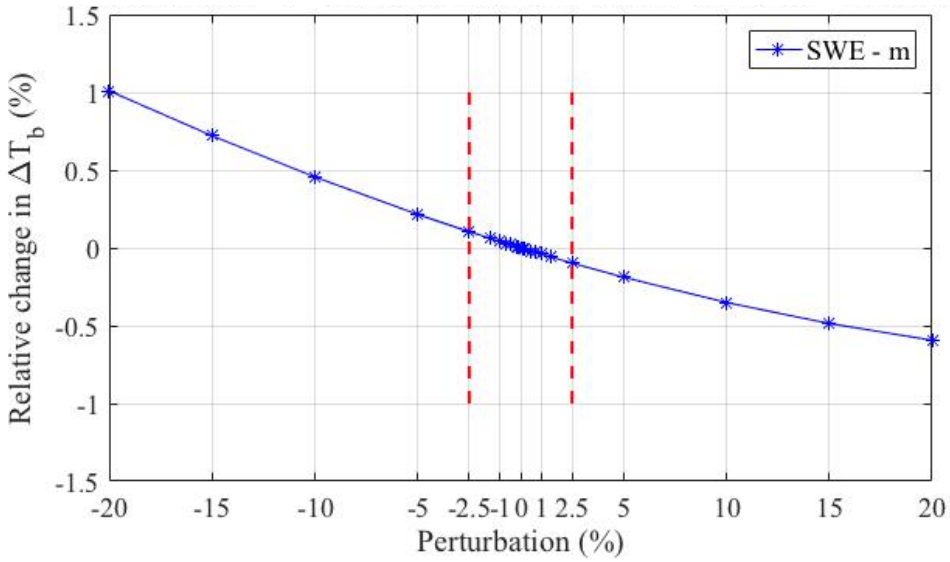

Figure 3.5. Variation in relative change observed in ΔT_b (18.7V − 36.5V) prediction as LIS modeled SWE input is perturbed for a point location in the Indus Basin (35.73^{O}N,76.28^{O}E) for one day (Jan 1, 2004). Red–dashed line shows the perturbation bounds that were ultimately chosen.

Apart from SWE, relative change versus perturbation plots for all the other 9 states were also generated (Figure 3.6). After studying a range of locations and days, a perturbation value of +/-2.5% (total 5%) was selected. The perturbation value is selected according to the model output. It is desired that the model behaves linearly within the perturbation range. Figure 3.5 and Figure 3.6 show that the SVM behaves approximately linearly with respect to change in any of the LIS input states within the perturbation bounds of +/-2.5%.

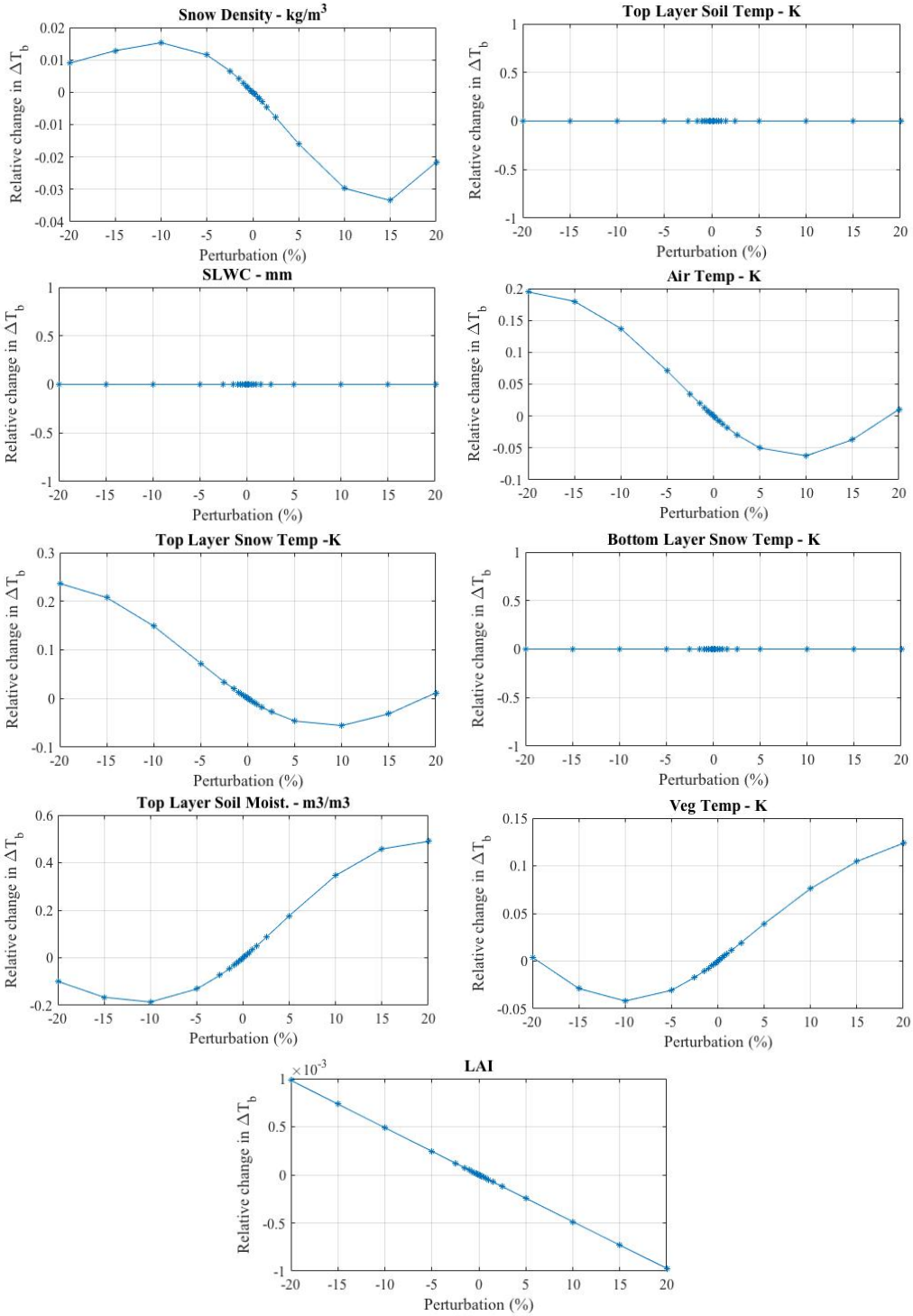

Figure 3.6. Variation in percent relative change observed in ΔT_b (18.7V − 36.5V) prediction as LIS modeled input states are perturbed for a point location in the Indus Basin (Lat: 35.73^{O}N, Lon: 76.28^{O}E) for one day (Jan 1, 2004).

The slope of the curves in Figure 3.6 represent the NSC value. Therefore, steeper slopes reflect higher calculated NSC magnitudes. The corresponding sign is determined by the direction of the slope, i.e., positive for an increasing slope.

Each LIS input state was perturbed by +/- 2.5% for each fortnightly SVM, and the corresponding NSC was calculated. A separate NSC was calculated for each of the 10 individually perturbed LIS input states, for each of the four SVM predicted ΔT_b outputs (multi-frequency and multi-polarization), for each day in the study period, for each pixel in the domain where LIS modeled SWE was > 1cm.

NSCs were calculated at the nominal state value, that is a positive and a negative (equal magnitude) perturbation was applied and a centered difference (eq. 3.4) was then calculated. This was done to remove the possibility of a biased NSC in case the SVM model behaved non-linearly within the perturbation limits for any day or location.

3.4.2 Variability in the Effect of Perturbation on SVM Prediction

Before delving into the NSC analysis, a brief overview of the effect of LIS input state perturbation on SVM prediction is presented here. The objective here is to understand the variability in the effect of LIS input state perturbation on SVM prediction observed throughout the domain. This variability is ultimately linked to variation in accuracy of SVM prediction in the domain, as will be explained in the coming sections. The direct physical relationship between SWE and ΔT_b (18.7V – 36.5V) is exploited to explain this effect.

In Figure 3.7, most of the areas show zero change, meaning SVM ΔT_b (18.7V – 36.5V) prediction is not very sensitive to LIS modeled SWE at these locations. The negative and positive perturbation propagates opposite effects within the domain. Areas that exhibit a negative difference value (blue color) for the negative perturbation show a positive difference value (red color) for the positive perturbation. Of these, the locations near lat: 34.5^{O}N, lon: 66^{O}E and the diagonal cluster of pixels from lat: 32.5^{O}N, lon: 77.5^{O}E to ~lat: 35^{O}N, lon: 76^{O}E show noticeable physically irrational results. This indicates physical irrationality because according to the principles of remote sensing of snow, as discussed in Chapter 1, SWE and ΔT_b (18.7V – 36.5V) have a directly proportional relation. If only SWE is decreased (negative perturbation) while

the other states are retained at their original values, then the resultant ΔT_b(18.7V - 36.5V) is expected to decrease as well and hence the difference between the nominal ($\Delta T_{b,nom}$) and negatively perturbed ($\Delta T_{b,neg}$) output should be a positive value. The locations identified here exhibit the opposite of this principle, i.e., decreasing SWE (negative perturbation) is resulting in an increased $\Delta T_{b,neg}$ and subsequently resulting in a positive value for difference between the nominal and negatively perturbed output ($\Delta T_{b,nom}$ - $\Delta T_{b,neg}$), indicating an inversely proportional relation between SWE and ΔT_b(18.7V - 36.5V).

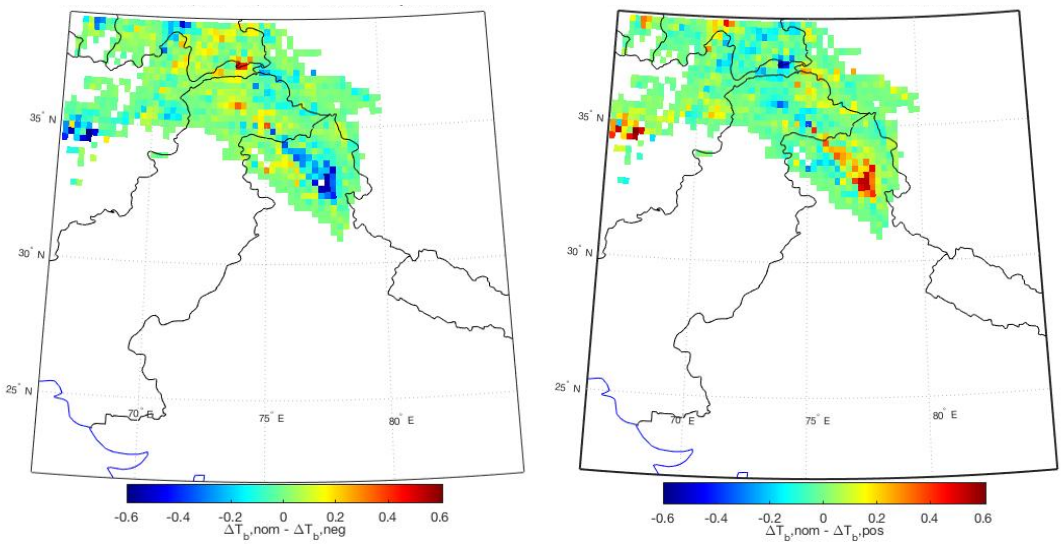

Figure 3.7. Maps of Indus Basin showing the difference between nominal and -2.5% perturbed SWE SVM prediction of ΔT_b (18.7V – 36.5V) (left), and nominal and +2.5% perturbed SWE SVM prediction of ΔT_b (18.7V – 36.5V) (right) for Jan 1, 2004. The solid black lines represent country boundaries and the solid blue line depicts the coastline.

Histograms (see Figure 3.8 and Figure 3.9) of all the differences between the nominal and perturbed SVM output values were generated from Figure 3.7. In Figure 3.8, for negative perturbation, physical rationality suggests all difference between the nominal and negatively perturbed SVM output values ($\Delta T_{b,nom}$ - $\Delta T_{b,neg}$) should be greater than 0. The results show that although all the values are not > 0, most of them are. In Figure 3.9 the case should physically be opposite to this and all differences between the nominal and positively perturbed SVM output values ($\Delta T_{b,nom}$ - $\Delta T_{b,pos}$)

are expected to be < 0. The results here are similar to Figure 3.8 and it is observed that although all the values are not < 0, most of them are.

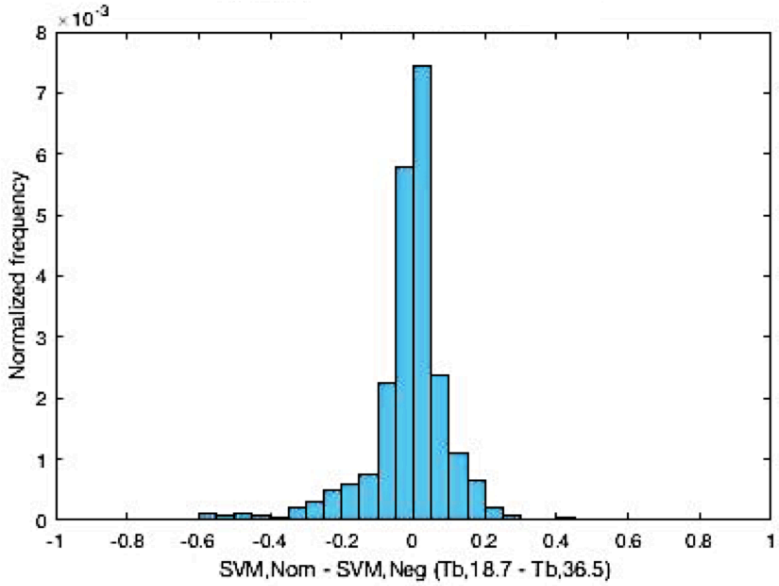

Figure 3.8. Histogram of difference between the nominal and -2.5% perturbed SWE SVM predicted ΔT_b (18.7V – 36.5V) values for areas where LIS modeled SWE > 1cm in the Indus Basin for Jan 1, 2004.

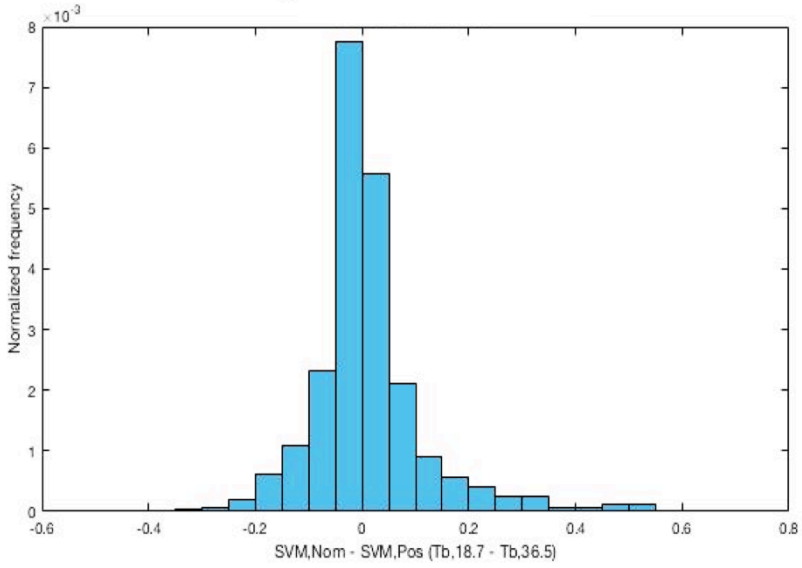

Figure 3.9. Histogram of difference between the nominal and +2.5% perturbed SWE SVM predicted ΔT_b (18.7V – 36.5V) values for areas where LIS modeled SWE > 1cm in the Indus Basin for Jan 1, 2004.

It is observed that the locations showing physically irrational perturbation effects in Figure 3.7 have high RMSE values (see Figure 3.4) and thus poor predicting capability. Apart from this, another reason is the inter-correlation between the LIS input states. When a single state is perturbed e.g. air temperature, physically rationalizing the soil temperature and the vegetation temperature is expected to undergo some change as well. During individual state perturbation, we ignore this physical phenomenon. The price of ignoring this inter-correlation between the LIS input states is paid in the form of misleading sensitivity analysis results (detailed in Chapter 4).

To gain some knowledge of the extent of inter-correlation (cross-correlation) present, the inter-correlation matrix of all LIS input states was computed (Table 3.4) using LIS model output for the whole domain for year 2004. Some very high correlation values are visible, highlighted in red. For example, air temperature and soil temperature have a correlation of 0.981. Most of the high correlation values are related to temperature variables. This is physically expected as well. As far as SWE is concerned, the highest correlation value is for snow density (=0.568). Since snow density is calculated from SWE and snow depth estimates, this relationship between these variables is expected.

Abbreviations used in Table 3.4:

SWE = Snow water equivalent, SLWC = Snow liquid water content,
ST = Soil temperature, TLST = Top-layer snow temperature, AT = Air temperature,
BLST = Bottom-layer snow temperature, SD = Snow density, SM = Soil moisture
VT = Vegetation temperature, LAI = Leaf area index

Table. 3.4. Inter-correlation matrix of 10 LIS input states, calculated using (daily) LIS modeled states over the Indus Basin for year 2004.

	SWE m	SLWC mm	TLST K	ST K	AT K	BLST K	SD kg/m^3	SM m^3/m^3	VT K	LAI
SWE m	1.000	0.202	-0.051	-0.041	-0.012	0.040	0.568	-0.051	-0.012	-0.042
SLWC mm	0.202	1.000	0.275	0.245	0.262	0.271	0.240	0.248	0.263	-0.025
TLST K	-0.051	0.275	1.000	0.723	0.717	0.818	0.068	0.460	0.765	0.183
ST K	-0.041	0.245	0.723	1.000	0.981	0.869	0.136	-0.389	0.976	0.313
AT K	-0.012	0.262	0.717	0.981	1.000	0.685	0.240	-0.403	0.992	0.335
BLST K	0.040	0.271	0.818	0.869	0.685	1.000	0.118	0.544	0.665	0.167
SD kg/m^3	0.568	0.240	0.068	0.136	0.240	0.118	1.000	0.174	0.243	0.285
SM. m^3/m^3	-0.051	0.248	0.460	-0.389	-0.403	0.544	0.174	1.000	-0.427	0.235
VT K	-0.012	0.263	0.765	0.976	0.992	0.665	0.243	-0.427	1.000	0.321
LAI	-0.042	-0.025	0.183	0.313	0.335	0.167	0.285	0.235	0.321	1.000

Supplementary to the LIS input states inter-correlation matrix, since the primary focus is on SWE, a graphical analysis of SWE vs. all the other (9) LIS input states (Fig 3.10) was performed. Based on these plots, all the other LIS input states, except snow density, do not seem to have any significant linear relationship with LIS SWE.

The graphs in Figure 3.10 do not give any indication of what the location specific relationship between the states is. To garner more information about the variation in the correlation between LIS SWE and other LIS input states throughout the domain, correlation maps were generated using LIS modeled data for the Indus Basin for year 2004, Figure 3.11. Correlation maps are developed using data from those pixels only where both variables had a non-zero value. The number of data points for each variable for each pixel is greater than 30 (8.22% of 365 days).

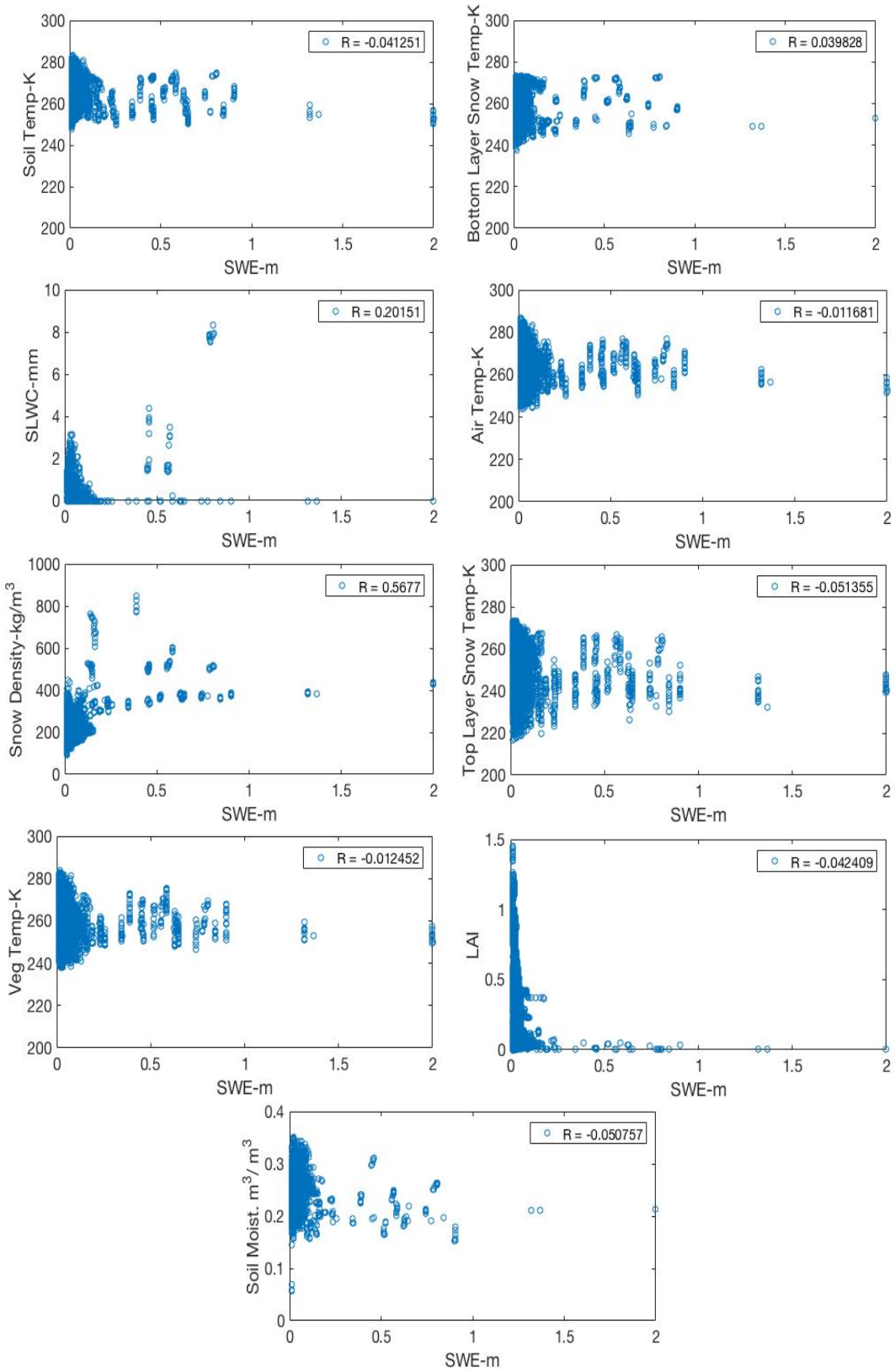

Figure 3.10. Graphical analysis of LIS SWE estimate vs. all other LIS input states used in SVM ΔT_b prediction over the Indus Basin for year 2004.

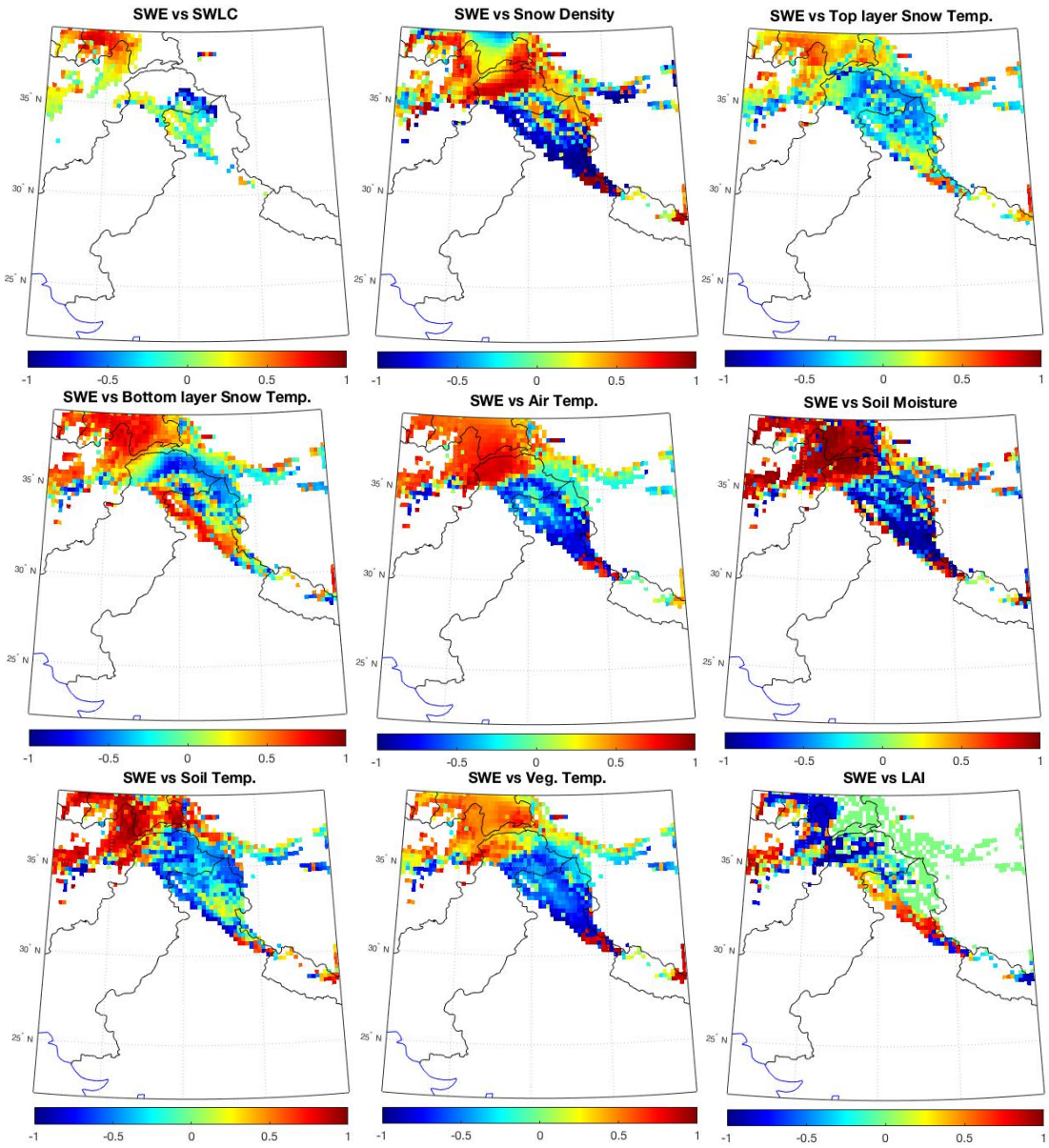

Figure 3.11. Correlation maps between LIS modeled SWE and all other (9) LIS input states for the Indus Basin for year 2004. The solid black line represents country boundaries and the solid blue line depicts the coastline.

The correlation maps (Figure 3.11) show a much different view of the relations between LIS SWE vs. other LIS input states compared to the data plots in Figure 3.10. There are high correlation values (positive as well as negative) evident in each map at different locations. These locations with high correlation values guide in understanding the sensitivity analysis results in Chapter 4.

Chapter 4: Sensitivity Analysis of SVM Prediction

In this chapter, the sensitivity analysis results are discussed. The relative sensitivity of SVM predicted brightness temperature (spectral difference) to the LIS modeled input states is studied spatially as well as temporally, using normalized sensitivity coefficients.

The sensitivity analysis results described in this chapter focus on ΔT_b (18.7V-36.5V), see Table 3.3, as this brightness temperature spectral difference is most significant for remote sensing of snow parameters (such as SWE) and thus very relevant to snow-covered areas of importance within the domain. Vertical polarization is selected for this analysis since horizontally polarized microwave radiation is comparatively more affected by ice-layers present within the snow pack than vertically polarized microwave radiation. The 10 LIS input states used for SVM prediction do not include any ice-layer characterizing state. Therefore, in order to achieve a better understanding of the effect of LIS input states on correctly trained SVMs, vertically polarized ΔT_b are focused upon.

4.1. Spatial Analysis of Normalized Sensitivity Coefficients

Considering that the primary focus was on snow covered areas, two main seasons for our sensitivity analysis i.e. snow ablation (April, May, and June) and snow accumulation period (December, January, and February) were used. These months were selected in accordance with the climatology of the Indus basin. April marks the end of the spring season and the advent of the summer season (in general) and witnesses the start of snow melt in most areas. This snow melt period continues throughout the summer season. The main snow accumulation period falls within the months of December, January and February for most places in the Indus Basin. The snow accumulation and ablation periods were restricted to the three most important months to lessen excessive temporal averaging of the NSCs. This exercise helped identify the extent of the effect change in snow season had on the sensitivity values.

The objective of this spatial analysis was to find location specific relative sensitivities. Areas that are collocated with glaciers are expected to give irrational sensitivity values (due to the effect of glaciers on T_b measurement, as discussed in Chapter 2). Figure 4.1 shows the locations identified by the Global Land Ice Measurements from Space (GLIMS) database where glaciers/ice masses are present.

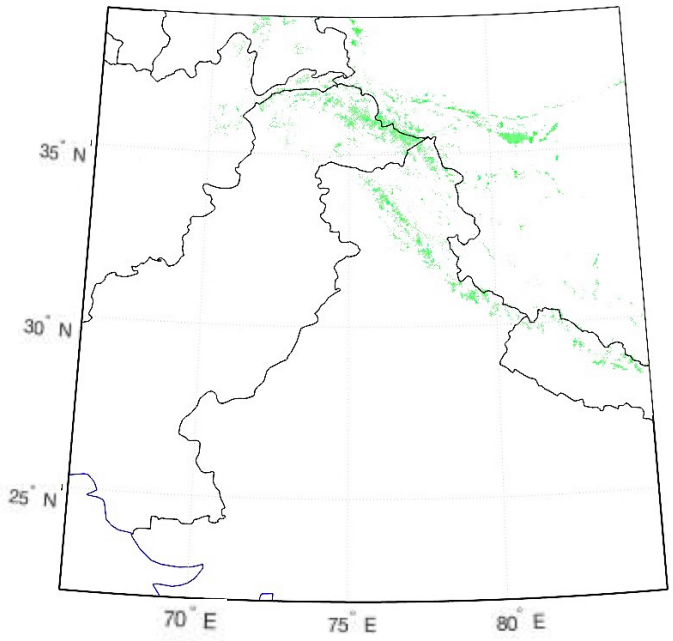

Figure 4.1. Map of locations, as identified by Global Land Ice Measurements from Space [68], where glaciers or ice masses are present within the Indus Basin.

NSCs were calculated only for those areas where SWE was greater than 1cm. These areas were identified as 'snow-covered'. The snow ablation and accumulation period NSC maps (Figure 4.2 and Figure 4.3) for the Indus Basin for year 2004 indicate the relationship between each LIS input state and the SVM predicted ΔT_b (18.7V-36.5V). Maps of soil moisture and the bottom-layer snow temperature are not included in Figure 4.2 and Figure 4.3. Soil moisture values modeled by LIS (Noah MP) under frozen soil conditions are not very accurate and thus have been removed from the sensitivity analyses. For some of the locations, the snow pack does not grow deep enough to have multiple layers, therefore a bottom-layer snow temperature does not exist in those cases. For locations that have values of snow temperature of the bottom-layer of snow pack, the NSCs are approximately zero.

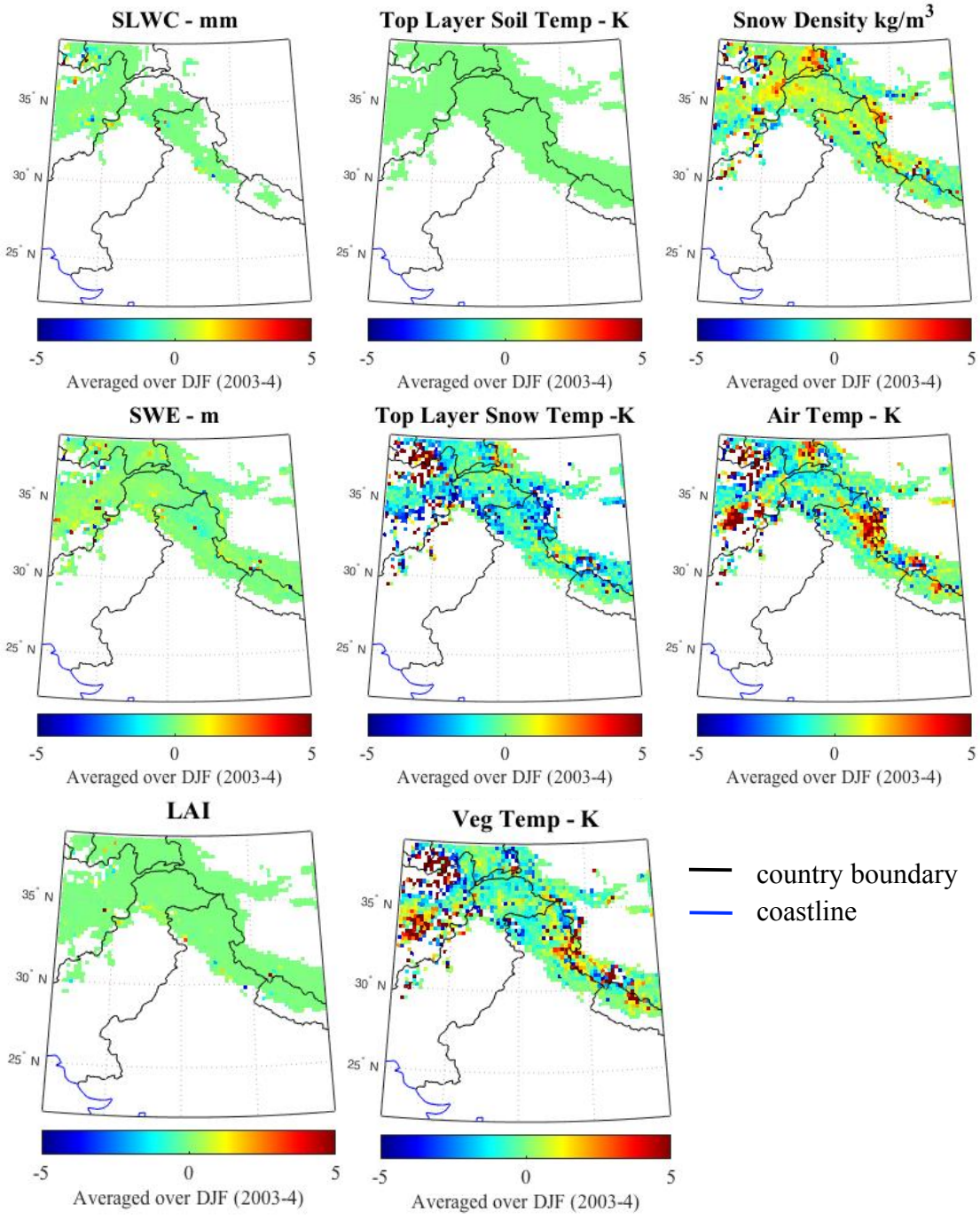

Figure 4.2. Maps of normalized sensitivity coefficients of SVM predicted ΔT_b(18.7V-36.5V) for LIS input states averaged over the snow accumulation period (Dec-2003, Jan-2004, Feb-2004) for snow-covered areas in the Indus Basin.

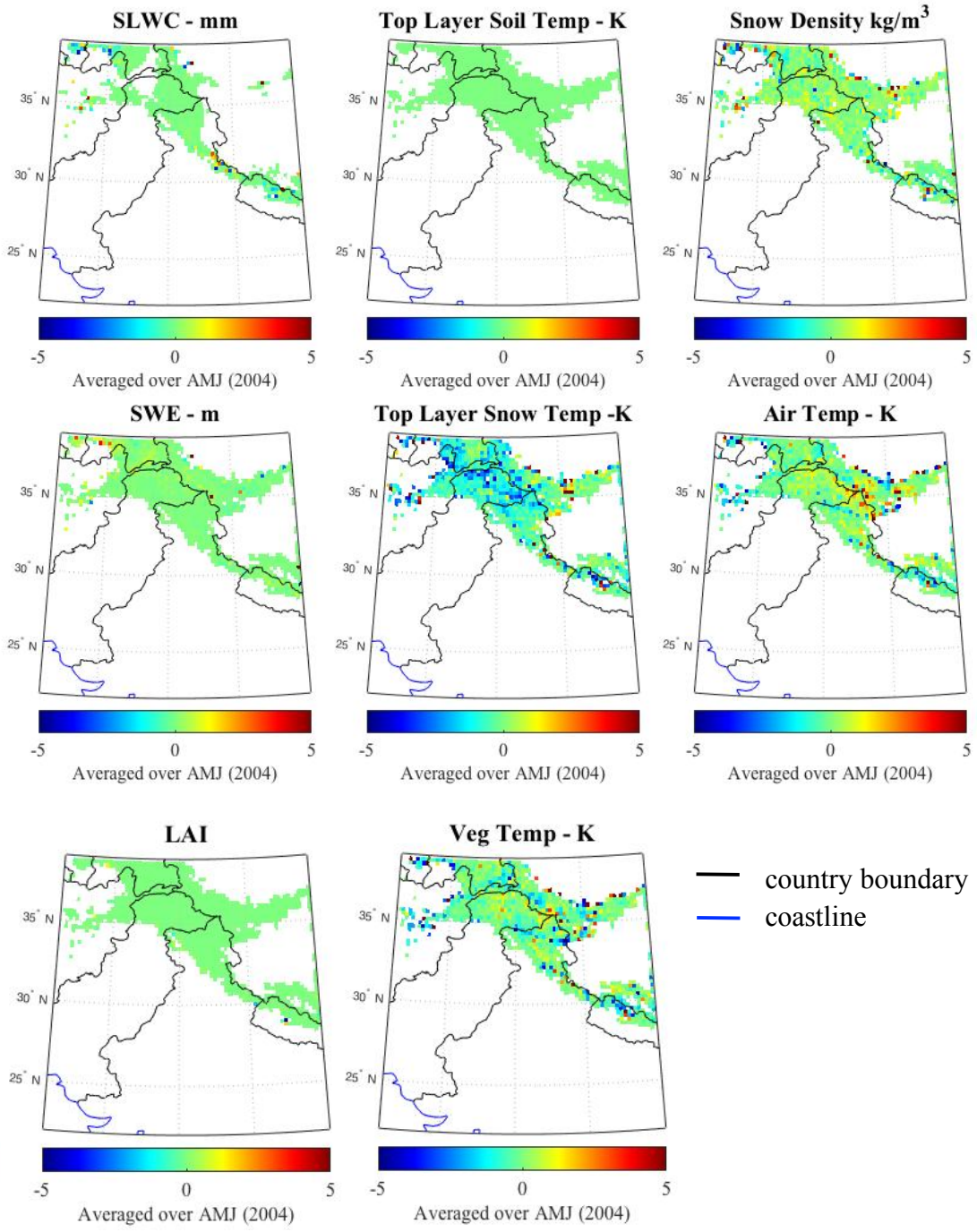

Figure 4.3. Maps of normalized sensitivity coefficients of SVM predicted ΔT_b(18.7V–36.5V) for LIS input states averaged over the snow ablation period (Apr, May, Jun; year = 2004) for snow-covered areas in the Indus Basin.

The highest NSC magnitudes are observed for snow-pack top-layer snow temperature. Most of the NSCs for snow temperature have a negative sign. This seems physically rational as for a higher snow temperature, the snow pack is expected to have a reduced amount of snow (SWE) and thus a decreased SVM predicted ΔT_b (18.7V − 36.5V) is expected. This results in an inverse relationship between snow temperature and ΔT_b (18.7V − 36.5V). The negative NSC sign highlights this inverse relationship.

Air temperature and vegetation temperature do not have a direct relationship with ΔT_b (18.7V − 36.5V), instead their relationship is defined through other states, e.g., if there is a cold (low) air temperature present, this will generally correspond with a lower snow temperature (ignoring effects of other physical processes). A lower snow temperature could indicate a deeper snow pack, which will correspond to a high ΔT_b (18.7V − 36.5V) value. This implies that (ignoring the effect of other physical processes) a decrease in air temperature would result in an increased ΔT_b (18.7V − 36.5V), indicating an inverse relation and a negative NSC sign. Figure 4.2 and Figure 4.3 show a majority of positive NSC values for air temperature. This can be explained by: (a) high cross-correlation of air temperature with other LIS input states such as snow temperature (Table 3.4), (b) effect of vegetation (increased in the snow ablation months) on T_b (as discussed in Chapter 2), and (c) presence of wet snow, especially in the summer months, and its effect on T_b.

In Figures 4.2 and 4.3, it is observed that soil temperature (top-layer of soil) has NSC values that are all equal to zero, indicating that it had no effect on ΔT_b (18.7V − 36.5V) SVM prediction. Vegetation temperature and air temperature also exhibit relatively high values as compared to SLWC (snow liquid water content) and LAI (leaf area index). Higher sensitivity to SLWC is expected during the snow ablation months (due to wetter snow) as compared to snow accumulation months (reduced presence of SLWC). Figures 4.2 and 4.3 corroborate this expectancy.

For SWE, rationality generally suggests positive values of NSCs. In Figures 4.2 and 4.3, we observe that although the magnitudes are relatively small throughout the domain (expect for few locations), majority of the values are positive. NSCs for locations collocated with glaciers (Figure 4.1) are not very reliable.

From the discussion above, it is concluded that the NSC value we obtain for each LIS input state is a result of a number of concurrent and interacting physical processes, high cross-correlation between the LIS input states and effect of location specific parameters such as dense vegetation and glaciers.

4.2. Temporal Analysis of Normalized Sensitivity Coefficients

A temporal analysis of NSCs was carried out to broaden our understanding of the change that the relationship between LIS input states and SVM predicted output undergoes throughout the year. The time-period of the analysis presented here spans from 1st September 2003 to 31st August 2004. This marks one complete water year.

4.2.1. Test Location Time-Series of NSCs

A test location was selected to study the time-series of NSCs for all LIS input states. The location selection criteria included:

- Not located on a glacier/ ice-mass
- absence of dense vegetation
- seasonal snow observed.

This list was developed to locate a spot that would help correctly identify the relationship between well-trained SVM predicted output and LIS input states by avoiding erroneous AMSR-E T_b measurement areas.

The time series in Figure 4.4 displays that the bottom-layer snow temperature and top-layer soil temperature NSCs had either no value (no data) or were equal to zero. The highest variation was seen in NSCs of snow density (values range between -3.8 to 3). Snow (top-layer), air, and vegetation temperature had more homogenous NSC values during the winter months and have highly varying NSCs during the summer months. SWE shows values that were mostly clustered between -1 and 1. SLWC had few available NSC values during the wet snow season and had no values present for the winter season (absence of moisture in the snow).

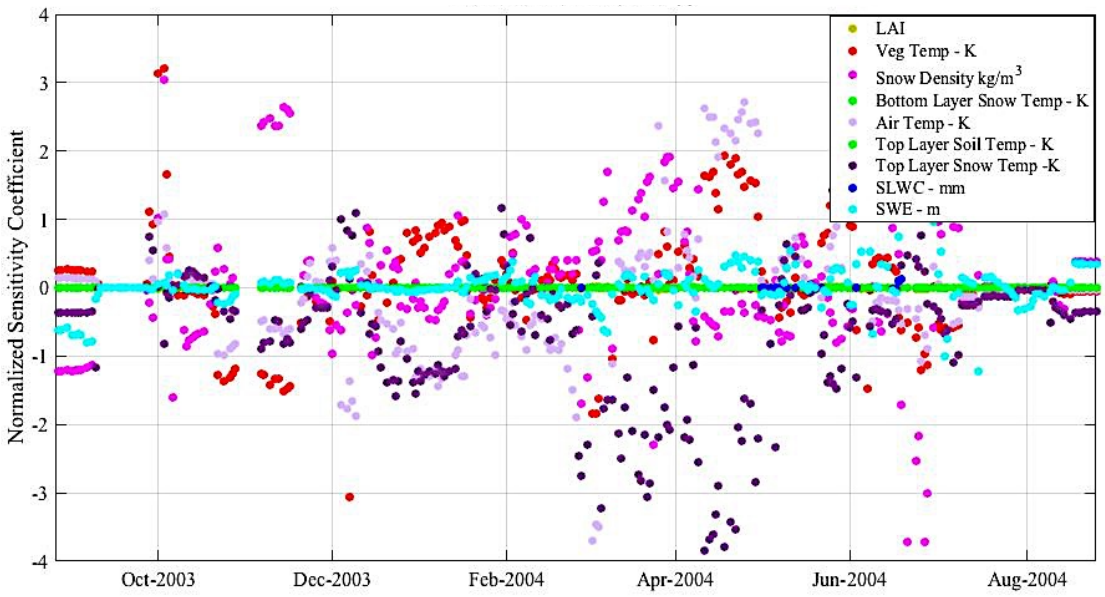

Figure 4.4. Time-series of NSCs of SVM predicted ΔT_b (18.7V – 36.5V) for LIS input states, calculated using a perturbation of +/-2.5% for test location (35.73^{O}N, 76.28^{O}E).

4.2.1.1. Absolute NSCs for test location

To generalize the highly varying time series in Figure 4.5 and to deduce some simple conclusions, the test location absolute average NSCs for the snow accumulation and ablation months were calculated for the more sensitive of LIS input states (states which have < 95% of the NSC values = 0).

As discussed in Chapter 1, T_b (in the microwave band) is dependent on emissivity and physical temperature. The first four LIS input states in Figure 4.5 affect emissivity while the last three are related to physical temperature.

In Figure 4.5, it is observed that for the snow accumulation period, snow temperature (top-layer of snow pack) had the highest absolute NSC of 1.4. All the physical temperature related states showed comparatively significant NSC values. Snow density and SWE had noticeable values due to their effect on emissivity, especially snow density which has the second-highest absolute NSC value. SLWC and LAI had negligible NSC values.

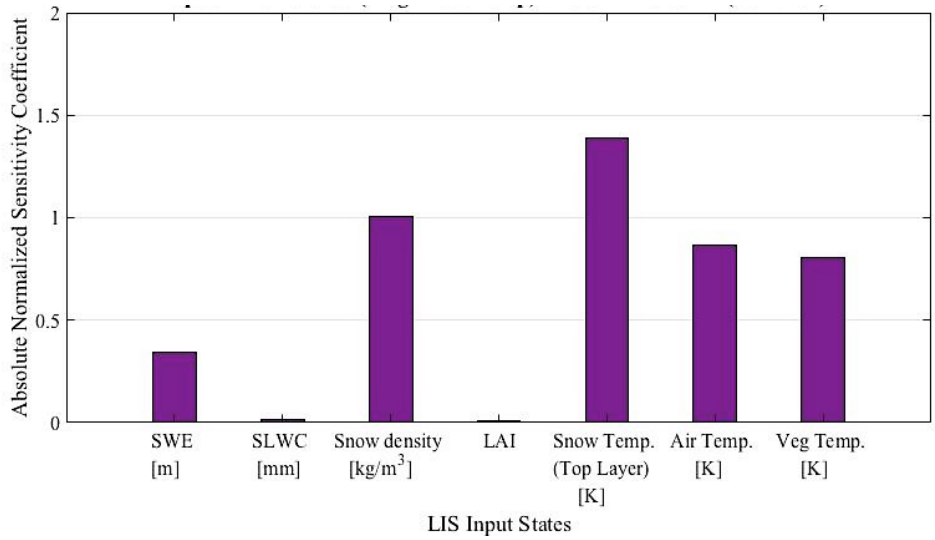

Figure 4.5. Absolute NSCs of SVM predicted ΔT_b (18.7V – 36.5V) averaged over the snow accumulation months during relatively dry snow conditions (Dec-2003, Jan-2004, Feb-2004) calculated using a perturbation value of +/-2.5% for test location (35.73^{O}N, 76.28^{O}E).

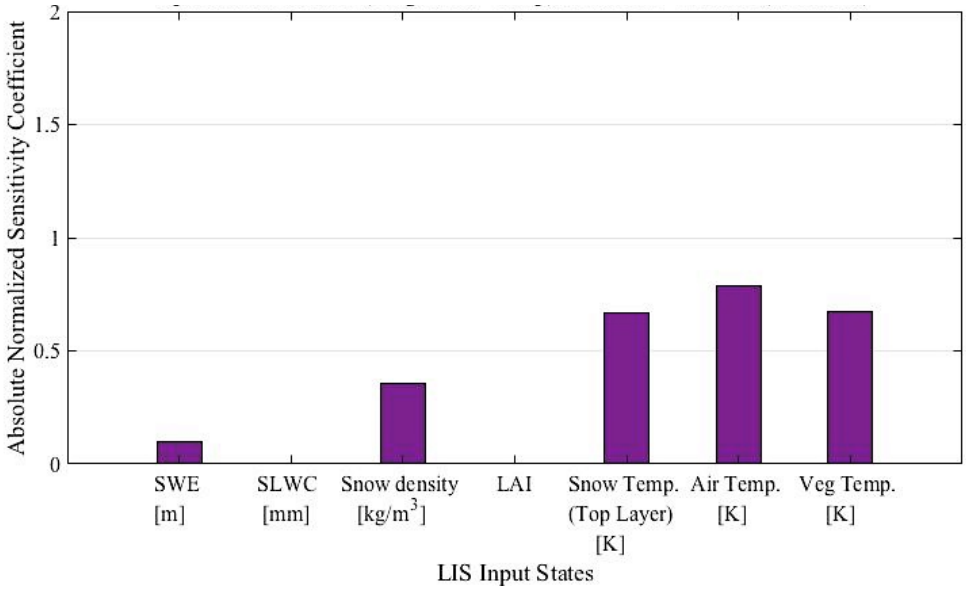

Figure 4.6. Absolute NSCs of SVM predicted ΔT_b (18.7V – 36.5V) averaged over the snow ablation months during relatively wet snow conditions (Apr, May, Jun -2004) calculated using a perturbation value of +/-2.5% for test location (35.73^{O}N, 76.28^{O}E).

In Figure 4.6 the snow ablation period averaged absolute NSC values were analyzed. Physical temperature related LIS input states had higher magnitudes as compared to the emissivity related states. SWLC and LAI had almost zero values.

4.2.2. Domain-wide Time Series of NSCs

One of the objectives was to find the domain-wide sensitivity of SVM predicted ΔT_b (18.7V − 36.5V) to LIS modeled SWE. This is intended to access the usability of SVM as the measurement operator (it maps the model states into the measurement space) in a SWE data assimilation framework. Therefore, if the sensitivity of SWE is significant enough and the accuracy of ΔT_b (18.7V − 36.5V) measurements being assimilated is appropriate, we expect an improvement in the ultimate SWE estimates for the domain. These estimates will be a combination of the LIS modeled SWE and the assimilated ΔT_b (18.7V − 36.5V) measurements.

Figure 4.7 presents the domain wide (snow-covered areas only) NSC range of SVM predicted ΔT_b (18.7V − 36.5V) for LIS SWE from Sep-2003 to Aug -2004. Outliers have been omitted in this figure to give a general idea of the snow-covered areas SVM prediction sensitivity to SWE.

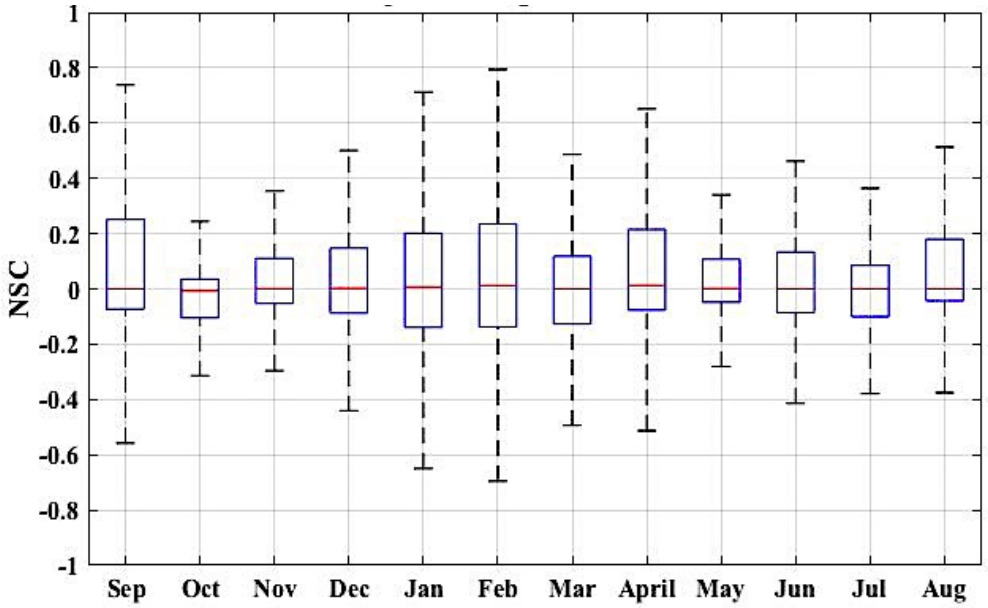

Figure 4.7. Monthly boxplots (outliers not shown) of NSCs of (LIS inputs states =10) SVM predicted ΔT_b (18.7V − 36.5V), using a perturbation value of +/-2.5%, for LIS modeled SWE from Sep-2003 to Aug-2004 for the snow-covered areas in the Indus Basin. Red line represents the median value while the blue box depicts the interquartile range.

The box-plots for each month differ considerably in their total range (-0.7 to 0.8 in Feb as compared to -0.3 to 0.25 in Oct). Most of the NSCs seem to lie between -0.175 and 0.3. This indicates that sensitivity of SVM output to LIS SWE for most of the areas within the domain is not very high.

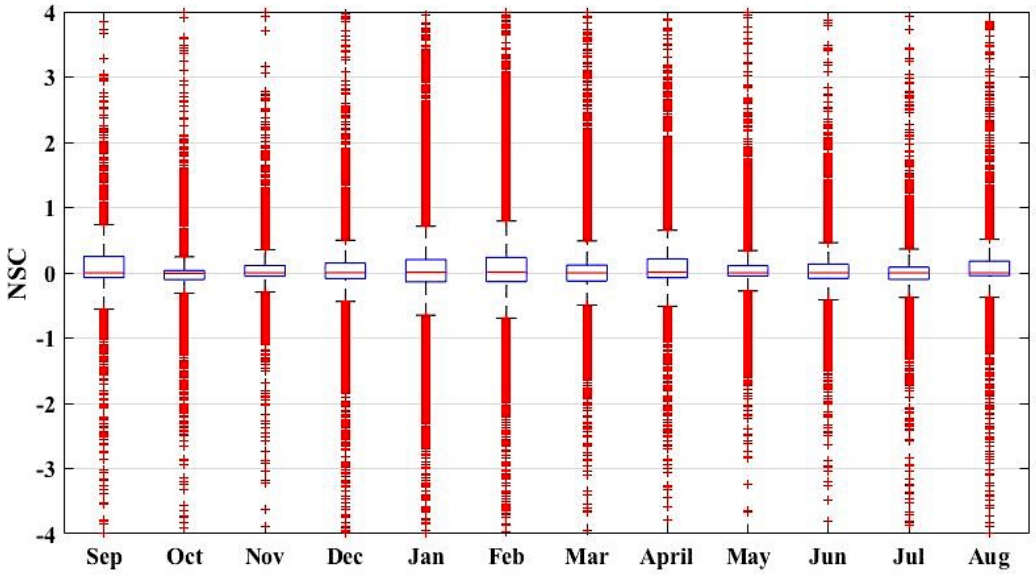

Figure 4.8. Monthly boxplots (outliers shown) of NSCs of (LIS inputs states =10) SVM predicted ΔT_b (18.7V – 36.5V), using a perturbation value of +/-2.5%, for LIS modeled SWE from Sep-2003 to Aug -2004 for the snow-covered areas in the Indus Basin.

Figure 4.8 shows the same boxplots as in Figure 4.7 with outliers visible. Outliers are classified as data points that are either 3 times the interquartile range or more above the third quartile or 3 times the interquartile range or more below the first quartile. Since a large amount of data (pixels) is being considered here, the number and range of outliers for each month is quite large. It can be inferred from Figures 4.7 and 4.8, that although SVM output (obtained using all 10 LIS input states) sensitivity to LIS SWE is not very high for all the snow-covered areas in the domain, it is significant enough for some places during some periods of the year. As far as SWE data assimilation is concerned, the results seen in Figure 4.7 and Figure 4.8 suggest that noticeable improvement in SWE estimates through ΔT_b (18.7V – 36.5V) assimilation is expected for some places during some periods of the year.

The total number of NSCs calculated for each month differs. NSCs were calculated for those pixels only which had a defined magnitude for each of the 10 LIS input states, apart from having LIS SWE > 1cm (Table 4.1).

Table 4.1. Total no. of pixels having a defined NSC value of SVM predicted ΔT_b (18.7V – 36.5V) for LIS SWE over snow-covered areas in Indus Basin (10 LIS input states used during prediction)

Month (Year 2003-4)	No. of NSC values
Sep	2009
Oct	5821
Nov	8062
Dec	19055
Jan	26211
Feb	27855
Mar	20479
Apr	11257
May	10146
Jun	3904
Jul	2825
Aug	2812

4.3. SVM Prediction using 4 LIS Input States

Continuing the SWE data assimilation discussion, it was analyzed whether the SVM output sensitivity to LIS SWE can be increased by using fewer LIS input states for SVM training and prediction. The 4 LIS input states selected for this analysis were according to the initial ordering of the states (Table 3.1).

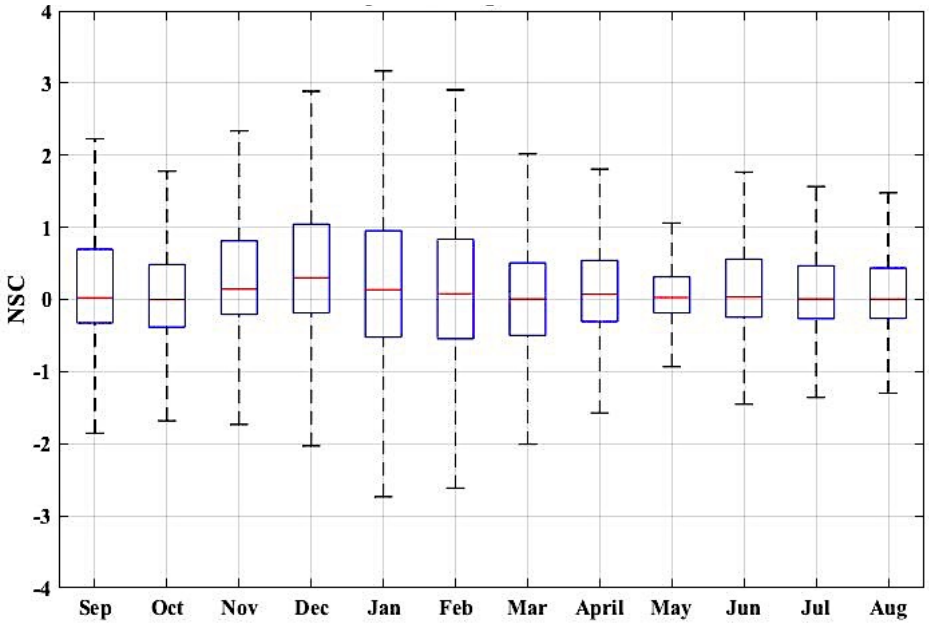

Figure 4.9. Monthly boxplots (outliers not shown) of NSCs of (LIS inputs states =4) SVM predicted ΔT_b (18.7V – 36.5V), using a perturbation value of +/-2.5%, for LIS modeled SWE from Sep-2003 to Aug-2004 for the snow-covered areas in the Indus Basin.

Figure 4.9 shows that the sensitivity can indeed be increased by decreasing the number of LIS inputs. Monthly boxplots of NSCs show increased range and interquartile NSC magnitudes for all months. Box-plots for each month still differ considerably in their total range (-2.7 to 3.2 in Jan as compared to -0.9 to 1.1 in May). Comparing each monthly box-plot to those in Figure 4.7, considerable change in the interquartile range is observed. The median monthly values however seem to remain unchanged except for the months of Dec and Jan.

Total number of NSCs calculated for each month differs. NSCs were calculated for those pixels only which had a defined magnitude for each of the 4 LIS input states, apart from having LIS SWE > 1cm (Table 4.2).

Table 4.2. Total no. of pixels having NSC value of SVM predicted ΔT_b (18.7V – 36.5V) for LIS SWE over snow-covered areas in Indus Basin (4-LIS input states used during prediction)

Month (Year 2003-4)	No. of NSC values
Sep	2823
Oct	7746
Nov	13937
Dec	25942
Jan	35515
Feb	37878
Mar	27910
Apr	15449
May	13731
Jun	5645
Jul	3902
Aug	3879

NSC spatial and temporal analysis is utilized to detect the effect of decreased number of LIS input states on sensitivity of SVM predicted ΔT_b (18.7V – 36.5V) to each remaining LIS input state (total = 4).

4.3.1. Spatial Analysis of SVM Prediction using 4 LIS Input States

Comparing Figure 4.10 with Figure 4.2, an increase in the NSCs for SWE and snow temperature (top-layer) is observed. NSC values for SLWC and soil temperature (top-layer) however remain almost the same. Thus, decreasing the number of LIS inputs visibly increased the sensitivity of two of the states (SWE and top-layer snow temperature) and confirmed the zero sensitivity of soil temperature. It can thus be concluded that SVM predicted ΔT_b (18.7V – 36.5V) is independent of soil temperature.

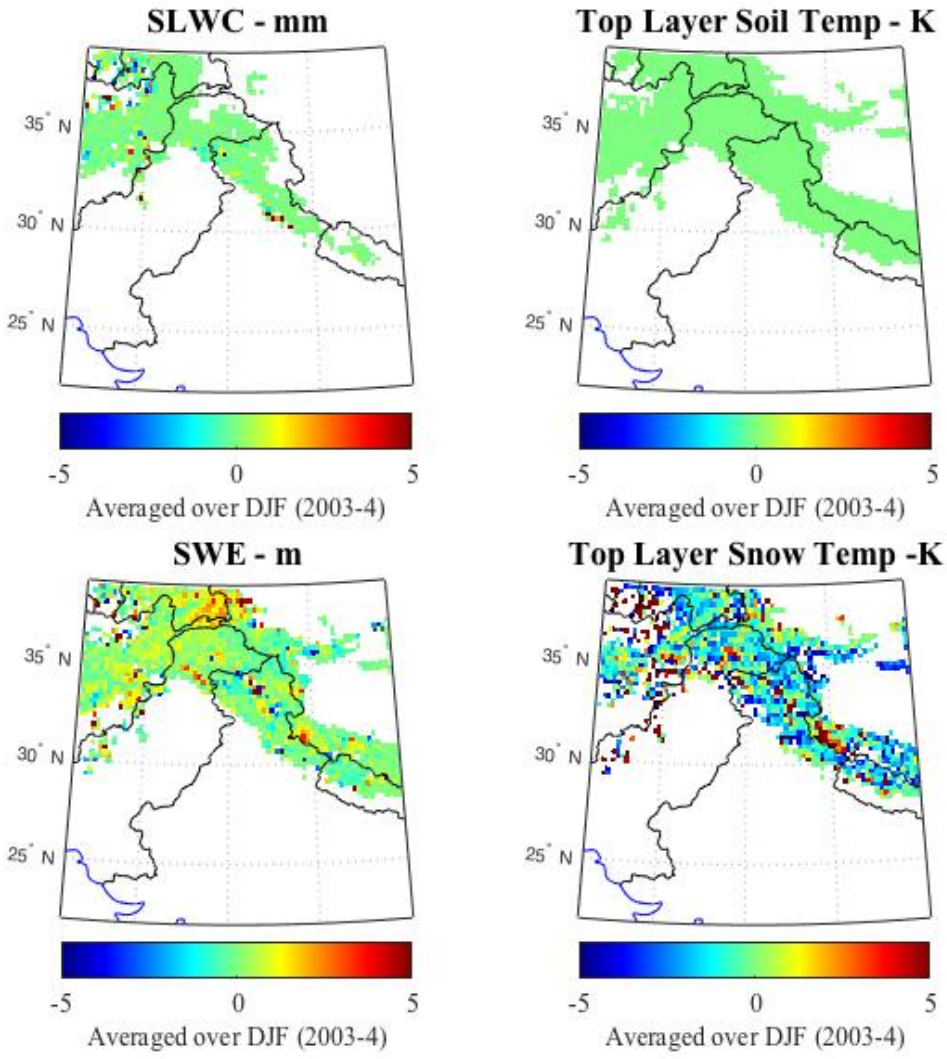

Figure 4.10. Maps of NSCs of SVM predicted ΔT_b (18.7V − 36.5V), using 4 LIS input states, averaged over the snow accumulation period (Dec-2003, Jan-2004, Feb-2004) for snow-covered areas in the Indus Basin. The solid black line represents country boundaries and the solid blue line depicts the coastline.

Similar comparisons are visible in Figure 4.11 and Figure 4.3. Sensitivity of two states (SWE, and top-layer snow temperature) is increased, whereas top-layer soil temperature maintains its zero sensitivity. SLWC NSC values do not demonstrate any visible change.

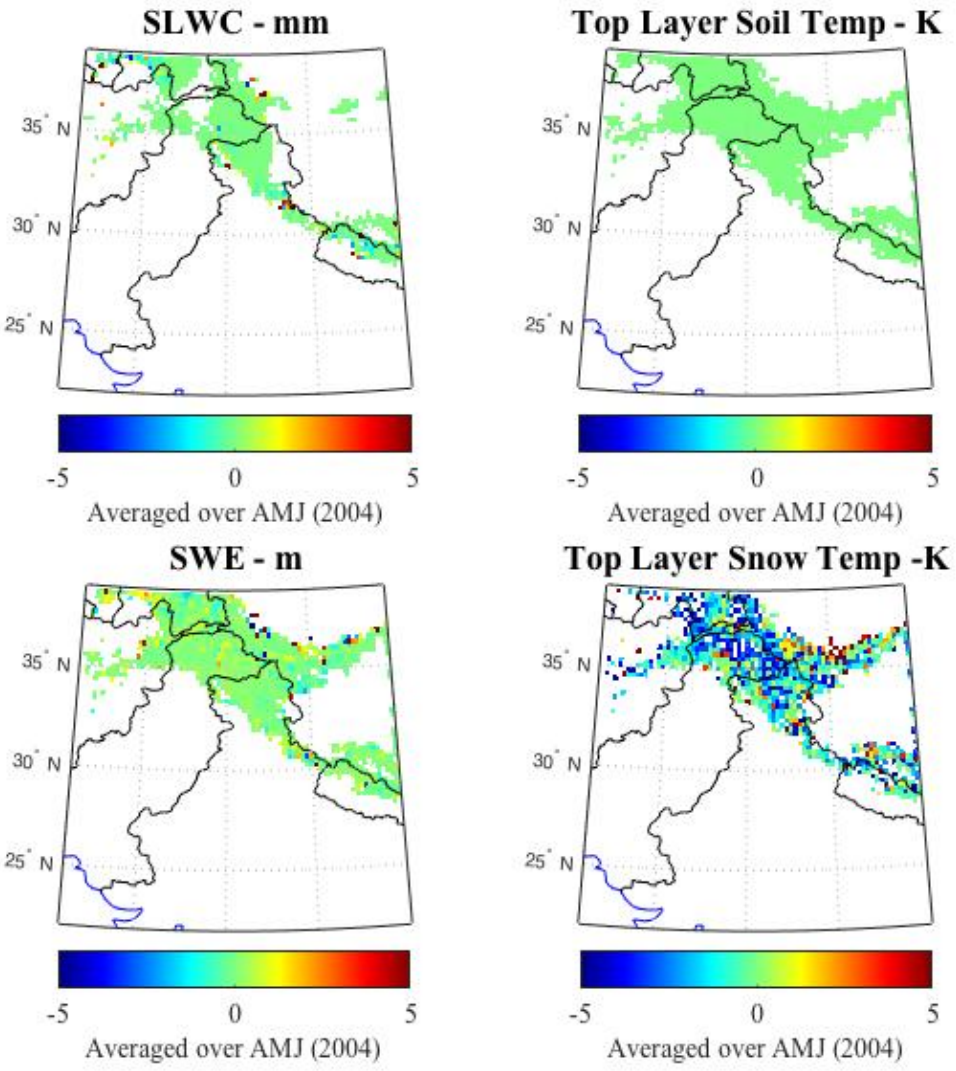

Figure 4.11. Maps of NSCs of SVM predicted ΔT_b (18.7V – 36.5V), using 4 LIS input states, averaged over the snow ablation period (Apr, May, Jun - 2004) for snow-covered areas in the Indus Basin. The solid black line represents country boundaries and the solid blue line depicts the coastline.

4.3.2. Temporal Analysis of SVM Prediction using 4 LIS Input States

Time-series of NSCs for the same test location described in Section 4.2.1 is presented in Figure 4.12. The magnitude range and variance of SWE NSCs increased visibly as compared to Figure 4.4. SLWC and top-layer soil temperature maintained their NSC values. Top-layer snow temperature NSCs displayed increased variance throughout the year.

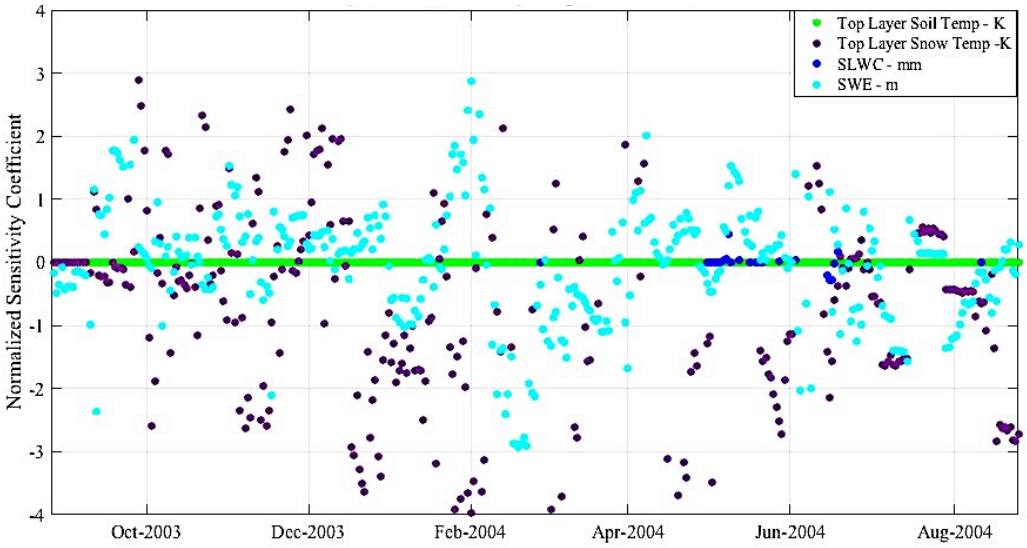

Figure 4.12. Time-series of NSCs of SVM predicted ΔT_b (18.7V – 36.5V), using 4 LIS input states, calculated using a perturbation of +/-2.5% for test location (35.73^{O}N, 76.28^{O}E).

Absolute average NSCs for snow accumulation and ablation periods in Figure 4.13 and Figure 4.14 corroborate the increase in SWE and top-layer snow temperature NSC values.

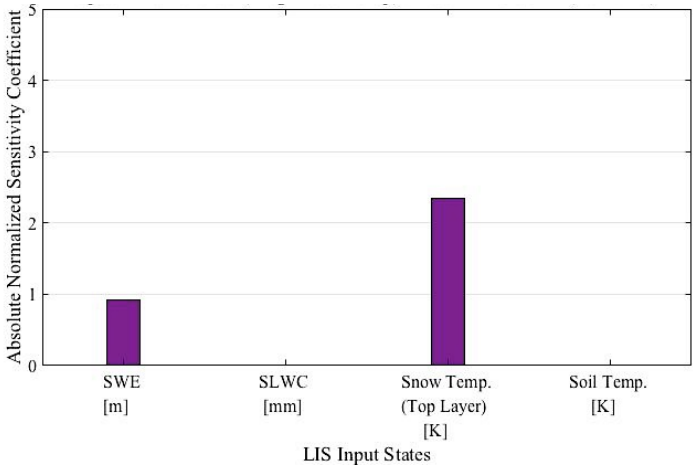

Figure 4.13. Absolute NSCs of SVM predicted ΔT_b (18.7V – 36.5V), using 4-LIS input states, averaged over the snow accumulation months (Dec-2003, Jan-2004, Feb-2004) calculated using a perturbation value of +/-2.5% for test location (35.73^{O}N, 76.28^{O}E).

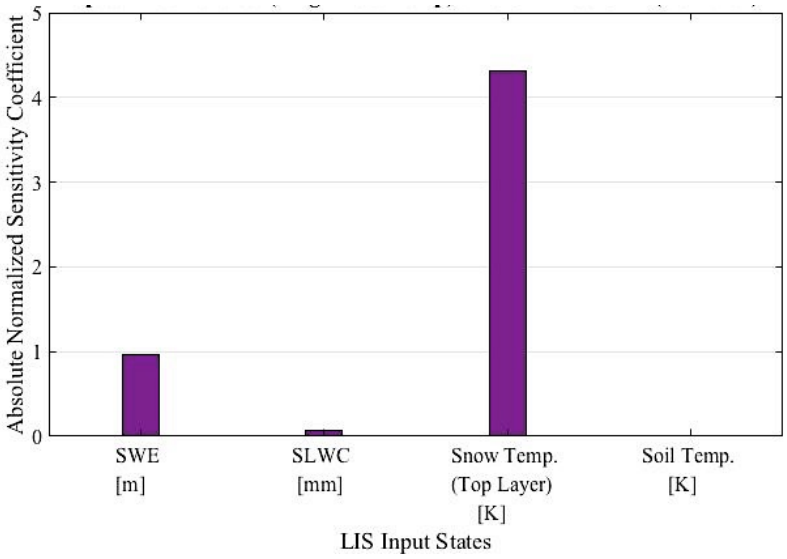

Figure 4.14. Absolute NSCs of SVM predicted ΔT_b (18.7V-36.5V), using 4-LIS input states, averaged over the snow ablation months (Apr, May, Jun -2004) calculated using a perturbation value of +/-2.5% for test location (35.73^{O}N, 76.28^{O}E).

4.4. Relative Importance of Predictors

It must be considered that although decreasing the number of LIS inputs increases the sensitivity to SWE, it can also affect the prediction accuracy of ΔT_b(18.7V-36.5V). If the prediction accuracy is significantly decreased, a poorer SWE estimate would result after ΔT_b(18.7V-36.5V) assimilation instead of an improved SWE estimate.

The objective is to achieve high SWE sensitivity while maintaining SVM prediction accuracy. For this purpose, the relative important of LIS input states is assessed, i.e., identification of LIS input states with the highest relative sensitivity. Therefore, if only those LIS input states (apart from SWE) are utilized for SVM prediction that have comparatively higher relative sensitivity, the objective of high SWE sensitivity (due to fewer LIS inputs states) without compromising considerably on SVM prediction accuracy is achieved.

Absolute NSC values of SVM predicted ΔT_b (18.7V - 36.5V) using all 10 LIS input states were calculated for this purpose. In Figure 4.15 and Figure 4.16, maps of 8 LIS input states are presented. NSC maps of snow temperature (bottom-layer of snow pack) and soil moisture are not included due to reasons discussed in Section 4.1.

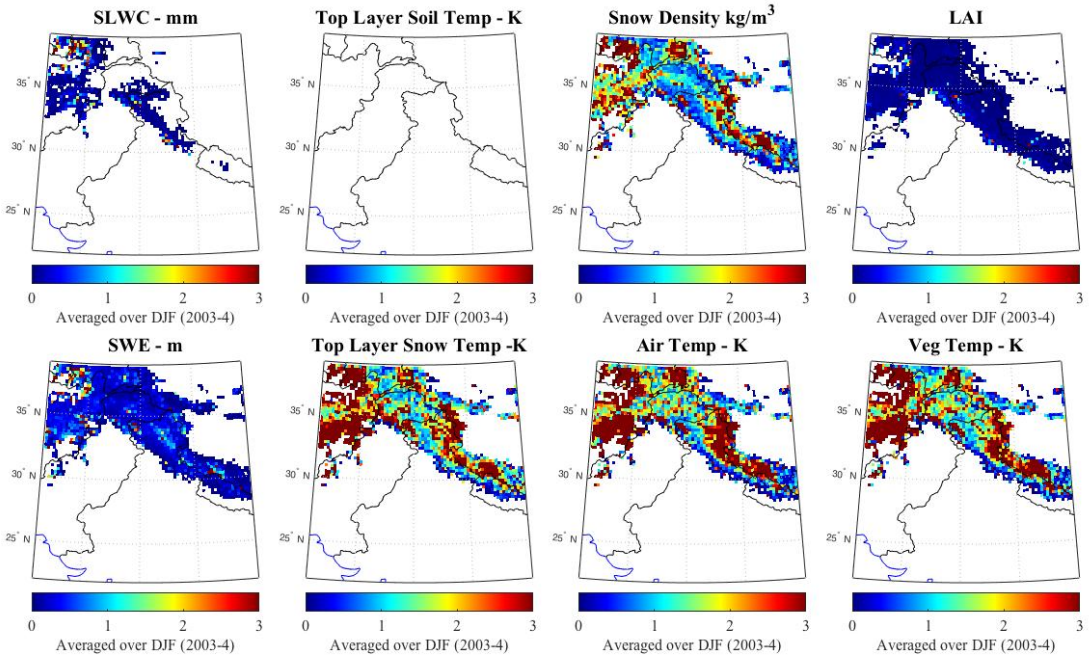

Figure 4.15. Maps of absolute NSCs of SVM predicted ΔT_b (18.7V – 36.5V) for LIS input states, averaged over the snow accumulation period (Dec-2003, Jan-2004, Feb-2004) for snow-covered areas in the Indus Basin.

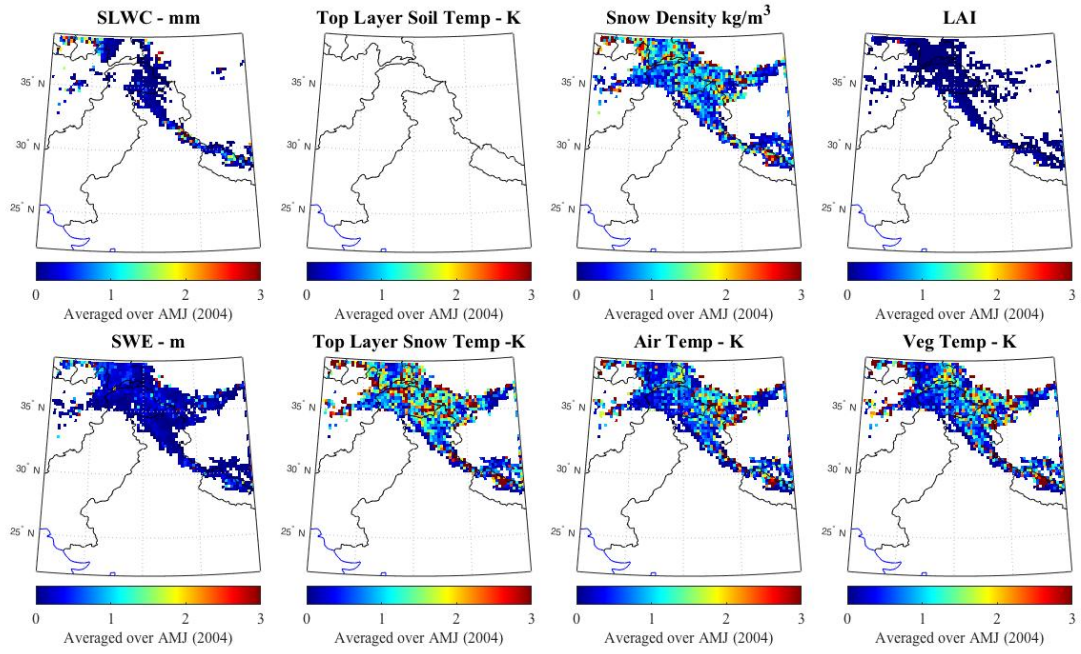

Figure 4.16. Maps of absolute NSCs of SVM predicted ΔT_b (18.7V – 36.5V) for LIS input states, averaged over the snow ablation period (Apr, May, Jun; year = 2004) for snow-covered areas in the Indus Basin.

Highest sensitivity for snow covered areas in the domain is dominated by the three temperature values; snow (top-layer) temperature, air temperature, and vegetation temperature. Snow density displays comparatively higher values than SWE, SLWC, and LAI. As observed previously, soil temperature maintains zero sensitivity throughout the domain. Comparing Figure 4.15 and Figure 4.16, NSC magnitudes are much higher for the snow accumulation period than the snow ablation period. One possible reason could be the increased snow presence during the accumulation months as compared to the ablation months.

Based on the discussion and figures presented in this chapter, the order of decreasing relative sensitivity of SVM predicted ΔT_b (18.7V − 36.5V) to LIS input states is:

1. Snow temperature
2. Air temperature
3. Vegetation temperature
4. Snow density
5. SWE
6. SLWC
7. LAI
8. Top-layer soil temperature

4.5. Limitations

Some of the limitations that inhibited this sensitivity analysis were:

- LIS (Noah MP) estimation of soil moisture under frozen soil conditions was physically erroneous and was thus excluded from the sensitivity analysis.
- Since the LIS model grid (0.01^{O}x0.01^{O}) used during the geophysical land surface states estimation was different from the AMSR-E brightness temperature grid resolution (25km x 25km), LIS modeled states were re-gridded to the AMSR-E EASE-Grid for consistency. This re-gridding added uncertainty into the LIS state estimates.
- Coarse resolution of the AMSR-E brightness temperature measurements can introduce representativeness errors.

- Most of machine learning techniques are black box methods. This means there are certain key aspects of the SVM model that are not observed or analyzed.
- Generalization of results is difficult for large varying domains, especially for complex models like the SVM.
- Cross-correlation between the LIS states is ignored in NSC generation. By perturbing one element only while maintaining the original value of all others, independence between the states is assumed. This is not true as seen by the cross-correlation matrix. Nonetheless, this correlation is ignored during individual perturbation and subsequent NSC calculation. This introduces irrationality in the NSC values, especially in terms of the sign associated with each magnitude.

Chapter 5: Assessment of LIS Modeled States

As discussed in Chapter 3, SVM training and (brightness temperature spectral difference) prediction accuracy is affected by the accuracy of the LIS geophysical states used as input. Error in LIS modeled states will translate into error in the subsequent SVM predictions. One of the ultimate goals is to refine the prediction skill of SVM. One approach towards achieving this goal is to improve the accuracy of LIS modeled states through comparison with measurement based data. This defines the fourth objective discussed in Section 1.4.

This chapter deals with the comparison between the Advanced Scatterometer (ASCAT) based categorical Freeze/Melt/Thaw dataset, developed by the NASA HiMAT sub-team lead by Dr. McDonald at CUNY, and a LIS-based Freeze/Melt/Thaw product.

5.1. Advanced Scatterometer (ASCAT)

A Scatterometer is a radar that transmits microwave pulses down to the Earth's surface and measures the power (backscatter) returned to the instrument. The backscatter is then analyzed to extract information about the land surface. Advanced Scatterometer (ASCAT) is installed on the EUMETSAT (European Union Meteorological Satellite) MetOp satellite. It is a polar orbiting satellite with a 3-day overpass and completes its global coverage in 1-2 days. Data availability period of ASCAT extends from 2007 to the present.

5.2. ASCAT based Freeze/Melt/Thaw Dataset

ASCAT F/M/T record is derived from C-Band (5.255 GHz, vertical polarization) normalized backscatter measurements taken by the ASCAT instrument. Resolution enhancement is done using the SIR algorithm [69]. Detection of snowmelt and soil freeze/thaw status are determined from time-series singularities as detailed in [70] [71].

This is a qualitative rather than a quantitative dataset. Each pixel is defined as experiencing one of the three states; 'freeze', 'melt' or 'thaw'. All the states are dimensionless and represent only the presence or absence without any magnitude or unit. The pixel size is defined as 4.45 km x 4.45 km. This is a coarse grid resolution and accurate depiction of land surface states at such a scale encounters numerous difficulties. This dataset has been developed for years 2009 to 2016 (out of the total ASCAT data available). The study domain covers the whole of high mountain Asia. Dataset consists of mean daily values for areas of importance in the study domain.

Considering the coarse resolution of the measurements as well as expected errors in the algorithm utilized for converting the radar backscatter into freeze/melt/thaw states, the imperfections of the dataset are acknowledged beforehand. These are grouped into measurement errors. These measurement errors will be discussed further during the temporal and spatial dataset comparison detailed in Section 5.4.

5.2.1 Relevant Definitions

Since this is a qualitative dataset, the definition of each state is very important. Following are the definitions of the three categories as stated by the dataset developing team at CUNY:

- Melt

Melt is defined exclusively for those areas that experience either seasonal snow or perennial glaciers. If the radar detects a loss of backscatter that resembles melt over areas of deep, seasonally permanent snow and ice, that pixel is categorized as undergoing 'melt'.

- Thaw

Thaw is defined for land areas (soil) where there is no snow. Thawed conditions are defined where there is a seasonal radar brightening from vegetation and surface moisture.

- Freeze

The frozen condition is where/when thaw and melt are not occurring, for both areas of land and snow/ice cover.

Figure 5.1 shows an example of the dataset for Jan 1, 2011. A lot of data-holes are visible in the map. This is due to the fact that no category was allocated to pixels that lay outside the area of interest within the study domain or where the algorithm didn't perform well.

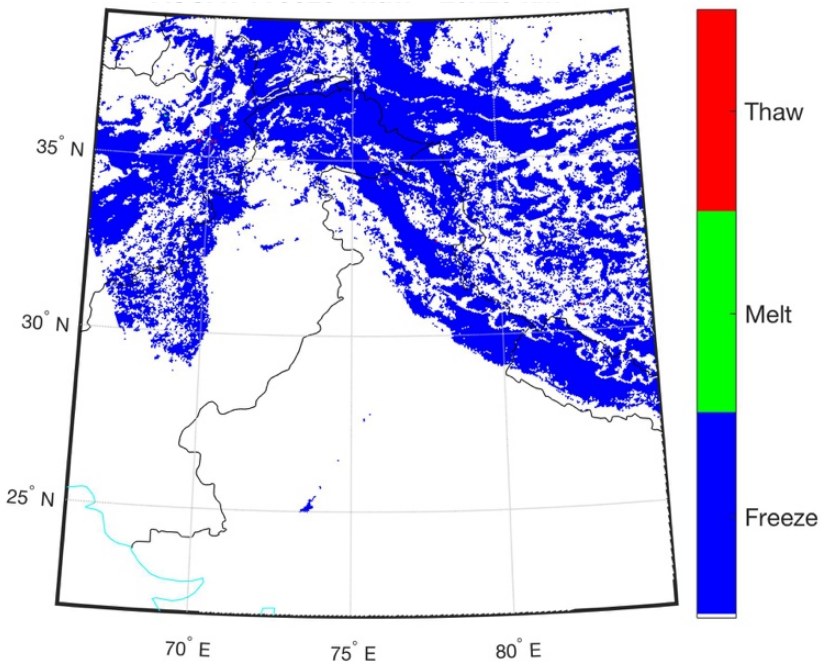

Figure 5.1. Map of the Indus Basin representing the ASCAT F/M/T pixel assignment for Jan 1, 2011. White color represents the pixels that lie outside the area of interest.

5.3. LIS Derived Freeze/Melt/Thaw Product

LIS is a land surface modeling framework and provides modeled geophysical land states as output. An algorithm based on the LIS modeled states was developed. Flowchart in Figure 5.2 provides an overview of the algorithm developed to define the LIS-based freeze/melt/thaw product.

Snow depth, air temperature, vegetation temperature, soil temperature, soil moisture and soil liquid water content used here belong to the same LIS (run) model output as was utilized previously in the sensitivity analysis of SVM predictions. Here, a very simplistic approach has been adopted to develop the pseudo-binary qualitative product. Since the fundamental objective was to achieve a first-order comparison

between the model (LIS) output and satellite measured data, this simple algorithm suited the purpose of this study.

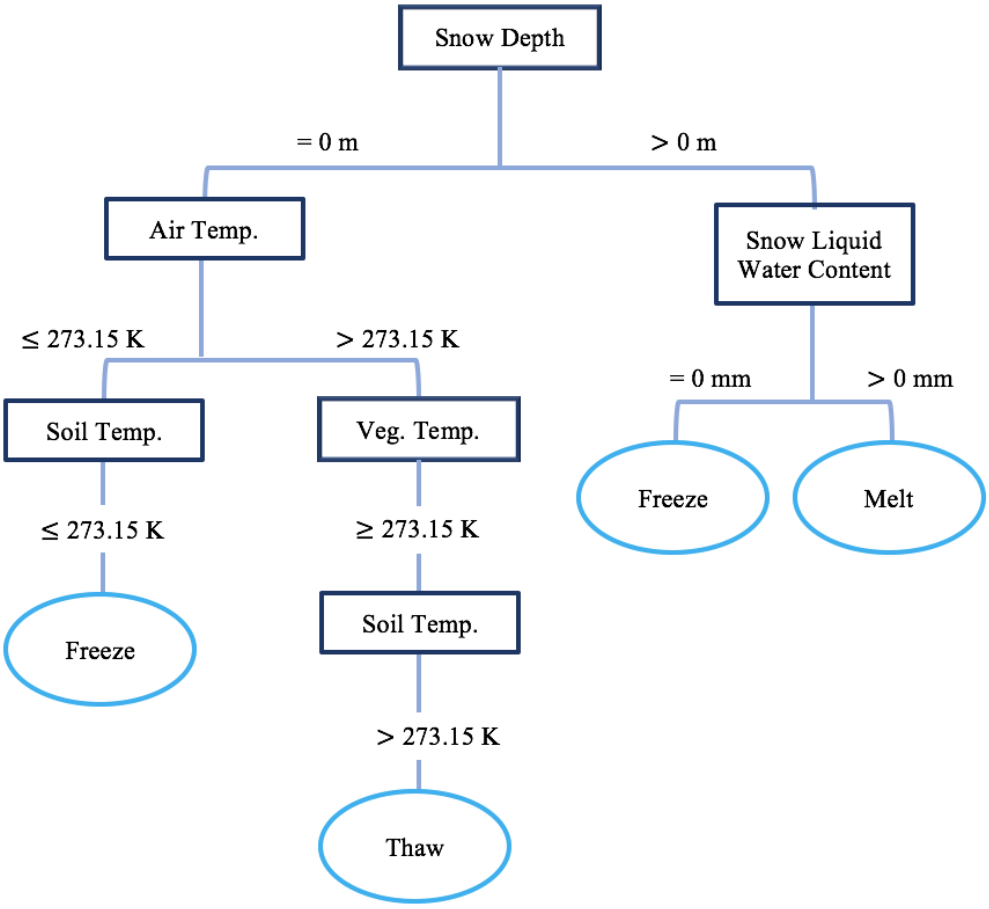

Figure 5.2. Flowchart defining the algorithm utilized for developing the LIS-F/M/T product

The three states categorized here have almost the same definitions as the ASCAT F/M/T product. This was particularly stressed upon to ensure an accurate comparison between the same data type obtained from two different sources i.e. land surface model and remotely sensed satellite measurements. The only difference between the two datasets lies in their characterization of 'freeze' state. ASCAT defines freeze as any place that is not undergoing thaw or melt within the area of interest. In the LIS algorithm, secondary restrictions are applied to the detection of 'freeze' state with consideration to some physical parameters (e.g. air temperature and soil temperature.)

A number of similar alternate algorithms where tested during the development phase (Figure 5.3). Selection criteria for the final algorithm included physical rationality and suitability to comparison with the ASCAT product.

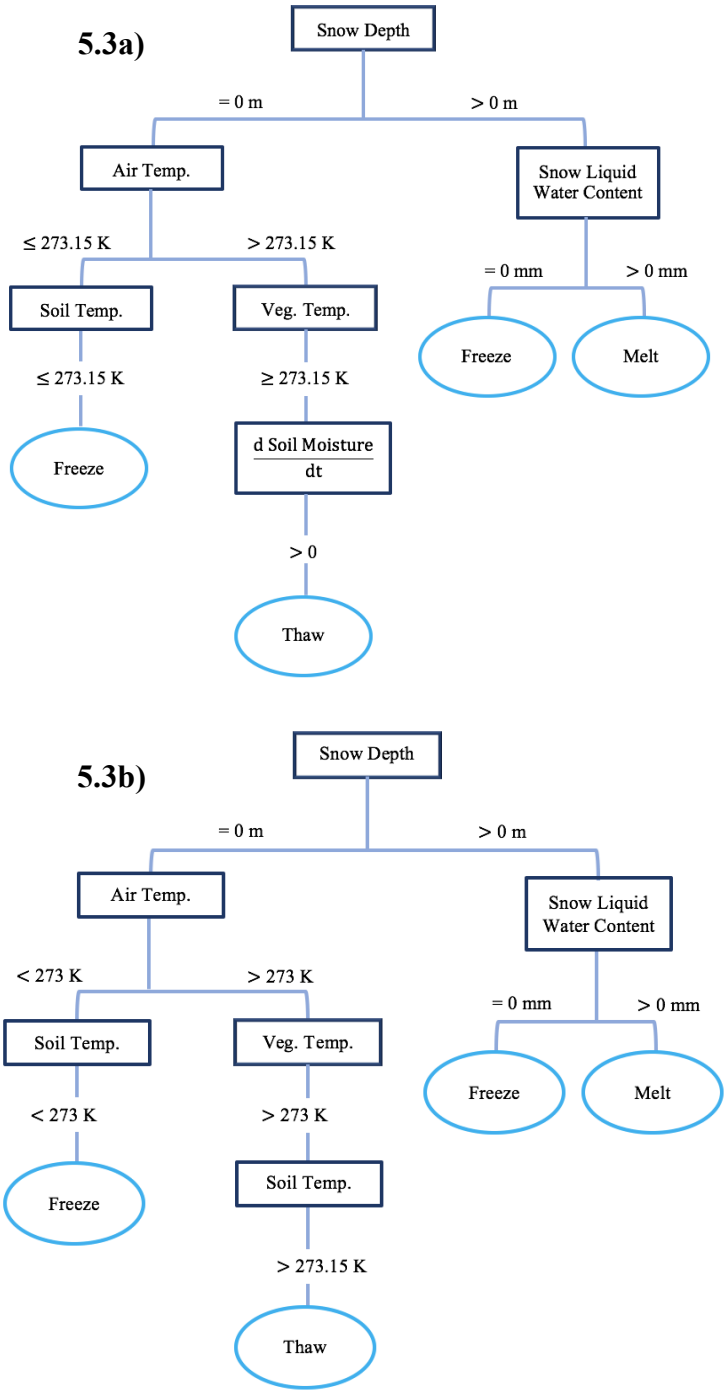

Figure 5.3. Flowcharts (5.3a and 5.3b) describe alternate algorithms tested for developing LIS F/M/T product.

5.3.1 Soil Temperature Preliminary Mask

Thaw was defined for those pixels only that experience a soil temperature below freezing point (<273.15K) at least once during a period extending from the preceding summer to the end of the study period. The rationale behind this restriction being that if the soil never froze, it cannot undergo thaw, since thaw is a physical process that succeeds freeze as temperatures begin to rise. A soil temperature mask was developed for this purpose. It is quite useful in identifying areas that are too warm throughout the year to undergo transition from one to the another of the three defined states. Such locations are ousted from the area of importance.

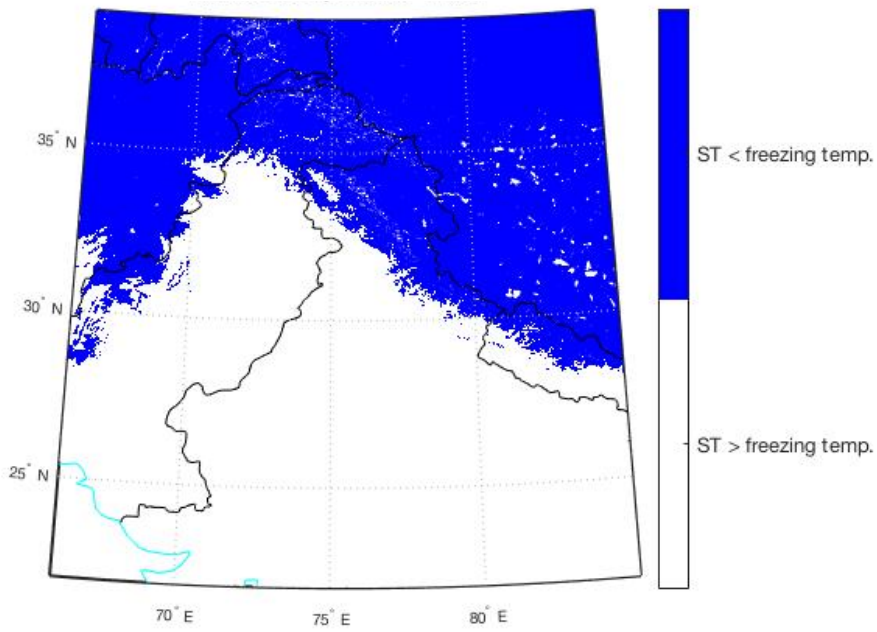

Figure 5.4. Soil temperature mask; blue color represents areas that experience below freezing soil temperature before or during the study period; white color represents areas that never experience below freezing temperature or that lie outside the dataset boundary.

5.3.2 LIS F/M/T Re-gridding

LIS F/M/T dataset was developed using the original 0.01^{O} x 0.01^{O} equidistant cylindrical grid. To maintain consistency between the two products, LIS dataset was re-gridded to the ASCAT F/M/T 4.45km x 4.45km grid. A 'drop in the bucket' re-gridding method was used. In this method, a pixel is categorized as one of the states

according to the average value of the original (0.01^{O} x 0.01^{O}) grid cells falling within its bounds. In this study, the idea of plurality was used instead of averaging. Thus, a pixel was defined by the state that had the highest plurality, with the lower threshold (percentage of pixels defined by that state) set at 30%.

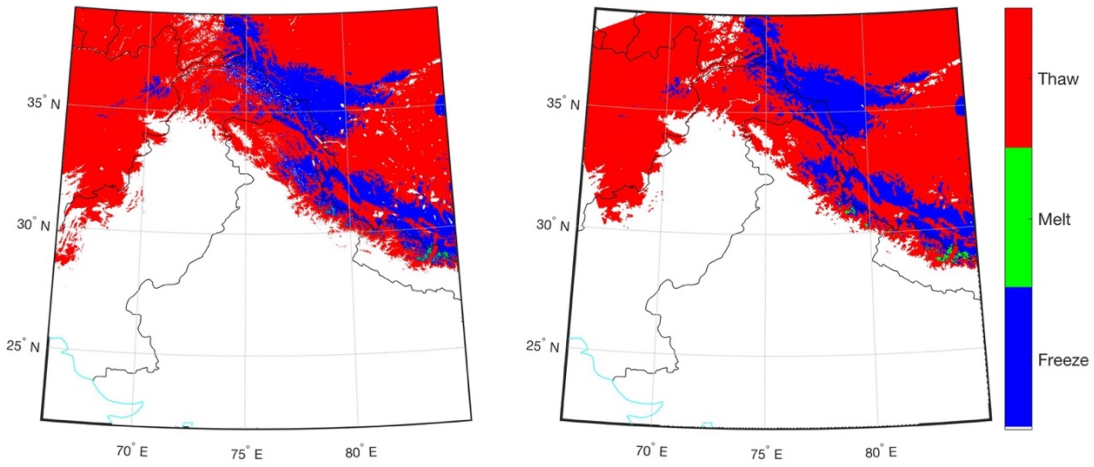

Figure 5.5. LIS F/M/T original dataset at 0.01^{O} x 0.01^{O} equidistant cylindrical grid (left) for Aug 1, 2011; LIS F/M/T dataset re-gridded to the ASCAT 4.45km x 4.45km grid (right) for Aug 1, 2011.

5.4. Comparative Analysis of ASCAT vs. LIS F/M/T

As discussed before, the ASCAT F/M/T product experiences data-holes. Similarly, due to the numerous quality checks applied on LIS geophysical states, there are some temporal and spatial gaps in the LIS F/M/T product as well. Thus, for this comparative analysis only those locations have been used where both datasets have one of the three states defined.

Here ASCAT F/M/T is considered as the observed data and LIS F/M/T as the modeled or forecast values. This is not a validation of the LIS modeled output as the ASCAT based dataset is also preliminary and is undergoing further refinement.

5.4.1 Spatial Analysis of ASCAT vs. LIS F/M/T

Figures 5.6, 5.7, 5.8 and 5.9 represent the spatial comparison between the two datasets. For Jan 1, 2011 (Figure 5.6), LIS F/M/T and ASCAT F/M/T seem to agree visibly regarding frozen areas. Both show that no snowmelt is detected. LIS F/M/T

identifies thawed soil locations in the upper-west corner of the study area, whereas ASCAT F/M/T shows no location under-going thaw conditions.

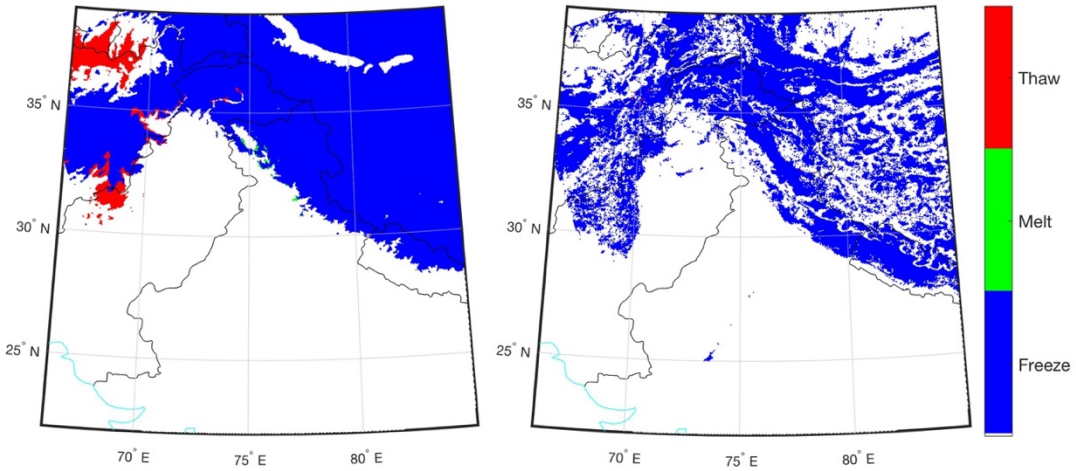

Figure 5.6. Maps of the Indus Basin displaying ASCAT F/M/T (right) and re-gridded LIS F/M/T (left) categorical pixel assignment for Jan 1, 2011.

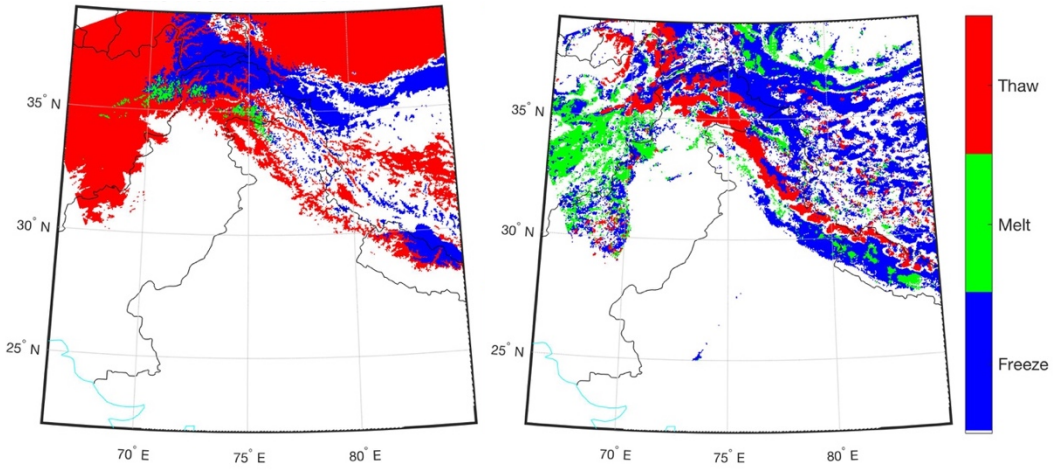

Figure 5.7. Maps of the Indus Basin displaying ASCAT F/M/T (right) and re-gridded LIS F/M/T (left) categorical pixel assignment for May 1, 2011.

For May 1, 2011 (Figure 5.7), both datasets display locations undergoing one of the three state conditions. Highest agreement between the datasets is observed in identifying frozen areas, whereas the least agreement is observable for snowmelt identification. Considering the climatology of the Indus Basin, 'freeze' is not expected

around lat: 30^{O}N and lon: 70^{O}E. For these locations, it seems that ASCAT F/M/T is over-estimating freeze. Above these co-ordinates (near 32^{O}N, 67.5^{O}E), ASCAT F/M/T shows snowmelt happening. In Figure 5.8, it is seen that LIS estimates zero snow depth at these locations. According to the LIS algorithm, snowmelt can only occur where there is a snowpack (snow depth > 0m) present. Since zero snow conditions are estimated by LIS, the possibility of snowmelt occurring in these areas is nullified. LIS F/M/T identifies these locations as undergoing thaw instead. This disagreement can indicate possible incorrect LIS snow depth estimation for these locations.

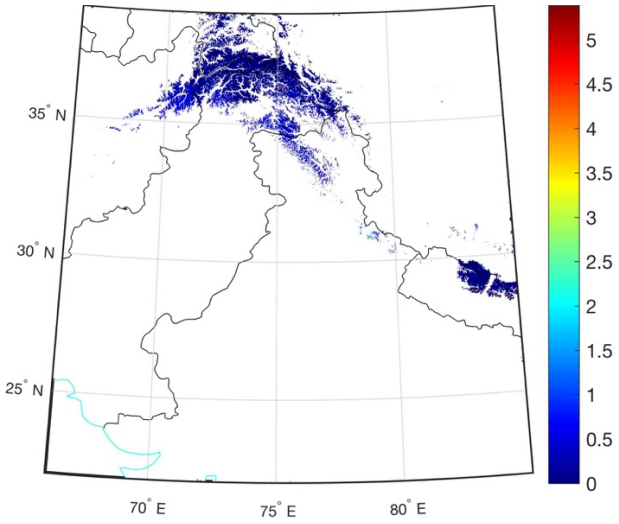

Figure 5.8. Map of Indus Basin showing LIS estimated snow depth [meters] for May 1, 2011.

The same snowmelt detection disagreement between the two datasets is perceived for Aug 1, 2011 (Figure 5.9). Here, freeze/thaw discord between the datasets is apparent as well. Locations identified as thawed in LIS F/M/T are categorized as frozen by ASCAT F/M/T. The Indus Basin experiences its peak summer temperatures around the end of July and start of August. This causes increased evaporation and subsequently decreased soil moisture. Apart from this, the rainfall season also initiates around this time for some parts of the basin. These events combine to result in an increased temporal variation in the surface moisture and hence affect the accuracy of the ASCAT backscatter analysis.

Figure 5.10 presents the land state conditions for Nov 1, 2011. November roughly marks the end of the autumn season and the beginning of the winter season for the Indus Basin. ASCAT F/M/T still identifies some locations as undergoing snowmelt whereas LIS F/M/T displays presence of freeze or thaw only. Recalling the state definitions, this could indicate possible error in LIS estimation of snow liquid water content since episodes of consecutive snowfall and snowmelt are expected in some areas.

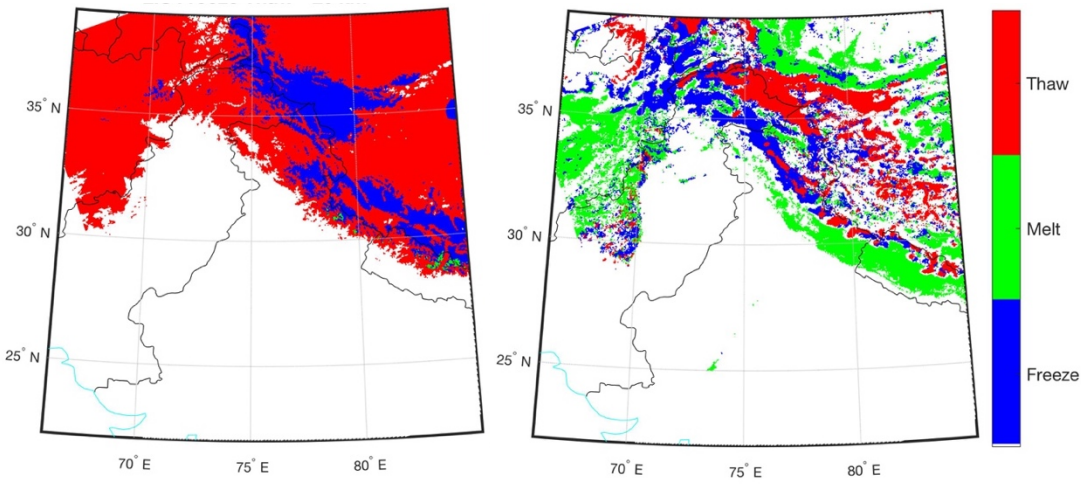

Figure 5.9. Maps of the Indus Basin displaying ASCAT F/M/T (right) and re-gridded LIS F/M/T (left) categorical pixel assignment for Aug 1, 2011.

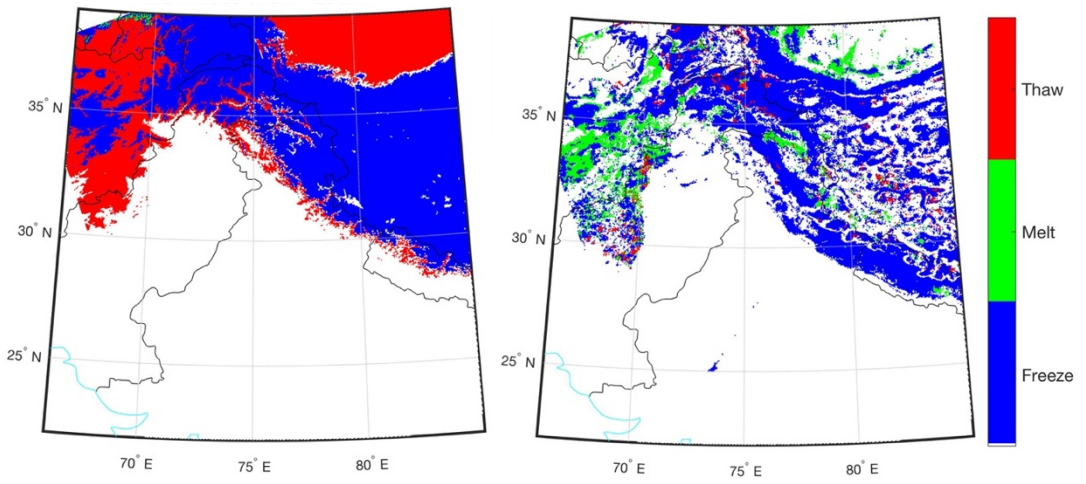

Figure 5.10. Maps of the Indus Basin displaying ASCAT F/M/T (right) and re-gridded LIS F/M/T (left) categorical pixel assignment for Nov 1, 2011.

5.4.2 Temporal Analysis of ASCAT vs. LIS F/M/T

Figures 5.11, 5.12, and 5.13 show time series for one year (2011) for three different locations. Test location-1, Figure 5.11, displays LIS and ASCAT F/M/T agreement for the months of January, May, November and December. Both datasets show almost no snowmelt for this pixel throughout the year. Considering the location of the pixel, presence of snow is expected at least during the main winter months of December, January and February. With the advent of summer, that snow is expected to melt. Therefore, this 'melt' state agreement between the two datasets must be viewed with caution and needs supplementary data from other sources to corroborate the plausibility of no snowmelt.

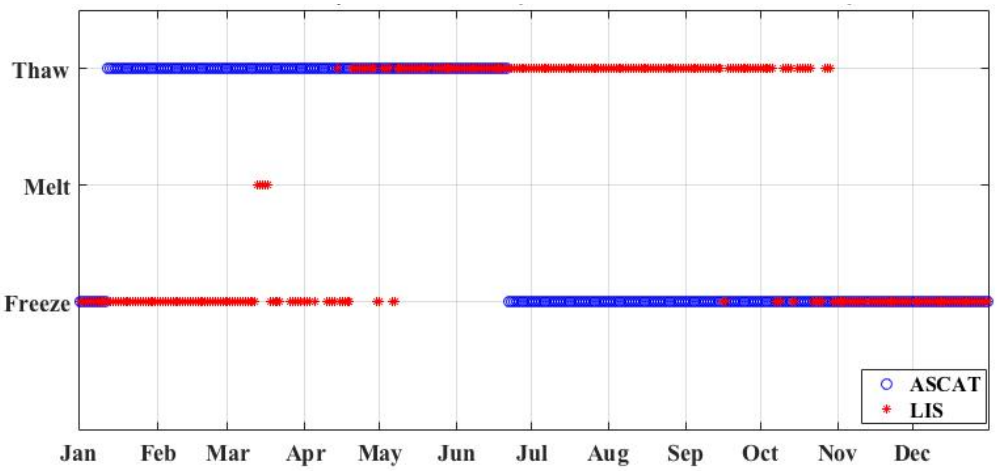

Figure 5.11. Test location-1 (36.1861^{O}N, 71.6222^{O}E) time-series of ASCAT vs. LIS F/M/T pixel state assignment for year: 2011.

The time series for test location-2, in Figure 5.12, highlights the same phenomenon discussed in the spatial analysis for days; May 1 and Aug 1, 2011. The melt and thaw disagreement between the two datasets indicates either inaccurate snow depth estimation in LIS or over-estimation of snow presence in ASCAT F/M/T. This assumption is based on the fact that snowmelt and thaw both indicate an increase in surface moisture. Thus, the ASCAT radar backscatter perceives an increased dielectric constant. In the presence of snow, this condition would be identified as snowmelt, whereas for snow-free areas it would be categorized as thaw.

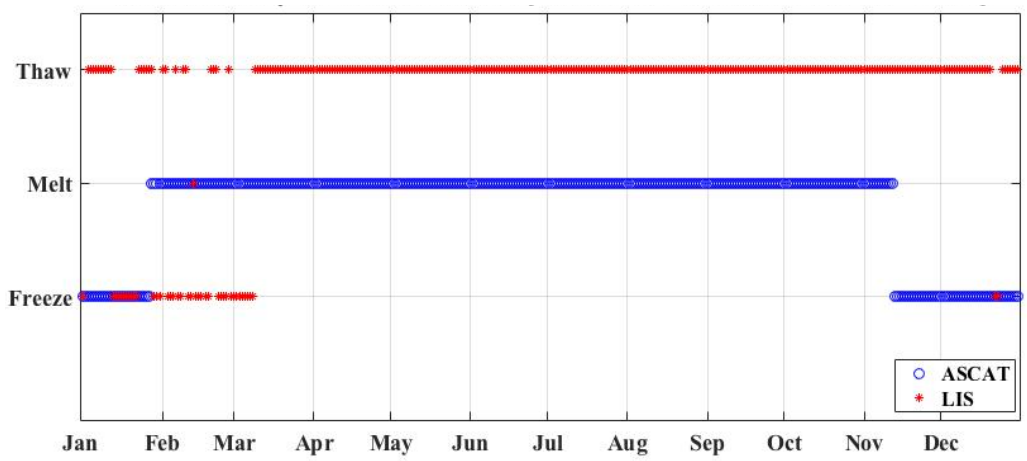

Figure 5.12. Test location-2 (32.5528^{O}N, 67.8278^{O}E) time-series of ASCAT vs. LIS F/M/T pixel state assignment for year: 2011.

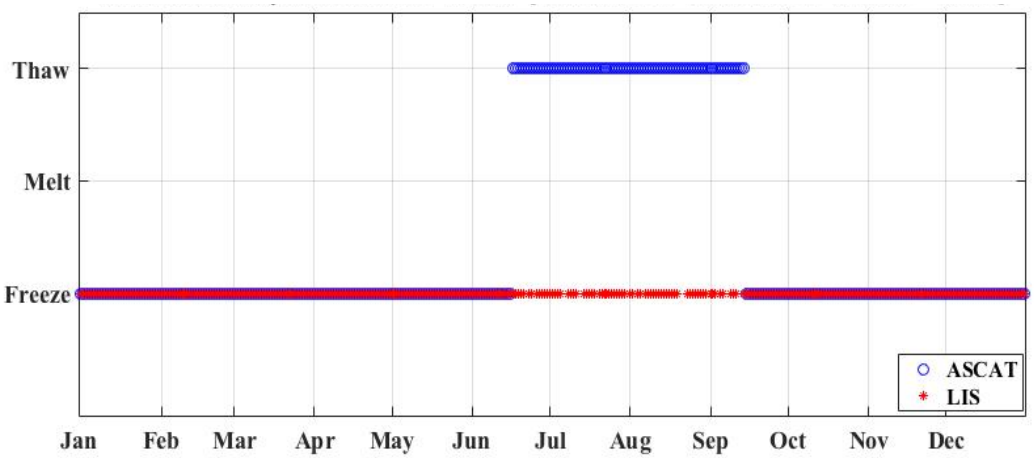

Figure 5.13. Test location-3 (35.4139^{O}N, 77.2944^{O}E) time-series of ASCAT vs. LIS F/M/T pixel state assignment for year: 2011.

The time-series for test location-3 in Figure 5.13 shows better agreement between the two data sets. Divergence between the two products is observed during the summer months of July, August, and September. This could be due to the highly increased temporal variability in surface moisture during these months which ultimately affects the ASCAT daily retrieval accuracy.

In all three time-series a discontinuity (data gaps) is perceived in LIS F/M/T point location time-series. One of the reasons behind this are the various quality checks that are applied on LIS state estimates. For example, during re-gridding if 75% of the

original (smaller) grid cells had no state defined, the re-gridded pixel was left blank. Different percent thresholds were tried such as 90%, 80%, and 60%. 75% seemed to give the best visible conformance between the original grid and the re-gridded dataset.

5.4.3 Statistical Analysis of ASCAT vs. LIS F/M/T

Conformance between the two datasets is evaluated using the following statistical approaches:

- Contingency tables
- Probability of detection

5.4.3.1 Contingency tables

Contingency tables (a.k.a. coincidence matrix or classification matrix) can be used to evaluate the agreement and disagreement between model forecast and observed values [72]. For this study, LIS F/M/T was labeled as forecast and ASCAT F/M/T as the observed values.

Since the goal is to investigate the monthly behavior of the two products, thus a more visually explanatory (graphical) approach was used instead of the conventional IxJ table (where IxJ is equal to the possible combinations of forecast and observation pairs).

5.4.3.1.1 Contingency table of four possible Forecast/ Observation events

Four possible combinations of the forecast/ observation pairs were identified (Table 5.1):

- Yes: Presence of a state
- No: Absence of that state
- YY: The observation identifies presence of a state and the model forecasts presence of that state as well.
- NY: The observation identifies absence of a state while the model forecasts presence of that state.
- YN: The observation identifies presence of a state whereas the model forecasts absence of that state (meaning presence of any other state).

- NN: The observation identifies absence of a state and the model also forecasts absence of that state

Table 5.1. '2x2' Contingency table of four possible Forecast/ Observation events

		Observation (ASCAT F/M/T)		
		Yes	*No*	
Forecast (LIS F/M/T)	*Yes*	**YY**	**NY**	YY + NY
	No	**YN**	**NN**	YN + NN
		YY + YN	NY + NN	Total no. of events (n)

Separate analysis was carried out for the three individual states, i.e., monthly contingency table values were calculated for each state separately as seen in Figures 5.14, 5.15, and 5.16. Due to the gaps in LIS data, the total number of events, n, varies for each month (ASCAT data-holes are constant). The total number of events for each month consists of all the points in the domain that had defined state values for both datasets during that month. In order to maintain consistency and to be able to compare the monthly values, relative frequency values are used instead of event counts. To obtain the relative frequency, the individual event counts are divided by the total number of events for each month. This way relative frequency values are achieved for each month and each state that are comparable to each other.

In Figure 5.14 the monthly change in agreement is observed between the two products for 'freeze' state. The highest 'YY' events occur in February while there are no discernable 'YY' events during the summer months (June, July, and August). This is expected as the frozen land condition is most prevalent during winter and lowest in

the summer months. Since these are relative frequency values, thus the total sum of the four events for each month should equal '1'.

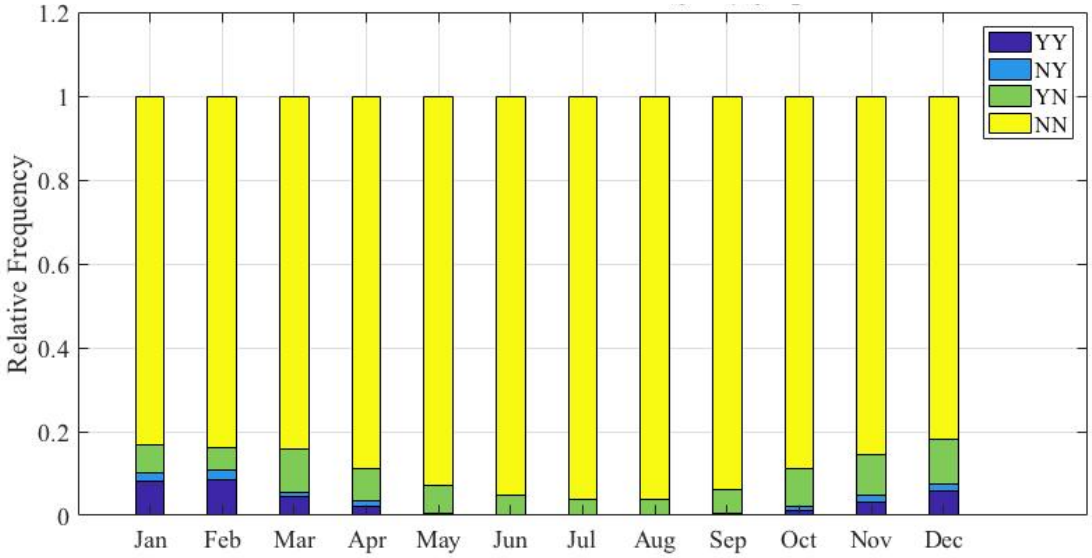

Figure 5.14. Contingency table relative frequency values of 'freeze' state for the Indus Basin (Year: 2011).

Figure 5.15 presents the monthly occurrence of the four O/F events for 'melt' state. The 'NN' event relative frequency values decrease as the summer months approach and both products start detecting presence of snowmelt. There are no substantial 'YY' or 'NY' events throughout the year. This seems to indicate that there is a comparative dearth of pixels that are identified as 'melt' by the LIS F/M/T product. The highest disagreement ('YN' events) is also observed during the summer months. This is attributed to the fact that 'melt' instances primarily occur during that season and hence the possibility of disagreement is also highest during that period.

Figure 5.16 presents results similar to Figure 5.15 for the 'NN' event. Combined relative frequency of 'NY' and 'NN' events indicate that the ASCAT F/M/T product identifies comparatively few 'thaw' occurrences. Observing the 'NY' value alone, higher pixel categorization of 'thaw' by LIS F/M/T relative to ASCAT F/M/T is realized.

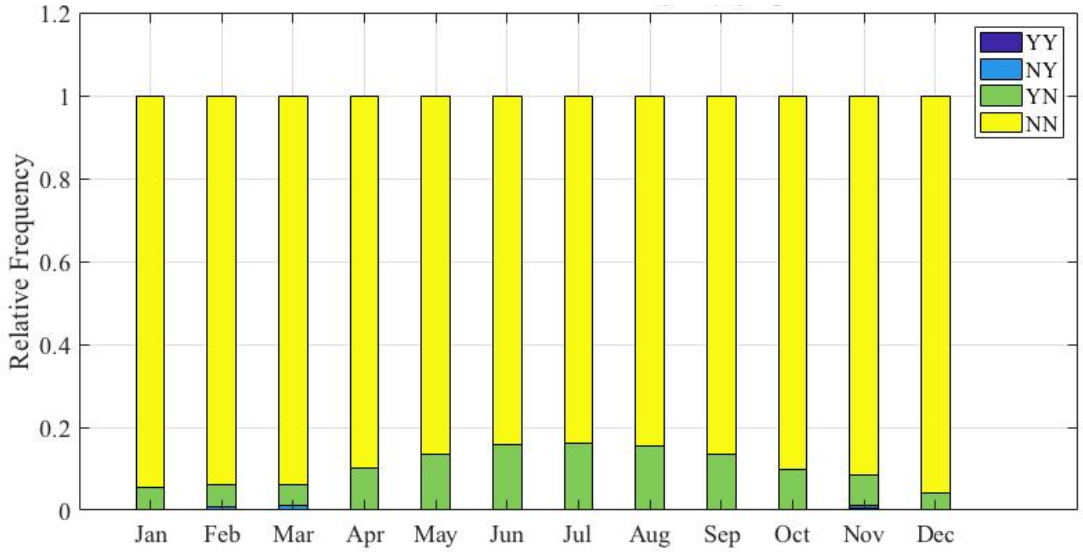

Figure 5.15. Contingency table relative frequency values of 'melt' state for the Indus Basin (Year: 2011).

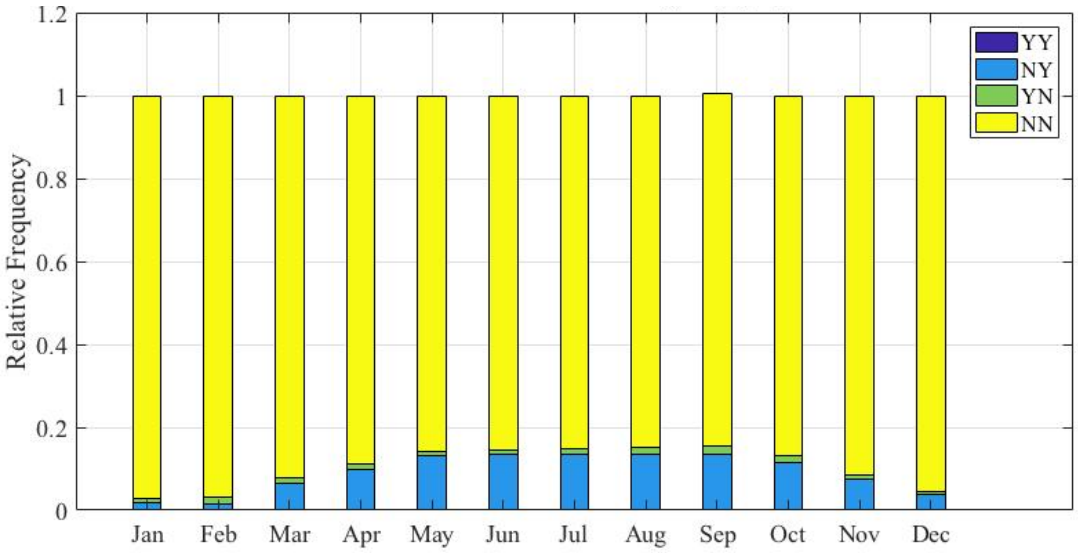

Figure 5.16. Contingency table relative frequency values of 'thaw' state for the Indus Basin (Year: 2011).

From the Figures 5.14, 5.15, 5.16, a first-order estimation is deduced, i.e., LIS F/M/T detects higher instances of 'thaw' than ASCAT F/M/T, whereas the ASCAT F/M/T detects a larger number of 'melt' occurrences than the LIS F/M/T product.

5.4.3.1.2. Contingency table of nine possible ASCAT vs. LIS F/M/T events

Continuing along a similar vein, the contingency table theory is approached from a slightly different perspective. Instead of defining relative event frequencies for each individual state, a 3x3 table (Table 5.3) was developed that compared relative presence and absence of all the three states. This helped hone in on the states that showed most ASCAT vs. LIS F/M/T agreement and disagreement, and gave insight into the accuracy of corresponding LIS modeled states (such as snow depth and air temperature).

Table 5.2. '3x3' Contingency table of nine possible ASCAT vs. LIS F/M/T events.

		ASCAT F/M/T		
		Freeze	*Melt*	*Thaw*
LIS F/M/T	*Freeze*	FF	MF	TF
	Melt	FM	MM	TM
	Thaw	FT	MT	TT

To facilitate the comparison, relative frequency values are calculated for each event, such that the summation of all (events) relative frequencies equals 1. In Figure 5.17, the highest relative frequency value belongs to 'MT' event. 'MT' is when ASCAT F/M/T categorizes 'melt' happening whereas LIS F/M/T indicates occurrence of 'thaw'. This means that 46% of all (nine) events that occur are categorized as 'thaw by LIS F/M/T while melt by ASCAT F/M/T'. Recalling the algorithm used for characterizing 'thaw' for LIS F/M/T, this MT percentage disagreement could indicate possible inaccurate 'snow depth' estimation by LIS.

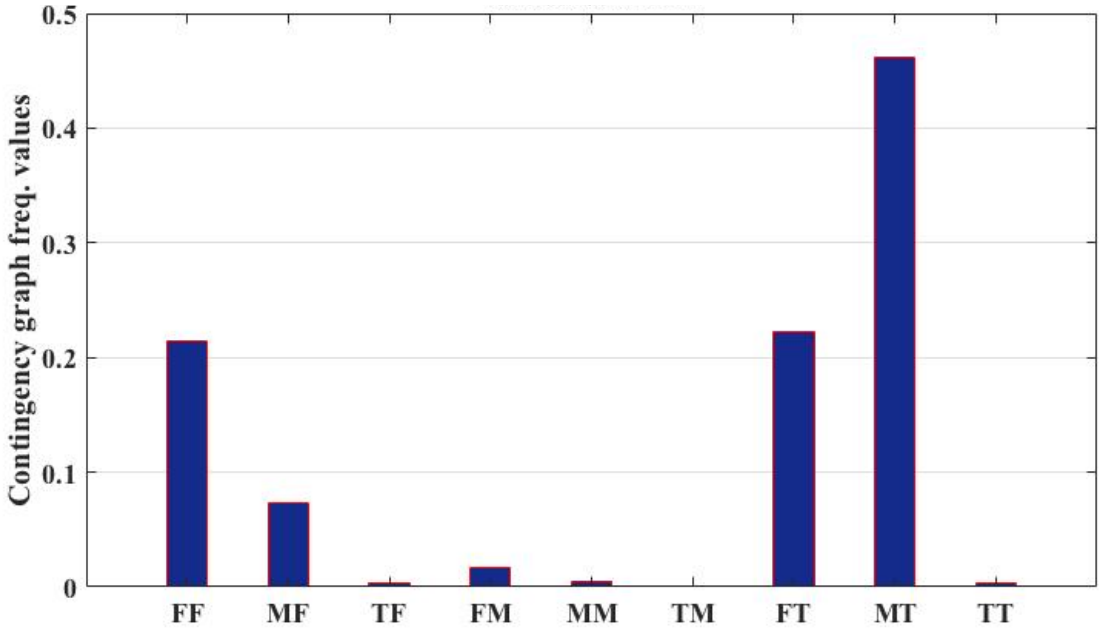

Figure 5.17. Contingency table relative frequency values of all ASCAT vs. LIS F/M/T event pairs for the Indus Basin (Year: 2011).

Interesting to note is the almost '0' percentage of TM events (ASCAT F/M/T declares thaw happening while LIS F/M/T indicates melt). Relative frequency of 'FT' event occurrence is also comparatively noticeable (22.5%). Combining 'MT' and 'FT' relative frequencies, it seems LIS F/M/T is predicting considerably higher occurrence of thaw as compared to ASCAT F/M/T. It could be due to deficiencies in the algorithm (Figure 5.2) used for developing the LIS F/M/T product or it could indicate possible inaccurate LIS modeling of the states used for characterizing 'thaw'. Highest agreement between the two datasets is observed for 'freeze', 22% of all (nine) event occurrences.

The relative frequency values in Figure 5.17 are representative of the whole year (2011). For a further detailed analysis, monthly contingency table (Table 5.2) values were calculated. Figure 5.18 displays monthly contingency graph values for January, March, August, and November 2011. These example months were selected considering the seasonal variation and climatology of the Indus Basin. Monthly results differ in the relative frequency magnitudes of all events yet still convey the same 'MT' and 'FT' disagreement between the two datasets as seen in Figure 5.17.

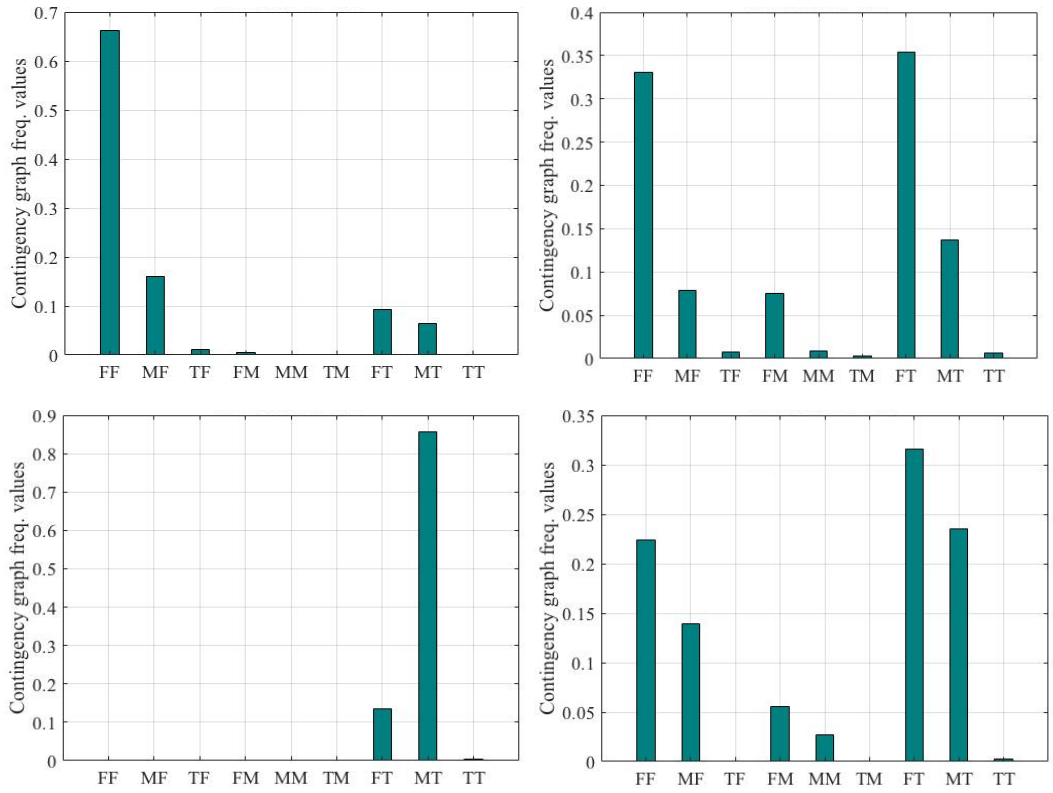

Figure 5.18. Monthly contingency table (Table 5.2) relative frequency values of all ASCAT vs. LIS F/M/T event pairs for the Indus Basin; January (top-left), March (top-right), August (bottom-left), November (bottom-right) – Year: 2011.

5.4.3.2 Probability of Detection (POD)

Also known as prefigurance, POD is the fraction of 'event being correctly forecast to occur' over the total events of its actually having occurred (as determined by observations). In this case, the formula used to calculate POD includes the 'YY' and 'YN' values from the contingency table (Table 5.1) values. It is given by [72]:

$$POD = \frac{YY}{YY + YN} \tag{5.1}$$

For this statistical value, it was assumed that ASCAT F/M/T are 'accurately observed measurement' values and POD is the probability of LIS F/M/T forecasting the same state as identified by the ASCAT F/M/T observation. A POD of '1' means perfect forecast while a '0' value signifies poor forecast.

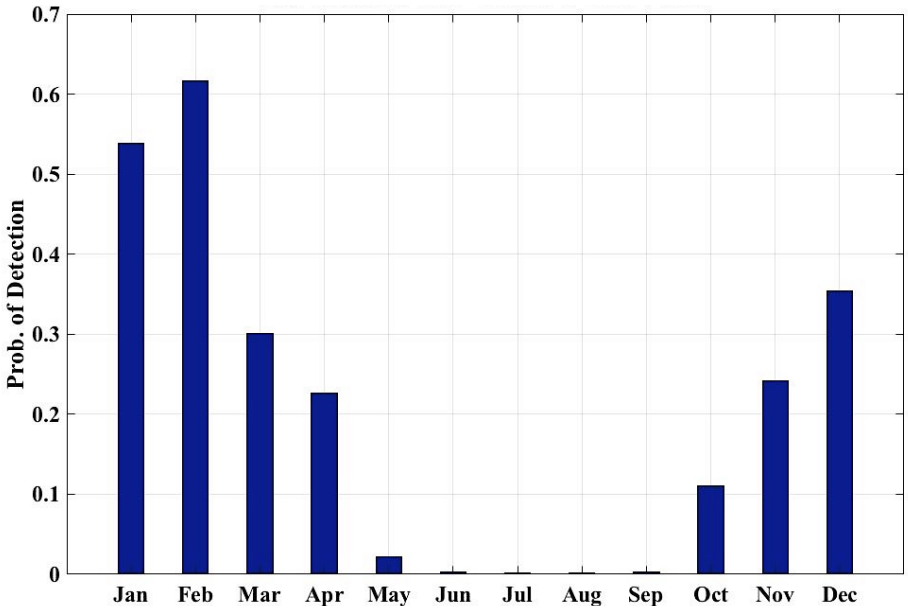

Figure 5.19. Monthly 'freeze state POD, ASCAT (observed) vs. LIS F/M/T (model forecast), for the Indus Basin (Year : 2011).

In Figure 5.19, highest POD value was observed for Feb while almost '0' values were calculated for June, July, August, and September. Low values coincide with the summer months when the total number of pixels undergoing 'freeze' was significantly reduced.

Highest value of POD observed in Figure 5.20 is 0.04575 (Nov 2011). This means the highest probability of LIS F/M/T detecting melt state in any month throughout the year 2011 (as detected by ASCAT F/M/T) over the Indus Basin is 4.57%. This percentage highlights the poor 'melt' detection by LIS F/M/T. Figure 5.21 presents POD values for 'thaw'. These values are comparatively higher than melt but still too trivial to indicate any significant correct thaw detection by LIS F/M/T.

POD is useful in accessing the agreement between the two data-sets with respect to each state. Figures 5.19, 5.20, and 5.21 suggest that for 'freeze' state, the agreement between the two products is the highest while the greatest difference is perceived in 'melt' state. This is partly due to the scarce identification of 'melt' in LIS F/M/T. POD values for 'thaw' are quite low and represent the discord between the datasets regarding occurrence of 'thaw'. One of the reasons is the comparatively high 'thaw' detection in LIS F/M/T, as discussed in Section 5.4.3.1.2.

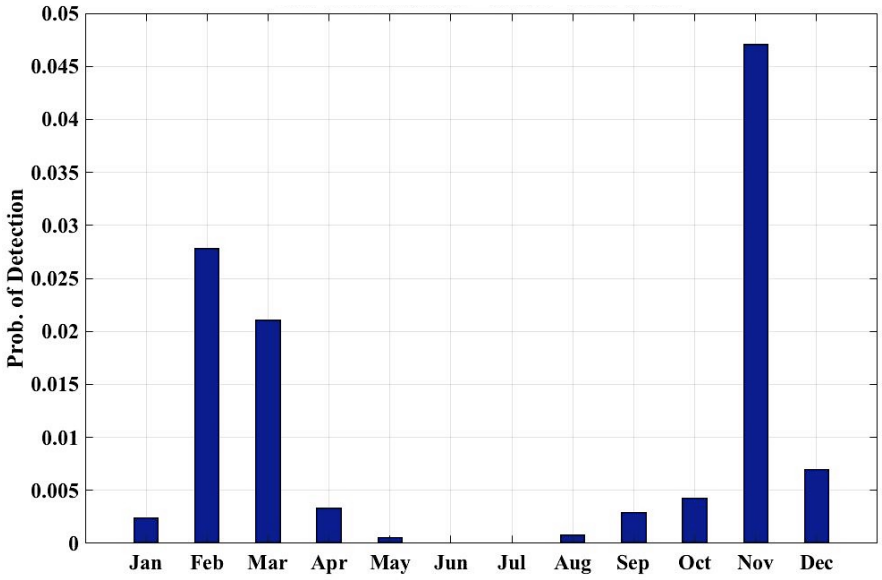

Figure 5.20. Monthly 'melt' state POD, ASCAT (Obs.) vs. LIS F/M/T (Model forecast), for the Indus Basin (Year : 2011).

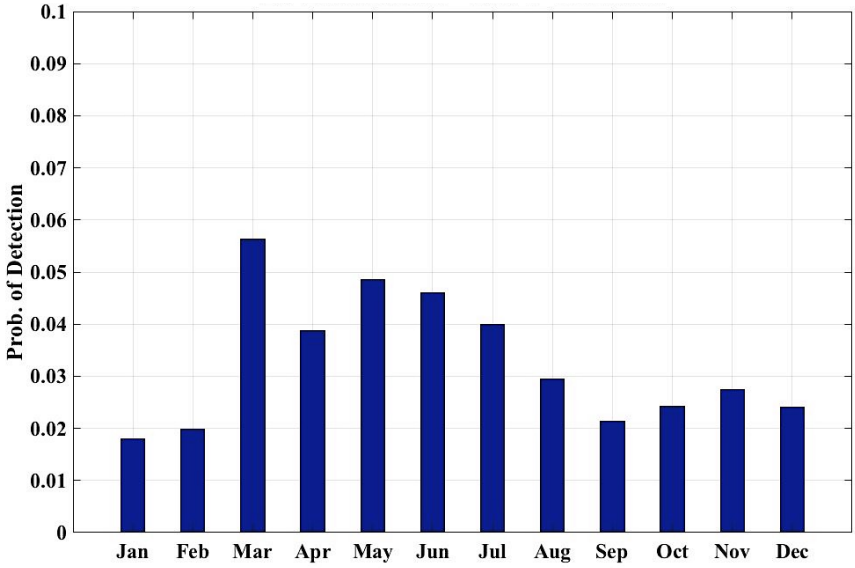

Figure 5.21. Monthly 'thaw' state POD, ASCAT (Obs.) vs. LIS F/M/T (Model forecast), for the Indus Basin (Year : 2011)

5.5. Limitations

Some of the limitations that constricted the effectiveness of this comparative analysis are:

- Both datasets are qualitative in nature and thus inhibit error analysis.

- According to the algorithm defined for LIS F/M/T, some pixels in the area of importance within the domain remain uncategorized. For example, pixels that have air temperature <= 273.15K and soil temperature > 273.15 K are not assigned any category. For special cases, similar to this, greater in-depth knowledge regarding the ASCAT F/M/T retrieval algorithm is required.
- The LIS re-gridding technique utilized is very basic and can lead to incorrect categorization in some cases. A more sophisticated re-gridding technique is expected to improve the accuracy of the LIS F/M/T product.
- Coarse grid resolution affects the accuracy of the ASCAT measurements and subsequent F/M/T retrieval.
- Both datasets are in their preliminary development stages. Further refinement of the respective algorithms is expected to improve the accuracy of both.
- Statistical methods described above ignore the serial correlation in both datasets. Accounting for serial correlation will decrease the sample size, and is hence expected to affect the outcome of the various statistics utilized here.

Chapter 6: Conclusion

In this study, NASA's Land Information System (LIS) was used to model the hydrologic cycle over the Indus basin. The ability of support vector machines (SVM), a machine learning technique, to predict passive microwave brightness temperature (spectral difference) as a function of LIS land surface model output was explored through a sensitivity analysis. Multi-frequency and multi-polarization passive microwave brightness temperatures measured by the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) over the Indus basin were used as training targets during the SVM training process. Normalized sensitivity coefficients (NSC) were then computed to assess the relative sensitivity of a well-trained SVM to each LIS-modeled state variable. Special focus was targeted towards SVM prediction sensitivity to LIS estimated snow water equivalent (SWE) for snow-covered areas in our domain. An assessment of LIS modeled states was carried out through a comparative analysis with a satellite based (ASCAT) dataset.

6.1. Sensitivity Analysis of a well-trained SVM to each LIS-Modeled State

The NSC values obtained for each LIS input state were representative of a number of concurrent and interacting physical processes, high cross-correlation between the LIS input states and effect of location specific parameters such as dense vegetation and glaciers.

According to the results we obtained from our sensitivity analysis, SVM ΔT_b prediction conforms with the known first-order physics. LIS input states that are directly linked to physical temperature like snow temperature, air temperature, and vegetation temperature generally displayed large absolute NSCs throughout the domain, signifying higher relative sensitivity.

Snow temperature (of the top-layer of the snow-pack) exhibited the largest sensitivity coefficient magnitudes. Near surface soil temperature displayed almost zero sensitivity. It can thus be concluded that SVM predicted ΔT_b (18.7V − 36.5V) is

insensitive to soil temperature. One possible reason could be the relatively low variability of soil temperature compared to other land surface parameters.

According to all the spatial and temporal analyses of NSCs carried out, the general order of decreasing relative sensitivity of SVM predicted ΔT_b to the LIS input states is:

1. Snow temperature
2. Air temperature
3. Vegetation temperature
4. Snow density
5. SWE
6. SLWC
7. LAI
8. Top-layer soil temperature

Bottom-layer snow temperature values did not exist for most locations and thus had no NSC values. Soil moisture was eliminated from the analyses due to the erroneous estimation by LIS under frozen soil conditions.

SWE had smaller NSC magnitudes compared to some of the other 9 LIS input states. Decreasing the total number of LIS input states (from '10' to '4') used for SVM training and prediction resulted in an increase in the relative sensitivity of SWE.

6.2. Assessment of LIS Modeled States

Accuracy of the LIS modeled geophysical states was assessed by comparing a LIS derived Freeze/Melt/Thaw dataset with the Advanced Scatterometer (ASCAT) based Freeze/Melt/Thaw product. In this assessment, the ASCAT based product was considered as 'observed' data whereas the LIS based F/M/T dataset was regarded as model forecast. Various spatial, temporal, and statistical analyses were carried out to study the agreement between the two datasets. The comparisons detailed in Chapter 5 suggest some common themes:

- ASCAT F/M/T overestimates freeze state condition in some areas due to its relaxed defining criteria for freeze. This is especially noticeable in the lower latitudes, around 32^{O}N, within the basin boundaries.
- ASCAT F/M/T displays greater data continuity, having static data-holes while LIS F/M/T has non-uniform data gaps.
- ASCAT F/M/T demonstrates less variability compared to LIS F/M/T which is significantly more (spatially and temporally) dynamic. This relates back to the ASCAT bearing MetOp satellite's overpass of 3 days. Hence, to develop a daily temporal resolution product some interpolations are required.
- Highest agreement between the datasets is realized for 'freeze' state.
- Considering the best agreement for 'freeze' state detection, it can be deduced that LIS modeled geophysical states (e.g. snow depth, air temperature) that are utilized in defining freeze (in the LIS algorithm) seem to have better accuracy.
- Based on the algorithm used, discrepancy in the datasets regarding 'melt' occurrence could be due to underestimation of snow liquid water content by LIS.
- LIS F/M/T detects higher instances of 'thaw' than ASCAT F/M/T, whereas the ASCAT F/M/T detects a larger number of 'melt' occurrences than the LIS F/M/T product.
- ASCAT F/M/T seems to underestimate thaw conditions, especially during the summer months. Considering the radar backscatter analysis principles, this could be related to the rapidly varying (at time interval < daily temporal resolution) soil moisture during those months.

6.3. Future Work

6.3.1. Data assimilation to improve SWE estimation in HMA.

This study was carried out to identify the relative importance of LIS input states to SVM brightness temperature (spectral difference) prediction. It will further contribute towards selection of LIS input states used in brightness temperature (spectral difference) prediction and assimilation to improve SWE estimation in high mountain

Asia. Previous studies [73] have successfully explored the usability of SVM as the observation operator in an assimilation framework. A similar SWE data assimilation framework will be developed for the high mountain Asia region.

6.3.2. Improvement of the ASCAT vs. LIS F/M/T Comparison

Further effort needs to be targeted towards overcoming the limitations identified in Chapter 5. Taking into account the serial correlation of ASCAT and LIS F/M/T products is likely to affect the comparison between the two datasets. Considering that both the datasets used were in their preliminary stages, the various analyses described in Chapter 5 need to be repeated for the newer and improved versions of LIS land surface geophysical state estimation and the ASCAT freeze/melt/thaw product.

6.3.3. Data assimilation of ASCAT Freeze/Melt/Thaw

One new avenue that can be explored is the possibility of ASCAT freeze/melt/thaw data assimilation in the LIS model estimates and analyses of the resulting change in LIS states.

6.3.4. Assessment of LIS land surface modeling accuracy with other datasets

The accuracy of LIS modeled states was assessed here using a LIS derived product. For an improved accuracy analysis, ground observations of the various states should be compared directly with the LIS model state estimates (ground observations were not available at the time this study was carried out and have been acquired recently).

Bibliography

[1] H. D. Pritchard, "Asia's glaciers are a regionally important buffer against drought," *Nature,* vol. 545, pp. 169-174, May 2017.

[2] J. Xu, R. E. Grumbine, A. Shrestha, M. Eriksson, X. Yang, Y. U. N. Wang and A. Wilkes, "The melting Himalayas: cascading effects of climate change on water, biodiversity, and livelihoods," *Conservation Biology,* vol. 23, no. 3, pp. 520-530, 2009.

[3] S. B. Kapnick, T. L. Delworth, M. Ashfaq, S. Malyshev and P. C. D. Milly, "Snowfall less sensitive to warming in Karakoram than in Himalayas due to a unique seasonal cycle," *Nature Geoscience ,* vol. 7, p. 834–840, 2014.

[4] W. K. M. Lau, M.-K. Kim, K.-M. Kim and W.-S. Lee, "Enhanced surface warming and accelerated snow melt in the Himalayas and Tibetan Plateau induced by absorbing aerosols," *Environmental Research Letters,* vol. 5, no. 2, 2010.

[5] K. Rühland, N. R. Phadtare, R. K. Pant, S. J. Sangode and J. P. Smol, "Accelerated melting of Himalayan snow and ice triggers pronounced changes in a valley peatland from northern India," *Geophysical Research Letters,* vol. 33, no. 15, 2006.

[6] B. A. Forman and Y. Xue, "Machine learning predictions of passive microwave brightness temperature over snow-covered land using the special sensor microwave imager (SSM/I)," *Physical Geography,* vol. 38, no. 2, pp. 176-196, 2017.

[7] B. A. Forman and R. H. Reichle, "Using a Support Vector Machine and a Land Surface Model to Estimate Large-Scale Passive Microwave Brightness Temperatures Over Snow-Covered Land in North America," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 8, no. 9, pp. 4431-4441, Sep 2015.

[8] B. A. Forman, R. H. Reichle and C. Derksen, "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,* vol. 52, no. 1, pp. 235-248, 2014.

[9] S. V. Kumar, C. D. Peters-Lidard, Y. Tian, P. R. Houser, J. Geiger, S. Olden, L. Lighty, J. Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. Wood and J. Sheffield, "Land information system: An interoperable framework for high resolution land surface modeling," *Environmental Modelling & Software,* vol. 21, pp. 1402-1415, 2006.

[10] M. B. Ek, K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno and J. D. Tarpley, "Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model," *Journal of Geophysical Research,* vol. 108, no. D22, 2003.

[11] R. Gelaro, W. McCarty, M. Suárez, R. Todling, A. Molod, L. Takacs, C. Randles, A. Darmenov, M. Bosilovich, R. Reichle and K. Wargan, "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)," *Journal of Climate,* vol. 30, no. 14, pp. 5419-5454, 2017.

[12] R. T. Hitchcock, Radio-frequency and Microwave Radiation, John Wiley & Sons, Inc., 2001.

[13] S. Kumar and S. Shukla, Concepts and Applications of Microwave Engineering, PHI Learning Pvt. Ltd.., 2014.

[14] K.-N. Liou, An introduction to atmospheric radiation, Academic Press, 2002.

[15] E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan and S. V. Nghiem, "Soil moisture retrieval from AMSR-E," *IEEE transactions on Geoscience and remote sensing* , vol. 41, no. 2, pp. 215-229, 2003.

[16] R. Kelly, "The AMSR-E snow depth algorithm: Description and initial results," *Journal of the Remote Sensing Society of Japan,* vol. 29, no. 1, pp. 307-317, 2009.

[17] H. J. Zwally and P. Gloersen, "Passive microwave images of the polar regions and research applications," *Polar Record* , vol. 18, no. 116, pp. 431-450, 1977.

[18] A. Chang, J. Foster and D. Hall, "NIMBUS- 7 SMMR derived global snow cover parameters," *Annals of Glaciology,* vol. 9, pp. 39-44, 1987.

[19] COMET® , " COMET PROGRAM, University Corporation for Atmospheric Research (UCAR)," 1997-2017 . [Online]. Available: http://meted.ucar.edu/ . [Accessed 6 March 2018].

[20] A. Chang, J. Foster, D. Hall, A. Rango and B. Hartline, "Snow water equivalent estimation by microwave radiometry," *Cold Regions Science and Technology,* vol. 5, no. 3, pp. 259-267, 1982.

[21] NASA, "AMSR-E," 2 March 2018. [Online]. Available: https://aqua.nasa.gov/content/amsr-e. [Accessed 7 March 2018].

[22] Science@NASA, "Water-Witching From Space," [Online]. Available: https://science.nasa.gov/science-news/science-at-nasa/2000/ast23may_1. [Accessed 7 March 2018].

[23] D. G. B. M. J. Long, "Optimum image formation for spaceborne microwave radiometer products," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 54, no. 5, pp. 2763-2779, 2016.

[24] W.-C. Hong, "Rainfall forecasting by technological machine learning models," *Applied Mathematics and Computation,* vol. 200, no. 1, pp. 41-57, June 2008.

[25] T. Petty and P. Dhingra, "Streamflow hydrology estimate using machine learning (SHEM)," *Journal of the American Water Resources Association,* vol. 54, no. 1, pp. 55-68, 2017.

[26] S. Ahmad, A. Kalra and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Advances in Water Resources,* vol. 33, no. 1, pp. 69-80, January 2010.

[27] M. Tedescoa, J. Pulliainen, M. Takalab, M. Hallikainen and P. Pampaloni, "Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data," vol. 90, p. 76–85, 2004.

[28] M. Sugiyama, Introduction to Statistical Machine Learning, Morgan Kaufmann, 2015, p. 534.

[29] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing,* vol. 14, no. 3, pp. 199-222, 2004.

[30] E. Viste and A. Sorteberg, "Snowfall in the Himalayas: an uncertain future from a little-known past," *The Cryosphere,* no. 9, pp. 1147-1167, 2015.

[31] S. Kumar, C. Peters-Lidard, K. R. Arsenault, A. Getirana, D. Mocko and Y. Liu, "Quantifying the Added Value of Snow Cover Area Observations in Passive Microwave Snow Depth Data Assimilation," *Journel of Hydrometeorology,* vol. 16, pp. 1736-1741, 2015.

[32] D. K. Hall, G. A. Riggs, V. V. Salomonson, N. E. DiGirolamo and K. J. Bayr, "MODIS snow-cover products," *Remote sensing of Environment,* vol. 83, no. 1, pp. 181-194, 2002.

[33] T. H. Painter, D. F. Berisford, J. W. Boardman, K. J. Bormann and J. S. Deems, "The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo," *Remote Sensing of Environment,* vol. 184, pp. 139-152, October 2016.

[34] L. J. P. Tsang, D. Liang, Z. Li, D. W. Cline and Y. Tan, "Modeling active microwave remote sensing of snow using dense media radiative transfer (DMRT) theory with multiple-scattering effects," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 4, pp. 990-1004, 2007.

[35] R. Kelly, "The AMSR-E Snow Depth Algorithm: Description and Initial Results," *Journal of The Remote Sensing Society of Japan,* vol. 29, no. 1, pp. 307-317, 2009.

[36] R. E. J. Kelly and A. T. C. Chang, "Development of a passive microwave global snow depth retrieval algorithm for Special Sensor Microwave Imager (SSM/I)

and Advanced Microwave Scanning Radiometer-EOS (AMSR-E) data," *Radio Science,* vol. 38, no. 4, p. 8076, 2003.

[37] R. Armstrong, A. Chang, A. Rango and E. Josberger, "Snow depths and grain-size relationships with relevance for passive microwave studies," *Annals of Glaciology,* vol. 17, no. 1, pp. 171-176, 1993.

[38] A. E. Walker and B. E. Goodison, "Discrimination of a wet snow cover using passive microwave satellite data," *Annals of Glaciology,* vol. 17, pp. 307-311, 1993.

[39] T. Che, L. Dai, X. Zheng, X. Li and K. Zhao, "Estimation of snow depth from passive microwave brightness temperature data in forest regions of northeast China," *Remote Sensing of Environment ,* vol. 183, p. 334–349, 2016.

[40] E. K. Smith, "Centimeter and millimeter wave attenuation and brightness temperature due to atmospheric oxygen and water vapor," *Radio Science,* vol. 17, no. 6, pp. 1455-1464, December 1982.

[41] D. G. Long and M. J. Brodzik, "Optimum Image Formation for Spaceborne Microwave Radiometer Products," *IEEE Transactions on Geoscience and remote sensing ,* vol. 54, no. 5, pp. 2763-2779, May 2016.

[42] C. D. Peters-Lidard, P. R. Houser, Y. Tian, S. V. Kumar, J. Geiger, S. Olden and L. Lighty, "High-performance Earth system modeling with NASA/GSFC's Land Information System," *Innovations in Systems and Software Engineering,* vol. 3, no. 3, pp. 157-165, 2007.

[43] G. Niu, Z. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari and Y. Xia, "The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements," *Journal of Geophysical Research ,* vol. 116, no. D12109, 2011.

[44] E. Alpaydin, Introduction to machine learning, MIT press, 2014.

[45] N. Chen, L. Wencong , J. Yang and G. Li, Support vector machine in chemistry, World Scientific Pub, 2004.

[46] V. Vapnik and A. Chervonenkis, Pattern Recognition Theory, Statistical Learning Problems, Moscow: Nauka, 1974.

[47] V. Vapnik, Estimation of Dependences Based on Empirical Data, Secaucus, NJ: Springer-Verlag New York, Inc, 1982.

[48] V. Vapnik, Nature of Statistical Learning Theory, Springer-Verlag New York, Inc, 1995, p. 314.

[49] S. R. Kulkarni and G. Harman, "Statistical Learning Theory: A Tutorial," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 3, no. 6, pp. 543-556, 2011.

[50] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[51] R. Fletcher, Practical methods of optimization, John Wiley & Sons, 2013.

[52] V. Kecman, "Support Vector Machines–An Introduction," in *Support Vector Machines: Theory and Applications (Studies in Fuzziness and Soft Computing)*, vol. 177, W. Lipo, Ed., Springer, Berlin, Heidelberg, 2005, pp. 1-47.

[53] H. W. Kuhn and A. W. Tucker, in *Proceedings of 2nd Berkeley Symposium*, Berkeley, 1951.

[54] V. Kecman, "Support vector machines–an introduction.," in *Support vector machines: theory and applications*, Berlin, Heidelberg: Springer, 2005, pp. 1-47.

[55] V. N. Vapnik, "An Overview of Statistical Learning Theory," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing ,* vol. 10, no. 5, pp. 988-999, September 1999.

[56] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology,* vol. 2, no. 3, p. 27, 2011.

[57] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Microsoft Research, 1998.

[58] ICIMOD, "Indus Basin Initiative," ICIMOD, 2018. [Online]. Available: http://www.icimod.org/indus. [Accessed 31 Jan 2018].

[59] M. Cheemaa and W. Bastiaanssena, "Land use and land cover classification in the irrigated Indus Basin using growth phenology information from satellite data to support water management analysis," *Agricultural Water Management,* vol. 97, p. 1541–1552, 2010.

[60] C. W. Hsu, C. C. Chang and C. J. Lin, "A practical guide to support vector classification," 2003.

[61] Y. Xue and B. Forman, "Comparison of passive microwave brightness temperature prediction sensitivities over snow-covered land in North America using machine learning algorithms and the Advanced Microwave Scanning Radiometer," *Remote Sensing of Environment,* vol. 170, pp. 153-165, 2015.

[62] D. Mattera and S. Haykin, "Support vector machines for dynamic reconstruction of a chaotic system," in *Advances in kernel methods*, MIT Press, 1999, pp. 211-241.

[63] R. McCuen, Modeling hydrologic change: statistical methods, Vancouver: CRC press, 2002.

[64] R. Willis and W. W. Yeh, Groundwater systems planning and management, Old Tappan , New Jersey: Prentice Hall Inc., 1987.

[65] H. Monod, C. Naud and D. Makowski, "Uncertainty and sensitivity analysis for crop models," in *Working with dynamic crop models: Evaluation, analysis, parameterization, and applications*, vol. 4, 2006, pp. 55-100.

[66] A. Etienne, M. Génard, P. Lobit and C. Bugaud, "Modeling the vacuolar storage of malate shed lights on pre- and post-harvest fruit acidity," *BMC Plant Biology,* vol. 14, no. 1, p. 310, 2014.

[67] J. Kirch, C. Thomaseth, A. Jensch and N. E. Radde, "The effect of model rescaling and normalization on sensitivity analysis on an example of a MAPK pathway model," *EPJ Nonlinear Biomedical Physics ,* vol. 4, no. 1, p. 3, 2016.

[68] GLIMS and NSIDC, " Global Land Ice Measurements from Space glacier database," 2013. [Online]. Available: DOI:10.7265/N5V98602.

[69] D. S. Early and D. G. Long, "Image Reconstruction and Enhanced Resolution Imaging from Irregular Samples," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 39, no. 2, pp. 291-302, 2001.

[70] N. Steiner and M. Tedesco, "A wavelet melt detection algorithm applied to enhanced-resolution scatterometer data over Antarctica (2000–2009)," *The Cryosphere,* vol. 8, p. 25–40, 2014.

[71] M. Tedesco, W. Abdalati and H. J. Zwally, "Persistent surface snowmelt over Antarctica (1987–2006) from 19.35 GHz brightness temperatures," *Geophysical Research Letters,* vol. 34, 2007.

[72] D. S. Wilks, Statistical methods in the atmospheric sciences, vol. 100, Academic press, 2011.

[73] Y. Xue and B. A. Forman, "Integration of satellite-based passive microwave brightness temperature observations and an ensemble-based land data assimilation framework to improve snow estimation in forested regions," in *Geoscience and Remote Sensing Symposium (IGARSS)*, 2017.