ABSTRACT

Title of Thesis:	ANALYSIS OF THREE DIFFERENT MACHINE LEARNING ALGORITHMS FOR SWE ESTIMATION OVER WESTERN COLORADO USING SPACE-BASED PASSIVE MICROWAVE RADIOMETRY
	Bincheng Yu, Master of Science, 2022
Thesis Directed By:	Associate Professor, Barton A. Forman Department of Civil and Environmental Engineering

This study compares the performance of three different machine learning algorithms used for snow water equivalent (SWE) estimation. Inputs to these algorithms include passive microwave (PMW) brightness temperature (Tb) observations at 10.65 GHz, 18.7 GHz, and 36.5 GHz at both vertical and horizontal polarization as collected by the Advanced Microwave Scanning Radiometer (AMSR-2). The three algorithms include: 1) support vector machine (SVM) regression, 2) long short-term memory (LSTM) networks, and 3) Gaussian process (GP) regression. In-situ SWE measurements from the SNOTEL network collected across western Colorado is used as the training "targets" during the training procedure. The performance of the algorithms is evaluated using a number of different metrics including, but not limited to correlation coefficient, mean

square error (MSE), and bias. The evaluation is conducted over a range of different elevations and different land cover classifications in order to assess algorithm performance across a broad range of snowpack conditions. Preliminary results suggest the LSTM algorithm is computationally more efficient during the training process as compared to the other algorithms, yet yields a similar level of performance. Some limitations, however, have been found in the study, including poor performance during deep snow conditions, which is likely related to signal “saturation” within the PMW Tb’s used during the supervised training process. Additionally, algorithm performance is strongly dependent on the amount of training data such that too little training data results in poor performance by the algorithm at successfully reproducing inter-annual variability. The strengths and limitations of these different machine learning algorithms for snow mass estimation will be discussed.

ANALYSIS OF THREE DIFFERENT MACHINE LEARNING ALGORITHMS FOR SWE ESTIMATION OVER WESTERN COLORADO USING SPACE-BASED PASSIVE MICROWAVE RADIOMETRY

by

Bincheng Yu

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
2022

Advisory Committee:

Associate Professor	Barton A. Forman, Chair
Professor	Kaye L. Brubaker
Assistant Professor	Michelle T. Bensi

Acknowledgements

Firstly, I would like to give a special thank you to my advisor, Dr. Barton A Forman, for his guidance, assistance and patience during my two-year-study. I would also like to thank him for providing me with the opportunity to come to the U.S., for introducing remote sensing and data assimilation techniques to me. He is an enthusiastic, responsible and caring professor. In class and life, he gave me a lot of help and advice, so that I can smoothly integrate into the unfamiliar American campus life. In terms of academic and research, he was rigorous in his studies and patiently guided me, so that I could become a qualified master's graduate.

I would like to thank Dr. Brubaker and Dr. Bensi for being my defending committee members and providing me with so many helpful suggestions. They are excellent professors I have ever met.

I would like to thank my parents for sponsoring me to come here and study in such a great university. Also, they inspired me to challenge myself to overcome difficulties when I encounter setbacks.

I would like to thank Colin Mclaughlin and Alieza Moghaddasi. As the members of Dr Barton Forman’s research group, they helped me a lot on my study and life in my first semester here and shared their study experience with me.

At last, I would like to thank the officers in UMD writing centers, they helped me a lot on improving my academic writing skills which is important to my thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AE_Land3	Level-3 land surface product
AMSR-2	Advanced Microwave Scanning Radiometer 2
GHz	Gigahertz
GP	Gaussian Process
GPR	Gaussian Process Regression
LIDAR	Light Detection and Ranging
LIS	NASA Land Information System
LSTM	Long Short Term Memory
MSE	Mean Square Error
NASA	National Aeronautics and Space Administration
PMW	Passive Microwave
RBF	Gaussian Radial Basis Function
RADAR	Radio Detection and Ranging
SWE	Snow Water Equivalent
SNOTEL	The U.S. Snowpack Telemetry network
SVM	Support Vector Machine
Tb	Brightness Temperature

Chapter1.INTRODUCTION AND MOTIVATION

1.1.Introduction to Snow

1.1.1. Importance of Snow

Terrestrial snow is an important component of the hydrologic cycle. It has a critical influence on the land-atmosphere interactions in hydrology, including the control of mass and energy exchange [Robinson et al. 1993]. Consequently, snow exerts a significant impact on global climate change and regional weather conditions [Barnett et al., 1989]. Additionally, water from melted snow and ice is the source of freshwater resources for many regions in the world. Over one billion people depend on it as a freshwater supply [Foster et al., 2011].

At peak accumulation, snow covers approximately 40 percent of the northern hemisphere each year [Hall and Martinec, 1986]. The importance of snow is not only for human water supply but also as a dominant control on the global energy balance[Armstrong and Brun, 2010]. Every object on the Earth can emit and reflect radiation across some range of radiation based on its physical properties [Campbell, 2002]. Snow and ice cover have a high albedo (reflection coefficient), which means they can reflect plenty of solar radiation.

Some scientists compare snow and ice cover to a minor of the Earth surface. Fresh snow with low density can reflect more than 75 percent of income solar radiation, which

is almost 15 times that of what wet earth can reflect [Thornes and Lydolph, 1987]. Hence, snow and ice cover have an important influence on the global energy balance.

Since snow plays such an important role in human life, we need to have a better understanding and characterization of terrestrial snow, particularly in the face of a warming world. To solve our water supply issues and carry out further research on snow, we must first quantify the amount of snow to better manage and protect this important resource.

1.1.2. Methods of Measuring Snow

Knowing the amount of water stored in the snow is extremely essential for resource management. Snow water equivalent (SWE) is a variable used to describe the amount of liquid water within a snowpack. It combines snow density and snow depth and represents the amount of water available for the possible runoff after melting [Larson et al., 2009]. The management of water supply in snow-dominated regions is highly related to the measurement of water stored in the snowpack and melting rate forecast [Dozier and Shi, 2000].

Ground-based snow measurements are the most traditional way to measure snow depth and SWE. Most simply, we can determine SWE with a ruler, a snow core, and a scale. Inserting the ruler straight into the snow until it hits the ground can tell us the snow depth. With a snow core of known dimension and volume along with a measuring scale, we can compute the corresponding weight of the SWE, which is the liquid

equivalent of snow after subtracting the weight of the snow core. The snow density, $\rho snow$, is then computed as:

$$\rho snow = \frac{Weight\ of\ snow}{volume\ of\ snow\ core} \quad (1.1\text{-}1)$$

Finally, we can calculate SWE from ground-based measurements as:

$$SWE = \frac{D * \rho snow}{\rho water} \quad (1.1\text{-}2)$$

where the unit of snow water equivalent (SWE) meters [m]; D is the snow depth in meters [m]; ρsnow is the snow density [Kg*m-3]; and ρwater is the water density [Kg*m-3].

The U.S. Snowpack Telemetry (SNOTEL) network provides direct measurements of SWE across much of the United States. The limitation of SNOTEL stations' data is that only temporal information at the point scale is provided [Serreze et al., 1999]. In some cases, the impact of the topography and other factors make regional characterization of SWE using SNOTEL stations a difficult and uncertain process [Molotch and Bales, 2006]. Ground-based snow measurements are relatively sparse and unavailable in most areas [Luce, Tarboton, and Cooley, 1998]. The emergence and development of remote sensing technology provide a new opportunity for research. Remote sensing instruments on aircraft and satellites complement the spatial sparsity of ground-based measurements of snow properties and can help provide a more robust characterization of SWE across the region scale [Larson et al., 2009].

1.2. Introduction to Remote Sensing

1.2.1. Snow Remote Sensing

Remote sensing is defined as the collection of information about natural objects and phenomena on the earth without any direct physical contact with these objects [Eckerstorfer, Bühler, Frauenfelder, and Marines, 2016]. It is the science of accessing, processing and generating measurements, and collecting data via detecting the interaction between matter and electromagnetic radiation [Sabins, 2007]. Remote sensing technology encompasses plenty of different forms, such as airborne, space-borne, and ground-based sensors. The potential of remote sensing in the research of snow was discovered by scientists in the last century. Since that time, space-borne sensors date has been widely used in the field of snow science [Chang et al., 1982].

Snow researchers started estimating snow cover and snow mass by collecting emitted and reflected radiation using space-borne sensors decades ago. With the development of aerospace, electronic physics, and other sciences, several different forms of sensors have been designed to record electromagnetic radiation. For instance, a radiometer, as one of the most widely used sensors, can be either an infrared radiometer or a microwave radiometer. It is a device designed to measure the radiant flux of electromagnetic radiation presented within a specific wavelength range based on its design [Kiedron, Michalsky, Berndt, and Harrison, 1999]. RADAR is another type of sensor, which was designed as an object detection system when it was invented.

The basic principle of RADAR is using electromagnetic waves to determine the range, altitude, direction, or speed of objects. There is also a technology called LIDAR. Instead of using low-frequency electromagnetic waves, LIDAR choose utilizes visible light from pulsed lasers for detection and ranging.

1.2.2. Passive Microwave Sensor

We use a passive microwave sensor (PMW) in this study, whose main structure is an antenna system collecting the power of an electromagnetic wave emitted by the snowpack and surrounding environment. Recorded data will be transmitted into brightness temperature by a built algorithm to calculate the strength of reflected radiation.

There are two important parameters of PMW sensors we care about except for the antenna size. Frequency is one of them. The designers determine the antenna size as a function of the frequency. The antenna size should be on the order of one-tenth or more of the wavelength of the frequency emitted by target objects [Lathi and Dao, 1989]. In addition, for the space-borne sensor used in this paper, the antenna size is much bigger than one-tenth of the wavelength to achieve a minimal signal-to-power ratio.

Polarization is another significant parameter to consider in the design of PMW sensors. The definition of polarization is the orientation of the electromagnetic wave relative to the Earth's surface [Mott, 1992]. Two of the most substantial polarizations that we use are horizontal polarization and vertical polarization. The combination of

data in two different polarizations can often provide users with really significant feedback information.

In addition, the wavelength of radiation emitted by snow cover is at a relatively low spectral frequency. The PMW sensor we use in this study is designed to detect microwaves emitted by objects like snow and ice. Compared to optical (visible) radiation, microwave radiation has a larger penetration depth through media. As a result of this physical property, microwave radiation can be detected by the sensor during both daytime and night under all-weather conditions, which means the PMW sensor can work during cloudy conditions.

1.2.3. Brightness Temperature

The electromagnetic attributes of snow change in different snowpack conditions. For example, the dielectric constant is an electromagnetic attribute of snow that we care about, is defined as a measurement of the polarization of the objects upon interaction with the electromagnetic wave. It varies with properties such as the amount of liquid water in the snow and snow structure [Mulders, 1987]. In general, snow has a dielectric constant between 1.2 and 2.0, which can be different during extreme conditions [Hallikainen, Ulaby, and Abdelrazik, 1986]. When the snow ablation season begins, a snowpack tends to have a larger dielectric constant. That is because snow contains a large amount of liquid water during the ablation season, and liquid water contains a larger dielectric constant [Rango, 1996].

Researchers use a parameter called brightness temperature (Tb) to characterize the signal information recorded by the space-borne PMW sensor. The definition of Tb is the equivalent temperature of the microwave radiation thermally emitted by an object [Chang et al., 1976]. To put it another way, brightness temperature is the temperature a black body should have to emit the same intensity of radiation as a gray body object when it is in the thermal equilibrium [Ahmad, Forman, and Kwon, 2019]. According to the Raleigh-Jean approximation for microwave radiation, the brightness temperature is related to objects' real temperature and their emissivity [Zwally and Gloersen, 1977] as:

$$Tb = \varepsilon \cdot Tphysical \quad \varepsilon \in [0\ 1] \tag{1.2-1}$$

where $T_{physical}$ is the real temperature of objects in [K], ε is the emissivity of objects [unitless] and Tb is brightness temperature in [K]. The emissivity of an object is frequency dependent and is a strong function of the dielectric constant of the object.

1.3. Goals and Objectives

In recent decades, machine learning is widely used as a data mining and prediction tool by snow researchers. The goal of the research is to understand the basic behavior of three different machine learning algorithms and analyze whether it is consistent with the physical characteristics of snow. This goal was achieved through the following objectives:

(1) Understand the basic physical principle of predicting SWE as a function of PMW brightness temperature data at 10 GHz, 18 GHz, and 36 GHz.

(2) Evaluate the performance of the three different machine learning algorithms (Long Short Term Memory, Support Vector Machine Regression, and Gaussian Process Regression) under different snow depth and vegetation conditions and point out their advantages and disadvantages.

(3) Analyze the prediction results of three machine learning algorithms and highlight the limitations of PMW snow remote sensing retrievals, including the indirect effects of vegetation and snow depth on the algorithm.

(4) Highlight the reasons why PMW snow remote sensing retrievals are less effective in regions with dense forest, regions with shallow snow, regions with deep snow, or in areas undergoing ablation when the snow is relatively wet.

Chapter2.BACKGROUND AND LITERATURE REVIEW

The following chapter talks about different types of snow measurements and principles related to remote sensing of snow. It also discusses the similarities and differences between three different machine learning algorithms as applied to snow

2.1. Ground-based Snow Measurements

In the days before remote sensing, ground-based snow measurements were the most relied upon by snow researchers. Although with the development of remote sensing technology, remote sensing and other advanced measurement methods are more commonly used. Ground-based snow measurements have the unique advantage of providing direct snowpack information and measurements [Armstrong and Brun, 2010]. There are several snow measurement techniques at the point scale. As we mentioned in Chapter 1, the simplest way is to use a snow ruler to measure the snow depth and collect a snow core at the observation time and location to calculate the snow density. We can get information on SWE by taking the product of snow depth and snow density data. With the advent of radiometers and sensors, ground-based snow measurements ushered in a new development. For example, an ultrasonic snow depth sensor whose working principle is using ultrasonic to measure snow depth. It calculates the snow depth based on the travel time and speed of the ultrasonic impulses emitted by the sensor [Lea, 1998]. In terms of large-scale ground-based snow measurements, an interpolation algorithm has some advantages in SWE estimation. It can reproduce the approximate SWE/snow

depth data in a large area based on the point-scale observations near to that area [Dyer and Mote, 2006]. However, the interpolation algorithm has many limitations, due to the spatial and temporal variability of the snow mass. Not only in the interpolation algorithm but also all types of ground-based snow measurements, the special and temporal variability of snow is a challenge. The spatial resolution of ground-based snow measurements highly depends on the proximity, and the number of available located stations [Bechle, Millet, and Marshall, 2013]. Because of the limitation of terrain and climate, we do not have high-quality ground-based measurements in regions that are hard for the human footprint to reach. Further, ground-based snow measurement data often lack temporal continuity given data gaps associated with sensor maintenance and replacement. Considering these limitations of ground-based snow measurements, remote sensing is an attractive technology for snow measurement across a range of spatial and temporal scales [Foster, Hall, and Chang, 1987].

2.2. Snow Remote Sensing

Remote sensing began to be widely used for snow measurements near the end of the 20th century [Kelly, Chang, Tsang, and Foster, 2003]. Given the time continuity and global scale required for snow cover research, the sensors aboard Earth observation satellites became increasingly popular. There are two general types of sensors used to measure snow: (1) optical sensors or (2) microwave sensors. Optical sensors are commonly used to map the distribution (i.e., spatial extent) of snow cover while

microwave sensors often focus on measuring snow depth or SWE.

Compared with broadband shortwave radiation, microwave radiation has a much longer wavelength, and hence, can travel much further through the snowpack as well as some low-density vegetation. [Ulaby and Stiles, 1980]. PMW sensors provide us with an indirect measurement of snow water (SWE) by measuring the brightness temperature. That is, PMW remote sensing of snow relies on a 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 snowpack. The increased scattering at high frequencies decreases the emissivity, and hence, lows the corresponding measured brightness temperature [Chang, Foster, and Hall, 1987]. In an idealized scenario, the intensity of radiation at 18.7GHz or 36.5GHz emitted by the soil surface should only have a small difference. When it comes to snow mass associated with volume scattering, the difference in scattering at a higher frequency (36.5GHz) relative to the lower frequency (18GHz) serves as the foundation for PMW remote sensing as illustrated in Figure 2.2-1.

SWE retrievals based on Tb measurements from space-borne microwave radiometers including the Scanning Multichannel Microwave Radiometer (SMMR) and the Advanced Microwave Scanning Radiometer for Earth Observing Systems (AMSR-E) have been proven to be successful tools in estimating SWE across some portions of the globe. The following study focuses on the data from Advanced Microwave Scanning Radiometer 2 (AMSR-2).

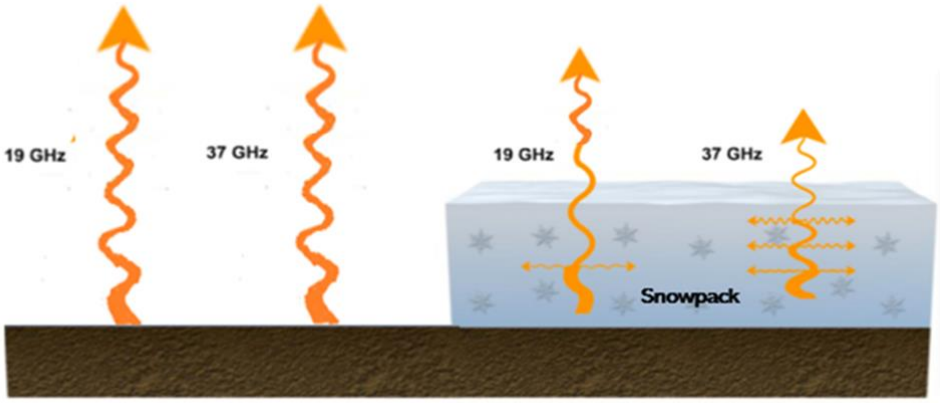


Figure 2.2-1 Preferential Scattering of Microwave Radiation Having Frequency 37GHz Compared to Microwave Radiation with Frequency 19GHz

2.3. *Existing Snow Estimation Products*

2.3.1. Spatial Interpolation

There are typically three ways to leverage snow mass information collected by space-based sensors. One of them is spatial interpolation, which combines the space-born PMW observation data of a relatively coarse spatial resolution with point-scale ground-based measurements. This method highly depends on the quality and quantity

of ground-based snow measurements data, and has a poor performance in regions with sparse spatial coverage of observations [Takala et al., 2011]. In addition, spatial interpolation is strongly affected by the strong variability of snow properties in complex terrain such as between mountains and valleys [Foppa, Stoffel and Meister, 2007].

2.3.2. SWE Retrieval

The second approach uses brightness temperature at a specific frequencies and in-situ measurements to calibrate regression coefficients in the retrieval algorithms. Chang et al. [1986] put forward the first snow depth retrieval of brightness temperature for a uniform snowpack with a fixed snow density of 300 kg/m3 and a mean snow grain size (radius) of 0.3mm, which was expressed as:

$$D = 1.59 \times (Tb18,H - Tb37,H) \quad (2.3.1)$$

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

With more research, Goodingson and Walker [1994] presented another commonly used form of SWE retrieval of brightness temperature for dry snow based on Chang's research, which was expressed as:

$$SWE = a + b(Tb37,V - Tb19,V) \quad (2.3.2)$$

where SWE is the snow water equivalent [mm]; a and b are fixed parameters; $Tb_{37,V}$ is the brightness temperature [K] at 37 GHz at vertical polarization; and $Tb_{19,V}$ is the

brightness temperature [K] at 19 GHz at vertical polarization; and a and b are fixed parameters from regression tests, a = -20.7 [mm] , b=-2.74 [K^{-1}].

Kelly et al. [2003] more closely considered the snow grain radius and volumetric fraction data from old models and developed a new generation of snow retrieval estimating snow depth based on SMM/I data at a constant snow temperature of 260 [K], which was expressed as:

$$D = b(Tb19,V - Tb37,V)\ 2\ + c(Tb19,V - Tb37,V) \quad (2.3.2)$$

where D is the snow depth [cm]; $Tb_{37,V}$ is the brightness temperature [K] at 37 GHz at vertical polarization; and $Tb_{19,V}$ is the brightness temperature [K] at 19 GHz at vertical polarization; b and c are regression coefficients related to snow grain radius and volumetric fraction data.

In addition to snow grain size, vegetation is another important factor that has a significant impact on snow retrieval algorithms [Tedesco and Narvekar, 2010]. Vegetations over snowpack will absorb and reflect the microwave radiation emitted by the snowpack under them. In addition, they also emit their own microwave radiation that contributes to the noise within the signal as measured by the PMW sensors [Derksen, Walker and Goodison, 2005]. After presenting the first snow depth retrieval using PMW brightness temperature, Chang et al. [1996] incorporated vegetation effects into his original snow retrieval and improve the SWE estimation accuracy in forested regions. This new retrieval uses the following expression:

$$SWE = \frac{a(Tb_{19,V} - Tb_{37,V})}{(1 - ff)} \quad (2.3.3)$$

where SWE is the snow water equivalent [mm]; $Tb_{37,V}$ is the brightness temperature [K] at 37 GHz at vertical polarization; $Tb_{19,V}$ is the brightness temperature [K] at 19 GHz at vertical polarization; a is a fixed regression coefficient; and ff is the forest fraction ranging from 0 to 0.75 [unitless].

With the launch of AMSR-E near the start of 21th century, Kelly et al. [2009] derived a unified expression for snow depth retrieval for both forested and non-forested regions:

$$D = ff * \left[P1 * \frac{(Tb_{18,V} - Tb_{36,V})}{1 - b * fd} \right] + (1 - ff) * [P1(Tb10,V - Tb36,V) + P2\ (Tb10,V - Tb18,V)] \quad (2.3.4)$$

where D is the snow depth [cm]; $Tb_{36,V}$ is the brightness temperature [K] at 36 GHz at vertical polarization; $Tb_{18,V}$ is the brightness temperature [K] at 18 GHz at vertical polarization; $Tb_{10,V}$ is the brightness temperature [K] at 10 GHz at vertical polarization; ff is the forest fraction [unitless]; fd is the forest density; b is the calibration coefficient; and P_1 and P_2 are two dynamic coefficients ranging from 1 to 2.

All these retrievals shown above are based on some assumptions such as uniform snow grain size and snow density, which is not reasonable in the real world. Factors shown below all have a significant impact on the accuracy of SWE retrieval algorithms and leads to the uncertainty of retrieval results.

Snow Grain Size

Snow grain size have influence on snow albedo, and is a challenging parameter to

characterize in the snow retrieval research given its variability in time and space [Armstrong, Chang, Rango and Josberger, 1993]. In addition, thick depth hoar layer composed with large loose snow grains near the bottom of the snowpack can attenuate microwave emission, which might cause the measured Tb to decrease [Hall and Martinec, 1986].

Snow Depth

Snow retrieval based on data from spaced-based PMW sensors can only work in a specific range of snow depth depending on the frequency. If the snowpack is deep, the radiation from the soil surface will be significantly scattered away in the snowpack, and the sensor will mistake the signal it receives from the middle of the snow layer as coming from the surface of the soil resulting in a retrieval result that is far less than the actual snow depth. In shallow snow conditions, the brightness temperature at two frequencies will be almost the same, which makes the retrieval completely fail.

Ice Layers

Ice crusts on the surface and within the snowpack will also contribute to increased scattering of radiation emitted by the underlying soil surface, hence, leading to a lower brightness temperature. As a result, snow retrieval may over-estimate snow depth and SWE.

Wet Snow

Wet snow has a completely different emissivity than dry snow. The liquid water in the wet snow emits more radiation itself than dry snow and thus increase the measure

brightness temperature [Walker and Goodison, 1993]. The reason is that liquid water has a higher r dielectric constant compared to that of dry snow, which causes the dielectric of snowpack to increase. Treating water in wet snow as dry snow can lead to significant errors in snow retrievals [Tedesco and Narvekar, 2010].

2.3.3. Land Surface Models

To reduce the uncertainty and error of existing satellite-based snow retrieval algorithms, the third method combines remote sensing observations and land surfacemodels' (including a snow physical module) estimation to improve the snow estimates [Reichle, 2008]. The basic principle of this technology is called data assimilation (DA), which merges the remote sensing measurements with the model estimates by weighing their respective uncertainties [McLaughlin, 2002]. Combing with the model estimation data makes the snow retrieval results more consistent with the physical characteristics of snow, and can effectively eliminate the wrong retrieval results that violate the physical characteristics.

Although this technology improves the performance of snow retrieval and overcomes some limitations of existing satellite-based snow retrieval algorithms, its practical application is still plagued by the spatial and temporal variability of snow [Pulliainen, Grandell and Hallikainen, 1999]. Wet snow, deep snow, complex terrain and mixed land cover still limit the utilization of this technology. Furthermore, because of the complexity and huge computational cost of this method, it is computationally

expensive to apply across large spatial domains [Durand, Kim and Margulis, 2008]..

2.4.Machine learning in SWE Retrieval

The uncertainty and limitations of all the existing snow retrieval algorithms mentioned in Section 2.3 motivate this research to explore a new approach to estimate SWE using space-based PMW radiometers.

Since the end of the last century, some snow researchers have tried to introduce machine learning algorithms into snow retrievals [Tsang et al. 1992; Davis et al. 1993; Tedesco et al. 2004]. In general, these studies focused on using model estimates (from physical models) along with ground-based measurements as input data to machine learning algorithms to predict SWE. Those predictions were then compared with the observation data not used during training to test the performance of those algorithms. Good agreement was found in the training area, which proved the feasibility of machine learning in snow retrieval. However, these applications are limited to certain regions such as the Antarctic region [Tsang et al. 1992] which has almost no vegetation and is relatively flat. A recent study conducted by Forman et al. [2013]; Forman and Reichle [2014]; and Xue and Forman [2015] applied machine learning based on snow properties(modeled) and brightness temperatures(measured) by utilizing ANN and SVM regression across large spatial scales. Based on their previous research, this article will explore and compare the performance of additional machine learning algorithms.

2.5. *Introduction to Machine Learning*

The definition of machine learning was proposed in 1959 by Arthur Samuel. Machine learning is a technology where computers are capable of learning patterns without programming expressly. It is more explicitly defined as the process of identifying a set of categories where a new observation is included in the basis of training data containing observations whose category membership is known [Ruppert, 2004].

Machine learning, which requires the analyst to label the training data, defines characteristic class signatures that are used to allocate labels to all other undefined areas in the model framework [Campbell,2002]. The difference between machine learning and unsupervised algorithms is that unsupervised algorithms are self-organizing and give models the ability to find natural data clusters, but machine learning refers to making computers automatically collect, analyze, and generalize the useful information based on the known training database in order to give accurate predictions with unused data in the future.

The purpose of machine learning is to generate classifying or regression expressions and functions in complex forms and simplify them in order to help humans understand them [Michie, Spiegelhalter, and Taylor, 1996]. One of the advantages of machine learning over traditional statistical methods is that it can make predictions and classifications without an explicit underlying probability model, so it has a promising application in many complex practical problems such as the stock market and Web

search [Domingos, 2012].

Since the theory of machine learning was proposed, a great variety of machine learning algorithms have been developed. Different algorithms differ in performance for different problems, so we will discuss the fundamentals of three different algorithms in the following sections.

2.5.1. Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) is an algorithm improved from the recurrent neural network (RNN), which is designed specifically for time series data [Gers, Schmidhuber and Cummins, 2000]. Compared to the original RNN algorithm, LSTM is much more efficient in dealing with long-term observation data. As the basic framework shown in Figure 2.5-1**,** LSTM contains several layers: (1) Input layer: these layers are used to load data from the original database and get information from outside. (2) Hidden layers: two or more hidden layers linked with each other builds the main body frame. They connect the input layers and output layers like a bridge, in which the information signals it chose to remember can remain and transfer between layers. (3) Output layers: layers respond to generate the output data we need and send it out of the network.

LSTM comes in several different forms of data transmission, including forward LSTM, backward LSTM, and duplex transmission LSTM. This paper chooses forward LSTM, which means the network does not contain any feedback connections. The

signals can only transfer in one direction: from the input layers assigned at first to the hidden layer and subsequently propagate into the output layer [Atkinson and Tatnall, 1997].

Xn is the input information in time step n. L1,n and L2,n are the hidden neurons for time step n. Instead of using the hidden information saved in neurons at the last time

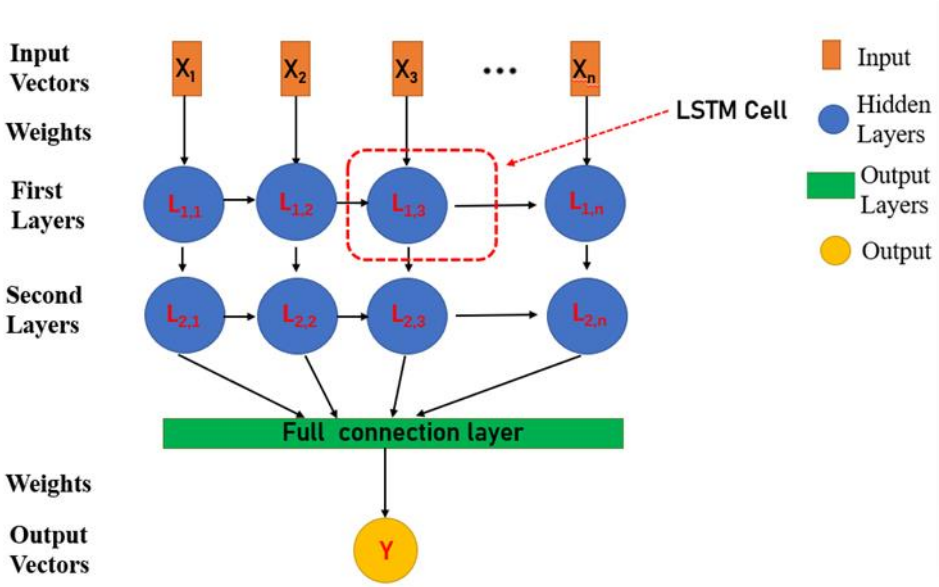


Figure 2.5-1 Basic Structure of Long Short Term Memory (adapted from Kratzert,2019)

step as an additional input like RNN, LSTM introduces a memory system called “gates” [Yin et al., 2022]. Based on the research of Kratzert et al. [2019], as Figure 2.5-2 shows, for each LSTM cell (hidden neuron), LSTM sets up three gates to control the information transmission: Forget gate F; Input gate I; Output gate O.

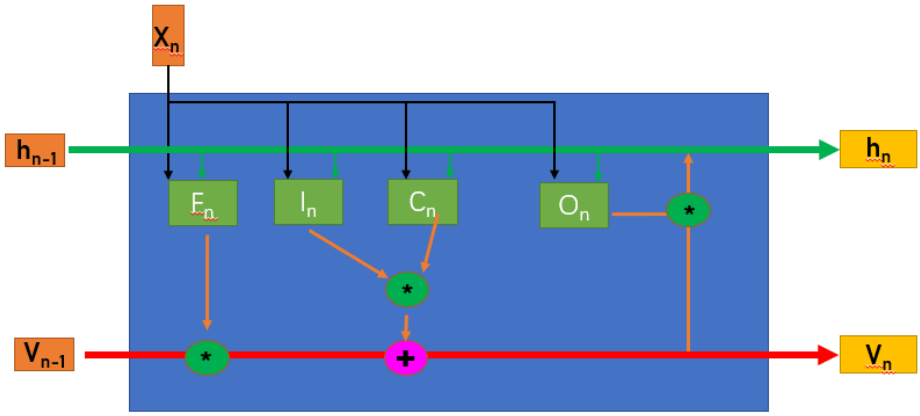


Figure 2.5-2 The Structure of Each LSTM Cell

Forget gate

The forget gate decides what information this cell is going to forget. It ranges from 0 to 1, 0 means totally forget the previous memory, 1 means completely keeping all information from the last time step. The function form is expressed below:

$$Fn = Ó(W_f(h_n - 1, Xn) + bf) \quad (2.5.1)$$

where Ó(.) is the sigmoid function; h_{n-1} is the hidden state information from the previous time step; Xn is the input data in running time step; W_f is a vector of weights and b_f is a vector of bias.

Input gate

The input gate controls how much new information from the input vector Xn is going to be added into the system. The expression is shown below:

$$In = Ó(Wi(hn - 1, Xn) + bi) \quad (2.5.2)$$

where Ó(.) is the sigmoid function; hn-1 is the hidden state information from the previous time step; Xn is the input data in the running time step; Wi is a vector of

weights, and bi is a vector of bias.

Cn is the cell input gate that generates cell state information. It can be expressed as:

$$Cn = tanh(Wc(hn-1, Xn) + bc) \quad (2.5.3)$$

where tanh(.) is the hyperbolic tangent function; h_{n-1} represents the hidden state information from the previous time step; Xn is the input data in the running time step; W_c is a vector of weights; and b_c is a vector of bias.

Output gate

The output gate decides which information stored in the cell states is sent into the output layer. The expression of the output gate is:

$$On = Ó(Wo(hn-1, Xn) + bo) \quad (2.5.4)$$

where Ó(.) is the sigmoid function; h_{n-1} is the hidden state information from the previous time step; Xn is the input data in the running time step; W_o is a vector of weights; and b_o is a vector of biases.

Cell states

V_n is the cell states that characterizes the memory of the system. It can be generated from the forget gate and input gate as

$$Vn = Ft * Vn-1 + It * Ct \quad (2.5.5)$$

where * represents element-wise multiplication.

In addition, the function for hidden state information is expressed as:

$$hn = On * tanh(Cn) \quad (2.5.6)$$

where tanh(.) is the hyperbolic tangent function.

With this memory system, LSTMs can avoid vanishing gradients problems existing in original RNN algorithms [Kratzert et al.2019]. That is, LSTMs have the ability to learn long-term relationships between input and output features [Zhang, Lindholm and Ratnaweera, 2018]. Consequently, LSTMs are thought to possess advantages in analyzing long timescale data such as snow cover and SWE.

2.5.2. Support Vector Machine Regression

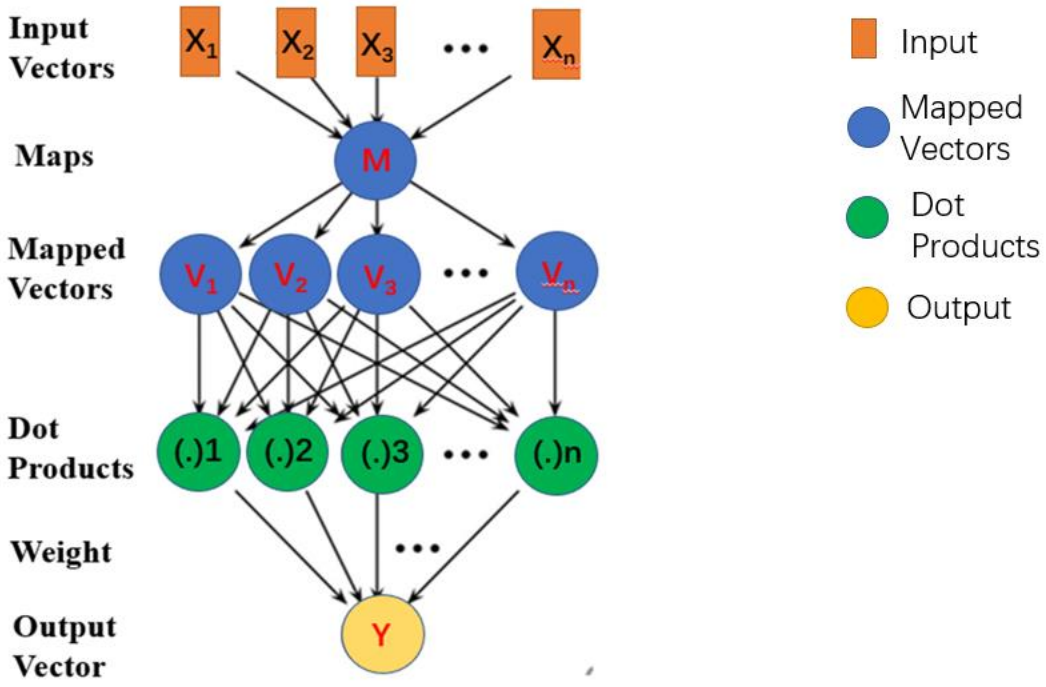


Figure 2.5-3 The Basic Structure of Support Vector Machine Regression (adapted from Barton A Forman,2014)

The Support Vector Machine (SVM) was first presented by Vapnik et al. [1998] in order to improve the efficiency of solving non-linear function problems. The theoretical foundation of SVM is the statistical learning theory put forward by Vladimir Vapnik

and Alexey Chervonenki [Cherkassky, 1997]; [Shoesmith, Vapnik and Kotz, 1984]; [Vapnik, 1999]. In terms of function, SVMs can be divided into SVM regression and SVM classification. This paper focuses on SVM regression whose principle is trying to find out the non-linear dependence between input matrix X and output vector Y.

Input matrix X and training targets Y corresponding to X together form the training data set t {(X1, Y1),…, (Xn, Yn)}, where Xn ∈ Rx and Yn ∈ RY where Rx represents the input space, RY represents real space where Yn belongs [Vapnik, 1999]. The basic framework of SVM regression can be seen in Figure 2.5-3. SVM regression assumes that M(x) (a.k.a.,"map" function) is a non-linear function to transmit input data set into a feature space. The input data sets that are transmitted into the feature space are called Mapped Vectors (Vn) which can be expressed as:

$$Vn = M(Xn) \tag{2.5.7}$$

Next, the F(.) is the function describes the linear combination of Mapped Vectors as:

$$F(Vn) = W * Vn + b\,;\ b \in RY \tag{2.5.8}$$

where W is a vector of weights and b is a vector of bias, and both of which are determined during the training process.

The training process of SVM in a given location can be treated as a convex optimization problem. The solution is shown as [Smola and Schölkopf, 2004]:

$$\begin{aligned} &Minimize\ (w,b,\xi)\quad \frac{1}{2}w2 + C\sum_{n=1}^{p}\xi n\ ,C > 0 \\ &Subject\ to\ \ WVn + b - Yn \le \xi n \\ &\xi n \ge 0, n = 1,2,\dots,i \end{aligned} \tag{2.5.9}$$

where W is the weight vector in the given location, C is a penalty parameter of the error term; p is the number of measurements in time; ξ is the slack variable, which intends to relax the boundary conditions and allow for a certain proportion of outliers; and Yn is the training target in time step n.

Parameter C determines how many data points within an abnormal deviation can be tolerated in the process. If C is given as a large number, the number of permitted outliers approaches zero. However, real measurements are difficult to achieve this requirement because real measurements cannot avoid deviation and errors. As a result, the algorithm many overfit the model by defining too many support vectors. Similarly, if C is given a as a small number, the number of support vector becomes too small, which causes the algorithm to generate an overly simple model. In practice, people use cross-validation in the algorithm to search for the best value of C [Vapnik, 1999].

The complexity of this convex optimization problem is determined by the dimensionality of training target Y [Fletcher, 1998]. The solution to this problem is using its dual formulation utilizing Lagrange multipliers, where time complexity can be transformed into the training targets' number. The original function (Equation 2.5.9) can be written as [Smola and Ikopf, 2004]:

$$Maximize\ (ai, ai*)\ \{-\frac{1}{2}\sum_{i,j=1}^{p}(ai-ai*)(aj-aj*) < M(xi).M(xj) > -\sum_{i=1}^{p} Yi(ai-ai*)\}$$

$$subject\ to\ \sum_{i=1}^{p}(ai-ai*) = 0\ , ai, ai* \in [0, C], i = 1,2,\dots,p \qquad (2.5.10)$$

where ai,ai* are Lagrangian multipliers; <M(xi) ·M(xj) > is dot product of M(xi) and M(xj); xi and xj are two training points; and C is the penalty parameter mentioned above.

In order to computer the dot products in Equation 2.5.10, a new method called "Kernel Function" is introduced in SVM. In this study, it can be defined as:

$$k(xi, xj) = < M(xi) \cdot M(xj) > \quad (2.5.11)$$

using kernel functions can help us avoid the risk of too large a dimensionality of M(x) making weight vectors difficult to represent explicitly in memory [Cortes and Vapnik, 1995].

There are four widely used kernel functions in SVM regression: linear kernel, polynomial kernel, hyperbolic tangent (sigmoid) kernel, and gaussian radial basis function (RBF) kernel [Smola and Schölkopf, 2004]. This study uses the RBF kernel as the kernel function, which is defined as:

$$k(xi, xj) = exp\left(-\gamma\left|\left|xi \cdot xj\right|\right|2\right) \quad (2.5.12)$$

where x_i and x_j are single instances of input X; ||xi · xj|| represents the Euclidean norm of x_i and x_j; γ is a positive parameter controls the width of Gaussian distribution [Ben-Hur and Weston, 2010]. If γ is too large, the RBF kernel function will behave like the linear function and lose its ability to represent the non-linear relationship. On the contrary, the decision boundary of SVM will become too sensitive to the noise in the training data when γ is underestimated [Haasdonk and Burkhardt, 2007].

2.5.3. Gaussian Process Regression

Gaussian Process Regression (GPR) is a nonparametric, kernel-based probabilistic model. A GPR model introduces latent variables to explain the response through a

Gaussian process (GP) [Wu and Wang, 2018].

The definition of Gaussian process (GP) will be explained below. There are n samples of input data X (x_1, x_2, …… x_n), and the output data Y can be written as $Y = f(x)$. Then based on the assumption that the joint distribution of f(x_1), f(x_2), ……, f(x_n) is multivariate normal (Gaussian), if this assumption is true for any n, the distribution of Y is defined as a Gaussian Process [Fearn, 2013]. There are two significant parameters used to specify a multivariate normal distribution. One is mean vector, and the other one is covariance matrix. To simplify the process, researchers usually assume the mean vector as 0 at first and use the kernel functions mentioned in SVM regression to describe the covariance matrix. Consequently, the basic structure of GPR is shown in Figure 2.5-4**.**

In a GPR algorithm, it is expected that the distribution of y at this observation depends on x, and the closer the two x are, the smaller y becomes. Hence, for each sample x_i, GPR introduces a latent variable f (x_i) following Gaussian distribution, and all f (x_i) together is a set of random variables with a joint Gaussian distribution. The GPR model can be expressed as:

$$Y = h(x)^T * \beta + f(x) \quad f(x) \sim GP(0, k(x, x')) \tag{2.5.13}$$

where f(x) is the latent variable; h (x) is a set of "basis functions" that transform the n-dimensional original eigenvectors into p-dimensional new eigenvectors to be consistent with the dimension of f (x) on the right. β is the parameter vector of p*1 [Raissi, Babaee and Karniadakis, 2019]. Based on Equation 2.5.14, the distribution of Y in the model

can be described as:

$$P(yi|f(xi),xi)\sim GP(yi|h(xi)^T * \beta + f(xi), ó2) \quad (2.5.14)$$

At the same time, the joint distribution of latent variable f (x_i) can be expressed as:

$$P(f|X)\sim GP(f|0,K(X,X)) \quad (2.5.15)$$

where $K(X,X)$ is a Gramme Matrix shown as:

$$K(X,X) = \begin{bmatrix} k(x1,x1) & \dots\dots & k(x1,xn) \\ \dots\dots & k(x_{0.5n},x_{0.5n}) & \dots\dots \\ k(xn,x1) & \dots\dots & k(xn,xn) \end{bmatrix} \quad (2.5.16)$$

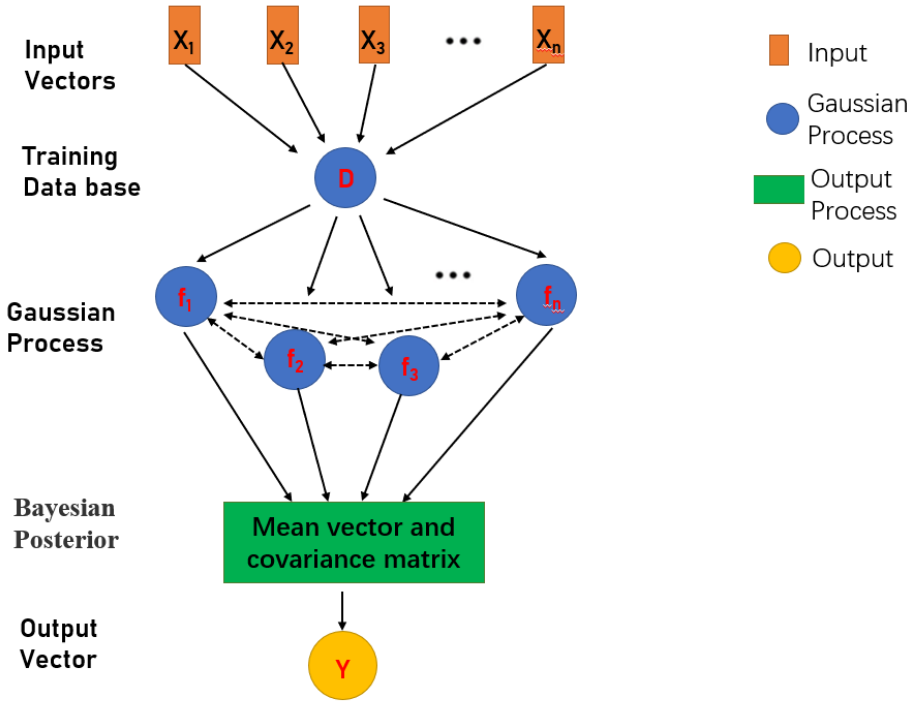


Figure 2.5-4 The Basic Structure of Gaussian Process Regression

During the training process, there are three main coefficients to be estimated: β, σ2 and θ (the hyperparameters of the kernel). They can be estimated by maximizing the following marginal log-likelihood [Wu and Wang, 2018]:

$$l(\beta, \sigma 2\ , \theta) = \ arg\ max\ log\ P(Y|\ X, \beta, \sigma 2\ , \theta) \qquad (2.5.17)$$

When applying the model for prediction, we also incorporate the $f(x_{new})$ of the new sample into f(X) according to the Bayes formula, and this population conforms to the following joint probability distribution [Verrelst et al., 2016]:

$$P(y\_new|Y, X, xnew) = \frac{P(y\_new, Y|X, x\_new)}{P(Y|X, x\_new)} \qquad (2.5.18)$$

The kernel function $k(xn, xn)$ is the same RBF kernel introduced in SVM (Section 2.5.2).

Chapter3.METHODOLOGY AND STUDY DOMAIN

3.1.Study Domain

This study focuses on the snow in Western Colorado, which has a relatively complete ground snow monitoring network as shown in Figure 3.1-1. Ground-based snow measurements are significant to this research as they serve as the training targets for these machine learning algorithms. The SNOTEL stations provide high-quality SWE information in Western Colorado, which leads to the determination of the study domain. In addition, Western Colorado is a region that has applicable snowfall annually. As mentioned in Chapter 2, extremely deep or extremely shallow snow will contribute to the invalidation of the physical principles of snow retrieval based on PMW brightness temperatures and associated volume scattering. The result of machine learning algorithms is dubious if it cannot be supported by the corresponding physical principles.

The study domain and the location of SNOTEL stations are shown in Figure 3.1-1. Black dots are the SNOTEL stations. Red dots represent the selected SNOTEL stations whose screening criteria will be explained in detail in Part 3.3.

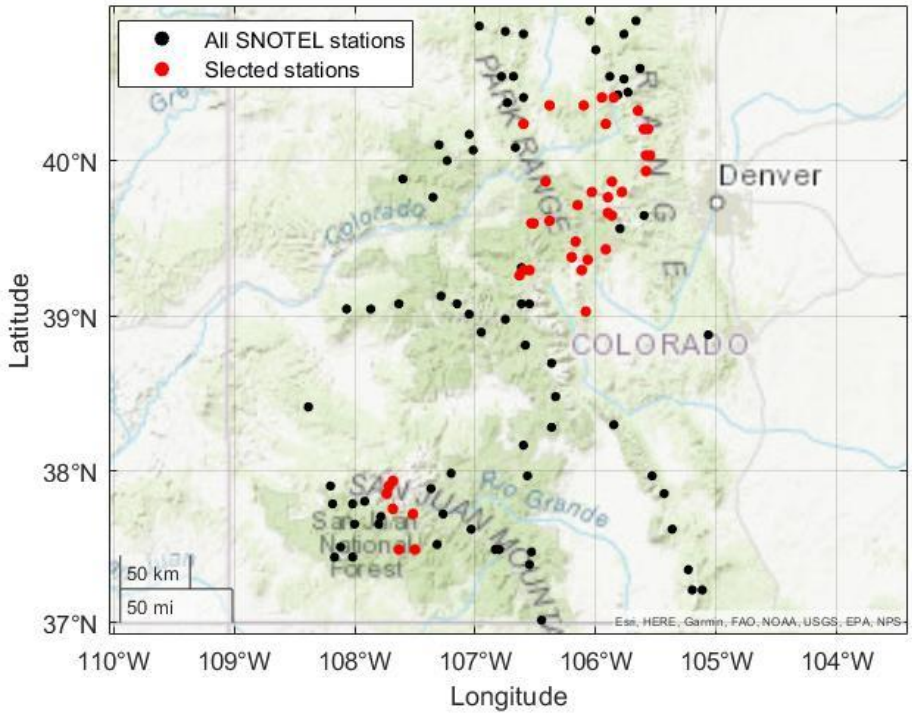


Figure 3.1-1 The Location of Selected Test SNOTEL Stations in Study Domain

The Rocky Mountains occupy most of the study domain. Abundant snow resources in western Colorado provide a good environment and opportunity to carry out this research. At the same time, complex mountain terrain and variable land surface types limit some of the research. Vegetation is one of the significant factors impacting SWE remote sensing using brightness temperature from PMW sensors in space [Langlois et al., 2011]. To better understand the impact of vegetation types on the study, land surface type data from the NASA Land Information System (LIS) [Waters et al., 2021] are incorporated into this study. Figure 3.1-2 shows the land surface types of the study domain and the location of SNOTEL stations.

The original LIS land surface data has 14 different land surface types. Based on the needs of this study, the original 14 land surface types are reclassified and combined

to achieve the purpose of simplification. Categories 1 to 5 including deciduous broad-leaved forest, evergreen broad-leaved forest, deciduous coniferous forest, and evergreen coniferous forest, are merged into category forest (dark blue in Figure 3.1-2). Categories 6, 7 and 14 including shrubs and subtrees are merged into the category shrubs (green in Figure 3.1-2). Categories 8, 9, 10, and 12 including different types of grassland are merged into the category grass (light blue in Figure 3.1-2). The urban area belongs to category 11, which is mainly found in the upper right corner of the study domain (dark yellow in Figure 3.1-2).

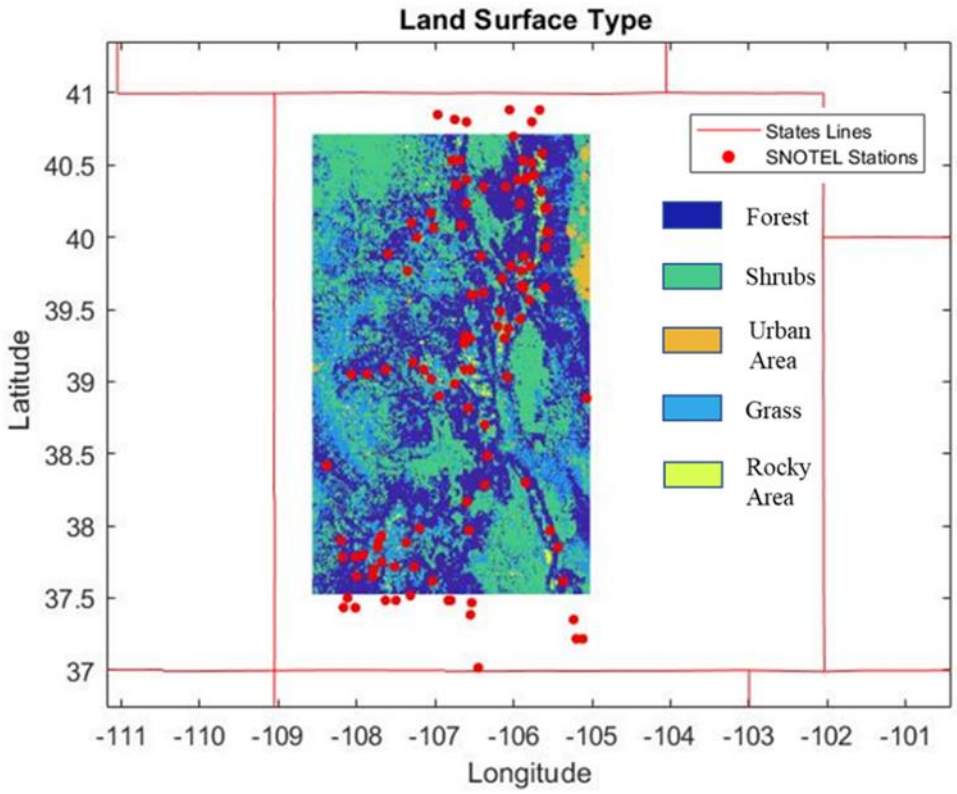


Figure 3.1-2 The Land Surface Type Map of All SNOTEL Stations in Study Domain

Besides vegetation, elevation is another important environmental parameter of the study domain. Snow depth and SWE collected by different SNOTEL stations may have very different values at different elevations. As a result, elevation will be an important environmental factor when evaluating the performance of machine learning algorithms.

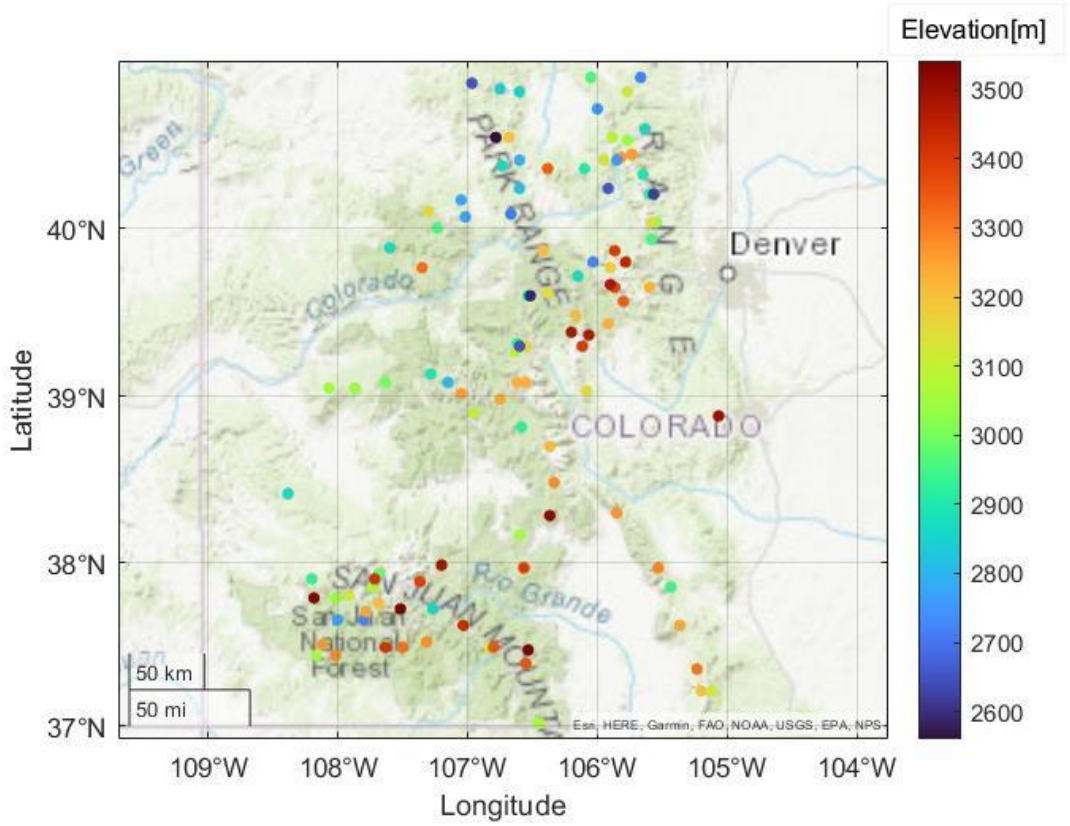


Figure 3.1-3 The Elevation Map of All SNOTEL Stations in Study Domain

Therefore, Figure 3.1-3 shows the elevation of all SNOTEL stations used in the study.

All of the stations have elevations of more than 2,600 meters, and more than 70% of the stations have elevations of more than 3,000 meters. All the stations are located at a higher elevation because snow accumulation is predominant in this area. However, there are still some differences in snow characteristics between stations.

3.2. *Input Data*

3.2.1. SNOTEL

Snow Telemetry Network (SNOTEL) is a network of snow and climate monitoring stations operated by the Natural Resources Conservation Service (NRCS) of the United States Department of Agriculture in the Western United States (SNOTEL Sensor Data (usda.gov)). SNOTEL has been collecting different climate information including SWE and snow depth since the 1960s [Serreze et al., 1999]. It provides publicly available and detailed snow cover data to scholars around the world and permits users to access a great amount of data at daily and hourly resolutions [Avanzi et al., 2014]. More than 40 years of complete time series, large amounts of data from more than 800 sites, and high temporal resolution of measurements (daily and hourly) make SNOTEL an excellent source of research data.

A typical SNOTEL station is equipped with multiple different sensors designed for several different climate data. An ultrasonic depth sensor is used to collect snow depth, a thermistor can measure the air temperature and a rain gauge is designed for annual total precipitation [Serreze et al., 1999]. The data of SWE used in this study is collected by a sensor called a "snow pillow". A snow pillow contains a $1m^2$ square pillow filled with a low freezing point liquid and a pressure transducer. It measures SWE by changes in pressure on sensors caused by overlying snow [Castle, 1969]. The general structure of a SNOTEL station is shown in Figure 3.2-2.

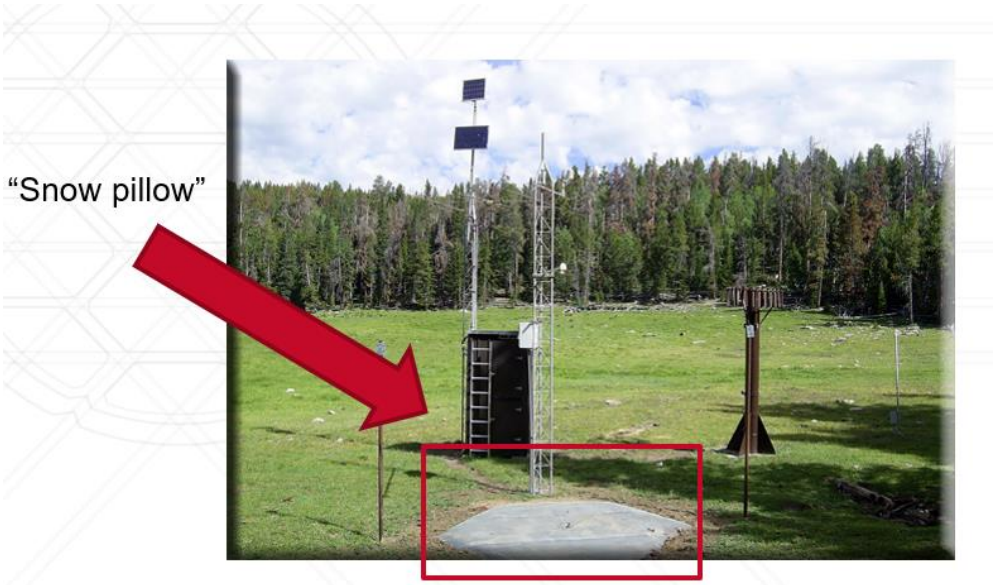


Figure 3.2-2 SNOTEL Station with Snow Pillow Measuring SWE (adapted from Bryan Allegretto,2020)

Factors such as damage from wildlife attacks and low battery power could cause SNOTEL stations to provide incorrect SWE data. Therefore, the quality control of SNOTEL data should be carried out first. By creating time series figures of SWE like

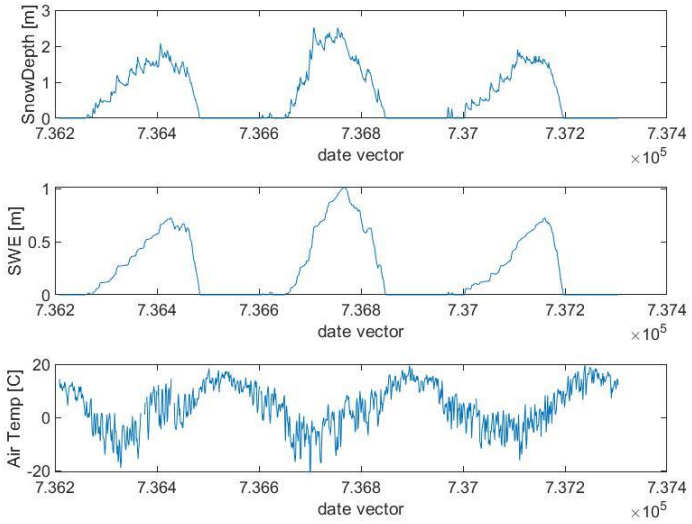


Figure 3.2-1 Time Series Plots of SWE Snow Depth and Air Temperature for Station 64

Figure 3.2-1 for all used stations, problematic data can be easily eliminated. Through

the examination of the time series plots of all 111 sites, the data of all used stations are in a qualified state.

3.2.2. AMSR2 Passive Microwave Brightness Temperature

The Advanced Microwave Scanning Radiometer 2 (AMSR2) was lunched abroad in May 18, 2012 [AMSR2 | Earthdata (nasa.gov)]. As the successor of to the Advanced Microwave Scanning Radiometer for Earth Observing Systems (AMSR-E), AMSR2 is also a passive microwave radiometer with 6 frequencies and 2 polarizations for a total of 12 channels, Figure 3.2-3. Some instrument specifications are presented in Table 3.2-

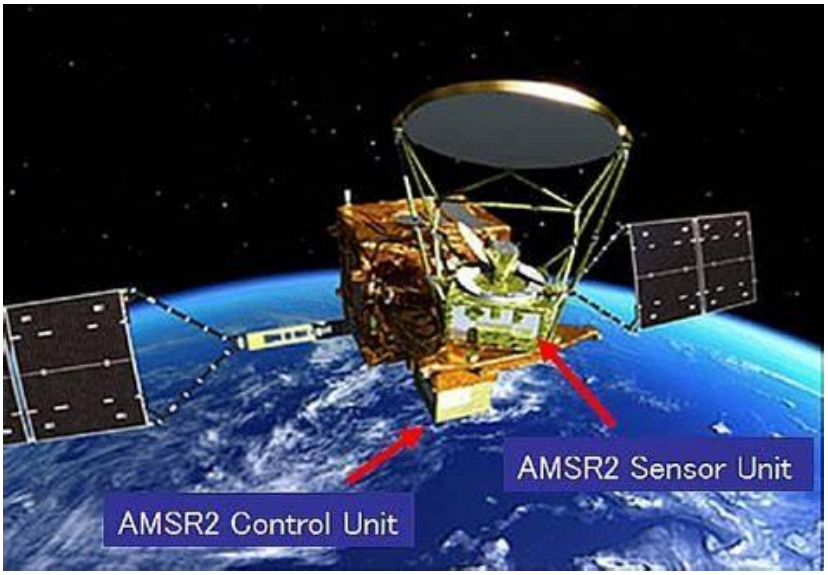


Figure 3.2-3 The Appearance of AMSR2 Sensor [image credit: JAXA, GCOM-W1]

1.

10.65GHz, 18.7GHz and 36.5GHz frequency bands are considered to be suitable frequencies for SWE retrieval. Microwave radiation at these frequencies can pass

through the atmosphere with relatively small attenuation [Chang et al., 1982]. AMSR2 observation is not only used for snow retrieval, it is also used for retrievals of other geophysical parameters such as precipitation, ocean surface temperature, soil moisture and water vapor [Kolassa et al., 2017].

Table 3-1 AMSR2 Instrument Specifications [NASA,2021]

Platform	GCOM-W1
Launch Date	May 18,2012
Swath width	1445km
Altitude	699.6 km (on equator)
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)

In this paper, the brightness temperature measurements come from the gridded Level-3 land surface product (AE_Land3) [Ashcroft，2000]. There are two available daily measurements at each frequency from descending (night) and ascending (day) overpasses [Tedesco and Jeyaratnam, 2016]. This study utilizes descending data as the input data because measurements from nighttime AMSR2 overpasses minimize the impact of liquid water present in the snow [Forman et al. 2013]. Data from all three frequencies are resampled into global cylindrical EASE-Grid cell spacing at a 25km × 25km resolution grid for consistency [Tedesco and Wang, 2006].

The 6.9 GHz channel and the 89.0 GHz channel are not used in this study. The 6.9 GHz channel has a relatively coarse resolution (75km × 43km), which will reduce the

accuracy of the machine learning algorithm and make it difficult to evaluate the performance of different algorithms on spatial differences in this paper. Although higher frequency channels such as 89 GHz mean higher resolution, they are usually designed for atmospheric research and are heavily influenced by water vapor and clouds [Durand and Margulis, 2007]. Some studies use 89GHz to estimate surface properties of snow such as surface grain size because it is more sensitive to surface properties of snow than snow depth [Durand et al., 2008]. The 23.9 GHz channel is also avoided being used in this study since it is highly affected by water vapor existing in atmosphere [Special issue on microwave radiometry and remote sensing applications, 2004]. In addition, as Kelly [2009] suggested, the difference of value between 10.65 GHz and 36.5 GHz can best retrieve medium snow SWE and vertically polarized Tb at 10.65 GHz and 18.7 GHz is more suitable for calculation of deeper snow SWE.

Therefore, this study uses both vertically and horizontally polarized Tb measurements at 10.65GHz, 18.7GHz, and 36.5GHz from AMSR2 in western Colorado from 1 September 2015 to 1 September 2018 as the training inputs, and uses SNOTEL SWE data as training targets, Table 3.2-2

Table 3-2 Machine Learning Inputs Symbols and Units

Inputs	Symbol	Unit
Brightness temperature at 10.65 GHz, V-polarization	10V	K
Brightness temperature at 10.65 GHz, H-polarization	10H	K
Brightness temperature at 18.7 GHz, V-polarization	18V	K

Brightness temperature at 18.7GHz, H-polarization	18H	K
Brightness temperature at 36.5 GHz, V-polarization	36V	K
Brightness temperature at 36.5 GHz, H-polarization	36H	K
Output	Symbol	Unit
Snow water equivalent	SWE	m

3.3. Training Data Selection

For most machine learning algorithms, it is essential to have sufficient training sets [Ruppert, 2004]. If the training set data is too small, the model established by the machine learning algorithm will have poor performance, and the final prediction result will also have large errors [Ben-Hur and Weston, 2010]. However, the training set with excessive data increases the computational cost, which makes the prediction efficiency of the model low. Therefore, how to set an appropriate size for the training set of machine learning algorithms has become the first problem to be faced in this study. This paper sets up a test scheme in order to select the most suitable number of training stations.

The test starts by setting a range to filter the training dataset, and a SNOTEL station is considered to be a usable training station if it is located within a 75-km radius of a circle centered around the test station. Turn all 111 SNOTEL stations into training stations and find the number of training stations each station has that can be used. According to the data obtained, Station 64 has the largest number of training stations available. For n from 1 to 35, this test randomly draws n of the training stations of

Station 64 as the training set and uses the LSTM algorithm to predict the SWE at station 64. The mean square error (MSE) between the predicted results and the actual SNOTEL data is then calculated, and its relationship with n is plotted. Repeating the above process 10 times, a boxplot of MSE and the number of training stations n can be generated.

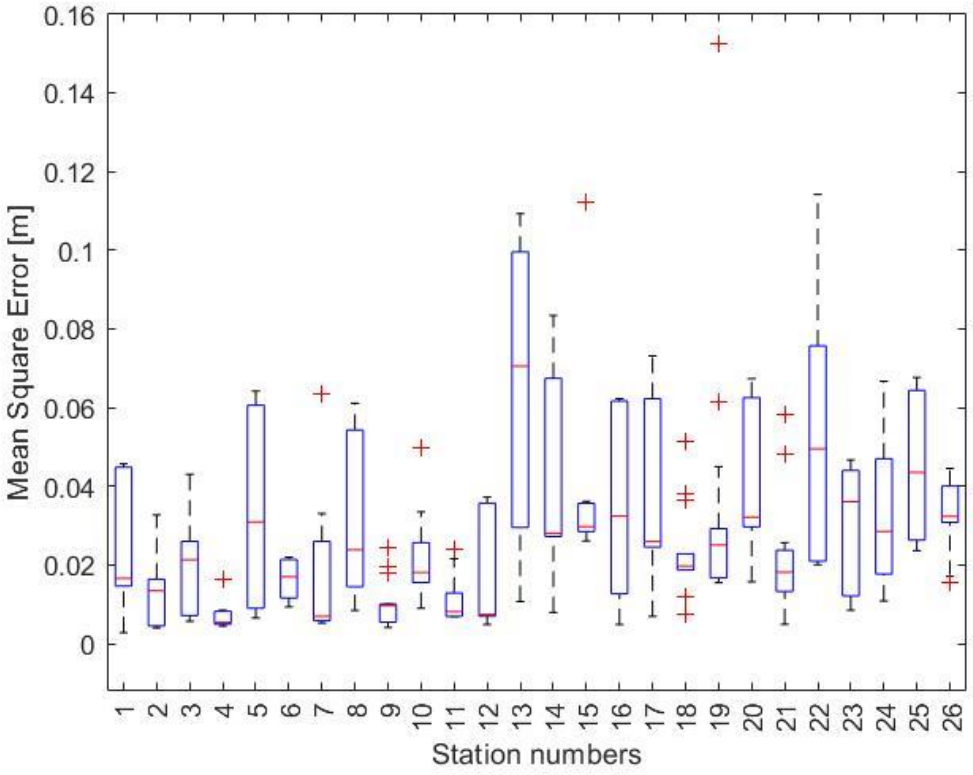


Figure 3.3-1 Mean Square Error during Station number tests

As can be seen from the Figure 3.3-1, when the number of training stations is small, the value of MSE changes greatly, and there is a big difference between the values of each test. This is because when the number of test stations is too small, if you happen to encounter a training station that has very similar snow data to the test station, it will lead to a very low MSE, while if you encounter a training station that has very different

snow data to the test station, the model will completely fail and produce a very large deviation. In addition, although some box plots such as stations 15 and 18 appear to have a good distribution, this is due to the presence of many outlier values (red crosses) beyond the confidence interval. After stripping these data with a large number of anomalies, we can observe that when the number of training stations exceeds 23, the overall prediction MSE tends to be stable and no longer has a large fluctuation. At the same time, the average value of MSE is also at a low level, and even if the number of sites is increased, the performance of the algorithm is difficult to be significantly improved. This is the method of determining how much input training data is enough to properly train the algorithm.

Once the appropriate training set size has been decided, it is time to filter the appropriate test stations. A suitable test station should meet the following conditions: at least 23 training stations can be used within a 75-km radius of the test station as the center. After screening, 39 test stations that can be used are selected in this paper. Their station numbers are 2, 3, 4, 7, 8, 9, 16, 18, 27, 29, 30, 31, 40, 41, 49, 50, 51, 52, 54, 57, 62, 67, 68, 69, 70, 75, 78, 86, 87, 88, 92, 95, 96, 98, 103, 107, 108, 109, 111.

3.4. Training Process

This study utilized the Machine Learning Toolbox and Deep Learning Toolbox provided by Matlab© to generate all three machine learning algorithms. Matlab© has powerful matrix computing capabilities, user-friendly code writing interface and

detailed interpretation help functions. Hence, Matlab© is an ideal tool for working with

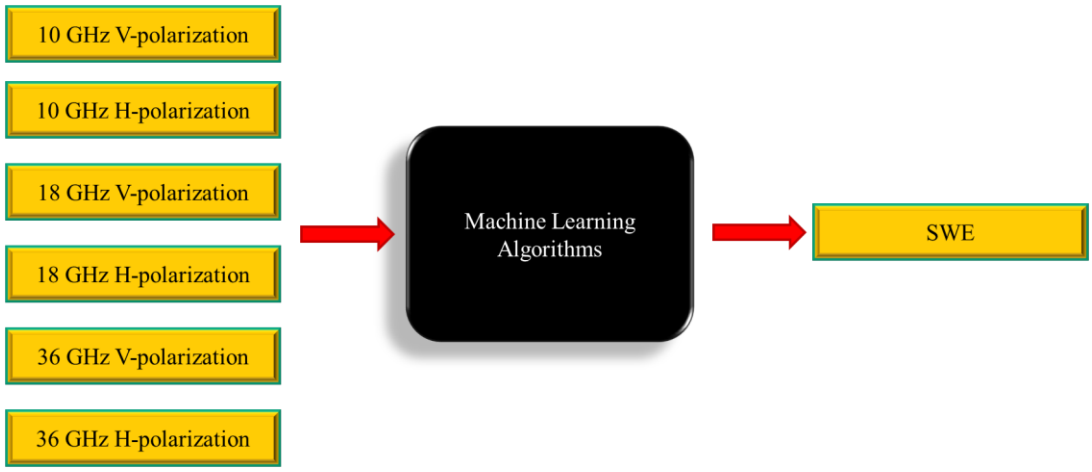


Figure 3.4-1 The Simplified Structure Plots of All Three Different Machine Learning Algorithms

machine learning algorithms. If the training process of machine learning is compared to a black box, the basic structure of three different algorithms is the same, with six inputs, one training process algorithm, and one output data (Figure 3.4.1).

3.4.1. Normalization

Input normalization is significant for all three machine learning algorithms since the training process and the determination of weight vectors are sensitive to the way features are scaled [BenHur and Weston, 2010]. In this study, there are six input vectors and one output vector in measured in a different scale with a different unit and has a different range of possible values. Normalizing all vectors into a common range has plenty of benefits. It can avoid greater numeric ranges dominating those in smaller ranges [Hsu et al. 2003]. Also, it can overcome the difficulty in calculating inner products of feature vectors when the input data has a very different range of value (which is not the case in this study) [Hsu et al. 2003]. The normalization methods

utilized in this study can be expressed by the following example. The normalization algorithms of $Tb_{36,V}$ data for a single station within a daily training period time can be written as:

$$\widehat{xi} = \frac{xi - min\,(x)}{max\,(x) - min(x)} * (b - a)\; + a \tag{3.4.1}$$

where $\widehat{xi}$ is the normalized brightness temperature data at 36.5 GHz vertically [dimensionless]; xi is the original $Tb_{36,V}$ data [K] at day i; min(x) is the minimum $Tb_{36,V}$ data across all three years daily measurements; max(x) is the maximum $Tb_{36,V}$ data across all three years daily measurements; a is the specified lower bound of the scaling range; and b is the upper bound of the defined range of scaling.

In order to define the scaling intervals and the upper and lower bound of scaling range, this study uses the methods of Sarle [1997] by scaling the data with the midrange of 0 and the range of 2 (in this study is [-1,1]). Sarle [1997] also mentioned another way in his study, which is to scale the data with the mean of zero and the standard deviation of one. In this study, as daily SWE is an irregular curve, the influence of extreme values and abnormal changes on the prediction results of the algorithm will be eliminated in this way. However, these extreme values and abnormal changes have important physical meaning in snow research (e.g., drought year versus wet year), and they may represent important environmental events and evidence of climate change in some years. As a result, this study chooses the first way to normalize the data in order to maintain the meaning of these extreme values and abnormal changes. In addition, the prediction data needs to be unscaled when it is output as the result, and the unnormalizing function

is the inverse function of normalizing function.

3.4.2. SVM

There are two parameters needed to be determined in the training process: C and γ. C is the penalty parameter which is determined by the range of training targets in this case as:

$$C = max(Y) - min(Y)$$

where Y is the raining targets which is the SWE [m] in this study. The selection of γ is based on a method called cross-validation [Hsu et al. 2003].

Unlike LSTM algorithm that can adjust parameters automatically during the training process, SVM needs to adjust the hyper-parameter γ manually. K-fold cross-validation method is introduced to SVM in order to find out the best hyper-parameters for the prediction model [Chan et al. 2013]. The K-fold cross-validation divides the training set into K subsets of equal size. One subset is extracted during training and used as a validation set, the other (K-1) subsets are used as the training set [Hsu et al. 2003]. The accuracy of cross-validation is calculated as the percentage of data that is in the 95 percent confidence interval of the regression. The parameters with the highest accuracy will be selected as the hyper-parameters of the prediction model. Based on the research of Forman and Xue [2017], when $K \geq 5$, the SVM algorithm performs best. Since this study uses a similar training data set, the value of K is chosen to be 5 as well in this study.

3.4.3. LSTM

The LSTM training was conducted based on the time step learning cycle to minimize the MSE between the LSTM-estimated SWE and the SNOTEL SWE value which is the training target in the algorithm. For a single test station, a training set {(P1, Y1) …… (P_n, Y_n)} including n pairs of input space P and output training space Y using the same time period from all of three years is given to the training process. In the training process, MSE for a single output neuron can be expressed as:

$$MSE = \frac{1}{2}\sum_{i=1}^{n} ||Pi - Yi||^2$$

where Pi is the i_{th} LSTM-estimated value of SWE [m]; Yi is the i_{th} SNOTEL measured value of SWE [m]; n is the total number of evaluated time steps; and ||.|| is the Euclidean norm operator.

In order to minimize the MSE, the initial weights at the first training time step are randomly selected. After that, the Levenberg-Marquardt optimization algorithm [Marquardt, 1963] is used to update the weights vectors during the training process until it reaches minimum for each output neuron. The Matlab© machine learning toolbox uses RMSE which is square root of MSE instead of MSE, shown as Figure 3.4-2.

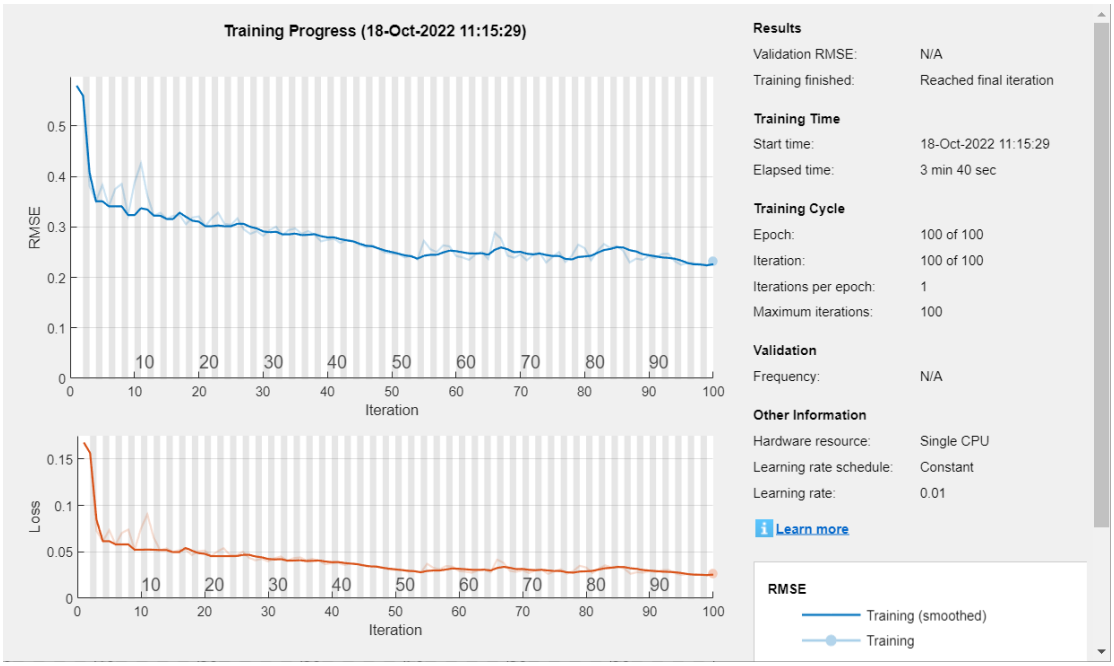


Figure 3.4-2 LSTM training process in Matlab©

3.4.4. Evaluation Methods

Statistical

There are many ways to evaluate the predictive performance of machine learning algorithms. From the statistical point of view, simple evaluation parameters such as bias and root mean square error, sensitivity analysis, and correlation analysis can be used to comprehensively evaluate the model from different perspectives. In this paper, mean square error (MSE), time-averaged average bias, and Pearson correlation coefficient are used to make an overall evaluation of the final prediction results. However, only using a single statistical parameter cannot fully evaluate the performance of the model. MSE and bias can be used to evaluate the deviation degree of the predicted value of the model, and the correlation coefficient can represent the correlation between the predicted value

and the actual observed value. The combination of the three can give a reasonable comprehensive evaluation of the model.

Plot

In addition to the statistical parameters above, images are also a good way to evaluate the performance of an algorithm. Images can elicit important information that statistical parameters cannot, and may lead to large differences in evaluation results. In this study, multiple evaluation maps were introduced to form a comprehensive analysis with statistical parameters. The scatter plot of MSE with respect to the training target is used to observe whether the distribution of MSE is independent or affected by the training target data. If its correlation with the training target is lower, it means that the performance of the algorithm is more reasonable.

The time series plots are also used in this study. By placing the prediction data of the algorithm and SNOTEL ground measurement data in the same time series graph, the performance and limitations of the prediction algorithm can be directly observed. The Y prediction figure is also used in this study. The SWE data estimated by the algorithm is used as the Y-axis, and the actual SNOTEL data is used as the X-axis to make the scatter chart. A reasonable and relatively well-behaved algorithm would observe that the scatter points are fairly evenly distributed on both sides of the line Y=X.

Chapter4.RESULTS AND ANALYSES

4.1.Spatial Analysis

The study domain of this paper, Western Colorado, is a combination of mountains and plateaus, and the varied topography causes significant differences in important factors such as snow cover and vegetation types among different stations. Therefore, the spatial analysis of the whole study domain is indispensable. In this paper, three different machine learning algorithms are used to test the 38 selected available stations, and regional MSE maps of the predicted results are produced. Since there is no snow cover in summer, there is still a large amount of vegetation and high soil moisture, so the introduction of summer data will affect the performance of the algorithm model. As a result, the snow season of this year is set from September 13 to April 1 of the following year in this study, and the data of the snow season of three years are selected as the training set and test set.

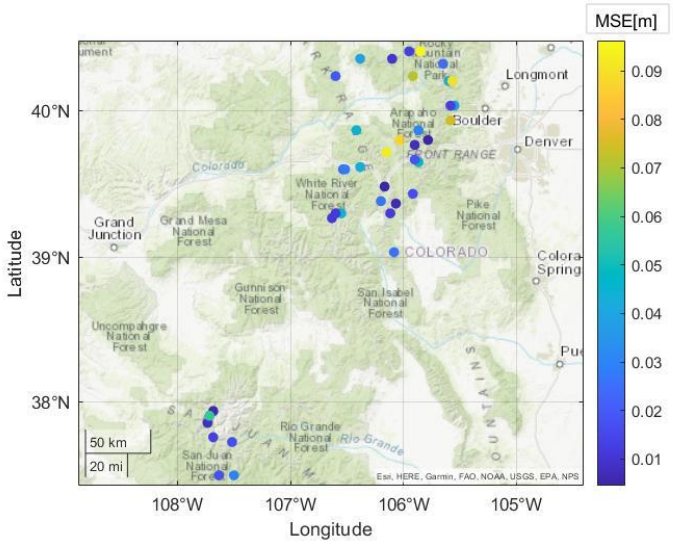


Figure 4.1-1 Regional MSE of Selected Stations Using LSTM

Figure 4.1-1 shows the MSE of all selected test stations in study domain calculated by LSTM. By combining the above figure with Figure 3.1-2 and Figure 3.1-3, it is seen that the stations in the grassland and shrub areas at high altitudes generally have a low mean square error, ranging from 0 to 0.03 meters, while the stations in the forest areas at high altitudes generally have some slightly higher MSE, with the value generally distributed around 0.06 meters. In the forest area of low elevation, the MSE was much higher than that of other sites, with values exceeding 0.1, showing a poor prediction result.

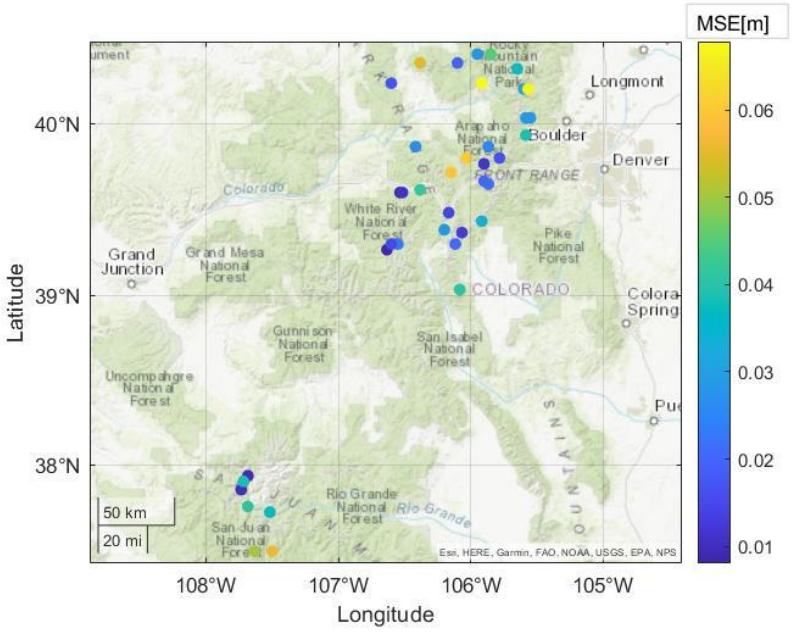


Figure 4.1-2 Regional MSE of Selected Stations Using SVM

Figure 4.1-2 shows the MSE of all selected test stations in study domain calculated with the trained SVM. From the overall point of view, the MSE of the predicted value of the SVM algorithm is smaller than that of LSTM algorithm. However, SVM and LSTM algorithms maintain a high degree of consistency in the MSE differences of different elevations and land surface types. In the high-altitude grassland and shrub

areas, the MSE is relatively low, ranging from 0 to 0.02 meters, which is very similar to the result of LSTM algorithm. In high-altitude forest areas, the range of MSE is about 0.04 meters, which is relatively small compared with the results of LSTM algorithm. However, when it comes to the low-altitude forest area, the MSE of the predicted value of SVM is significantly different from that of LSTM algorithm, which is only about half of that of LSTM algorithm. This indicates that SVM algorithm may have better performance than LSTM in low-altitude forest areas.

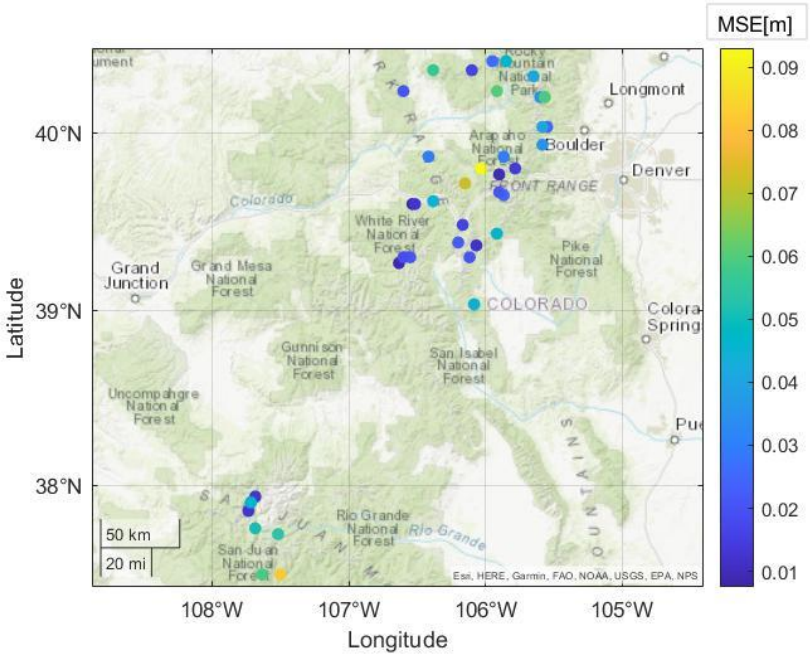


Figure 4.1-3 Regional MSE of Selected Stations Using GP Regression

Figure 4.1-3 shows the MSE of all selected test stations in study domain calculated by GP regression. GP regression and SVM regression have almost the same MSE for predicted SWE. However, it is worth noting that although the MSE of GP regression and SVM regression are almost the same in the grassland and shrub areas at high altitude and the forest areas at high altitude, the MSE of GP regression will be slightly

higher than that of SVM regression in the forest areas at low altitude. It can be speculated that, compared with SVM algorithm, the performance of GP regression is more affected by vegetation, snow depth and other factors.

In general, there are significant differences in the performance of these three algorithms under different elevations and surface types, which is determined by the PMW remote sensing technology used in this paper and the physical characteristics of snow cover. It also reflects that the machine learning results must be consistent with the actual physical principles before they can be evaluated and used. At the same time, the difference of MSE also shows that the three different algorithms have their own performance characteristics. The complete tables of statistical values of all test stations and their boxplots are available in Appendix A. In the following discussion, typical stations in different elevations and with different land surface types will be selected for more specific analysis and comparison.

4.1.1. SWE in the Regions with High Elevation and Low Forest Cover

As mentioned in Chapter 2, vegetation and snow depth are factors which have significant impact on PMW snow remote sensing. Low forest cover regions in this study means the areas whose land surface type is grass or shrub. Since most of the stations used in this paper are located in mountainous areas with high altitude, and their elevation generally exceeds 2500m and snow depth is generally large, the stations with altitude over 3000m are set as high elevation stations. This study choose station 111 as

the typical station which has high elevation and low forest cover. The basic situation of station 111 is shown below:

Table 4-1 Basic Information of Station 111

Station number	latitude	longitude	elevation	Land surface type
111	39.7667 N	105.9 W	3169.9m	Grass

Mean square error, bias, and correlation coefficient are three statistical number this study used to evaluate the performance of different machine learning algorithms, which are shown as bar plots below.

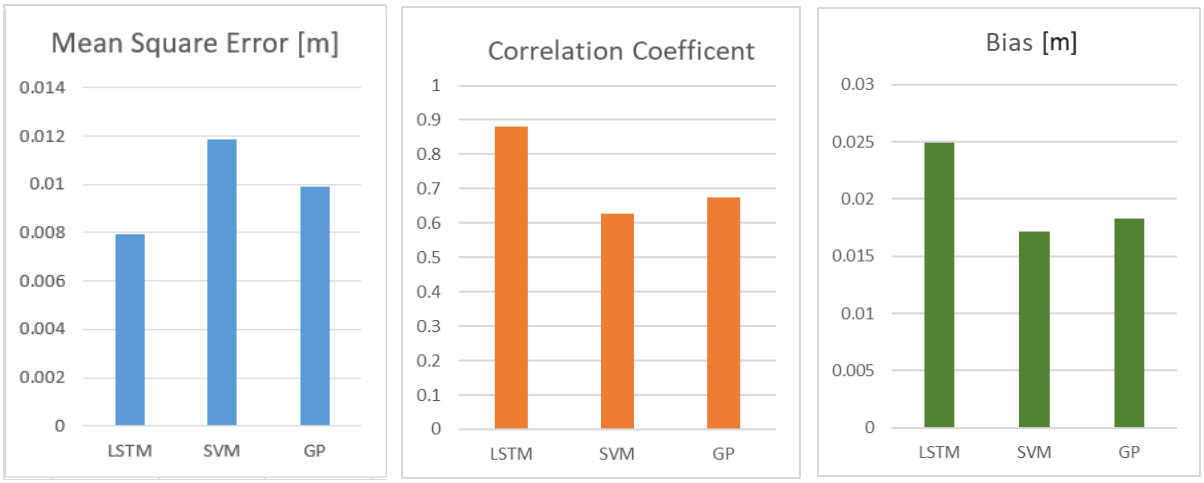


Figure 4.1-4 Statistical Parameters of Station 111 for All Three Algorithms

According to the bar plots of statistical data (Figure 4.1-4), the LSTM performs better in this region. Compared with the other two algorithms, its MSE is lower, but the average error was higher than the other two algorithms. LSTM has the highest correlation coefficient value of 0.88, which means the prediction SWE calculated by LSTM Is highly correlated with the actual SNOTEL SWE data and causes both to have very similar trends. SVM and GP regression almost have the same performance, the

values of their three statistical parameters are quite similar. The subplot representing bias shows the absolute value of the bias, and it can be observed from the subplot that the predicted values of the three algorithms are about 0.02m smaller than the actual

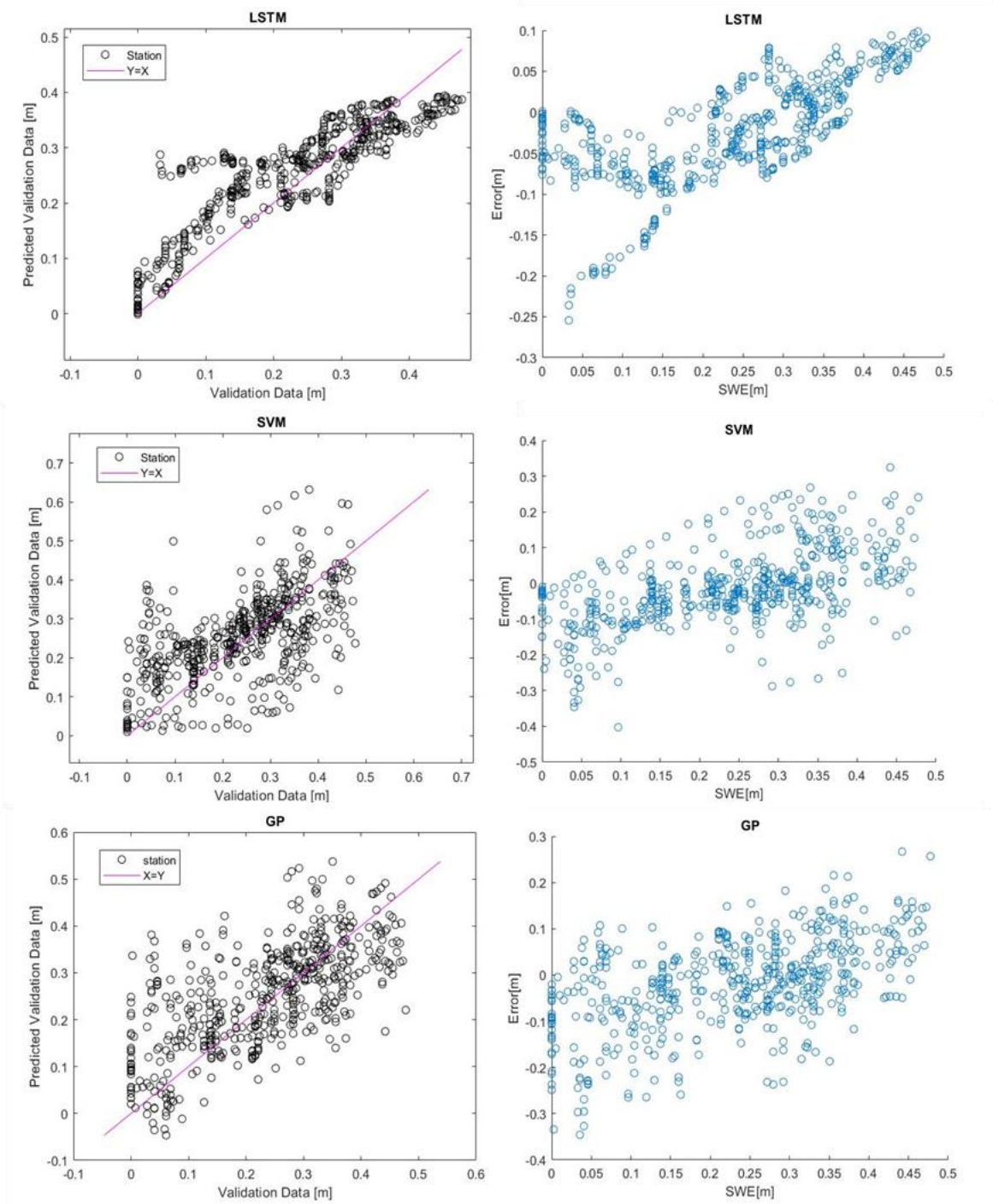


Figure 4.1-5 Prediction and Validation Plots and MSE Dot Plots for All Three Algorithms in Station 111

SNOTEL data on average.

In addition, the scatter plot of the predicted value of the algorithm on the SNOTEL data and the scatter plot of the errors (bias) in SWE are used to evaluate the performance of the algorithm (Figure 4.1-5).

As can be observed from the scatter plot of the predicted value of the algorithm on the SNOTEL data, the predicted SWE value from LSTM is more concentrated on both sides of the line than the other two algorithms. This is consistent with the data of statistical parameters and explains why the MSE of LSTM algorithm is lower and the Pearson correlation coefficient is higher. Compared with GP regression, SVM regression has a higher degree of concentration, but at the same time, it also has many outliers (i.e., predicted SWE and the actual SNOTEL SWE have a large deviation), which leads to the phenomenon of high MSE in the statistical parameters.

From the scatter plots of the errors (bias) on the SWE on the right, the deviation of an algorithm model should reflect independence and randomness, specifically, it should be relatively evenly distributed on both sides of 0 and not change with the change of the value of the training target. It can be clearly observed that the deviation of the LSTM algorithm is almost in the region less than 0 when the SWE is less than 0.25m while most of it becomes in the region more than 0 when the SWE is more than 0.3m. That is, intends to be larger with an increase in SWE. This shows that the LSTM algorithm has some defects. For SVM regression, it has the same problems as the LSTM algorithm. When SWE is less than 0.3 meters, the bias is generally less than 0. When

SWE is more than 0.3 meters, the bias is generally greater than 0. In this respect, the GP regression algorithm performs better. Although when SWE is less than 0.2m, the deviation of GP regression is more distributed in the region less than 0, but when SWE is in the moderate region between 0.2m and 0.4m, the deviation of GP regression shows a relatively uniform distribution on both sides of 0 and does not gradually increase with the increase of SWE.

Times series analysis is another important tool to evaluate the performance of algorithms (Figure 4.1-6). From the time series diagram, it can be clearly found that LSTM has an obvious advantage compared with the other two algorithms, that is, less noise. However, LSTM's prediction of the first half of the snow accumulation season in the first year has a larger deviation than the other two algorithms. On the contrary, SVM regression and GP regression algorithms showed relatively accurate predictions during the snow accumulation season of the first years while relatively large deviations occurred in the snow accumulation season of the third year. Compared with the GP regression algorithm, SVM regression shows a higher fit with the actual SNOTEL SWE data during the first two years, but its disadvantage is that a large amount of abnormal noise appears in the peak period of snow cover when SWE is large.

In general, in high-altitude areas with low vegetation cover, the predicted SWE values of the three algorithms show relatively good agreement with the actual SNOTEL SWE measurements. The three different algorithms also show their respective strengths and weaknesses in this region. The predicted value of LSTM has a high correlation with

the actual data and has less noise, while the predicted value of GP regression and SVM in the first year has a higher degree of agreement with the actual data.

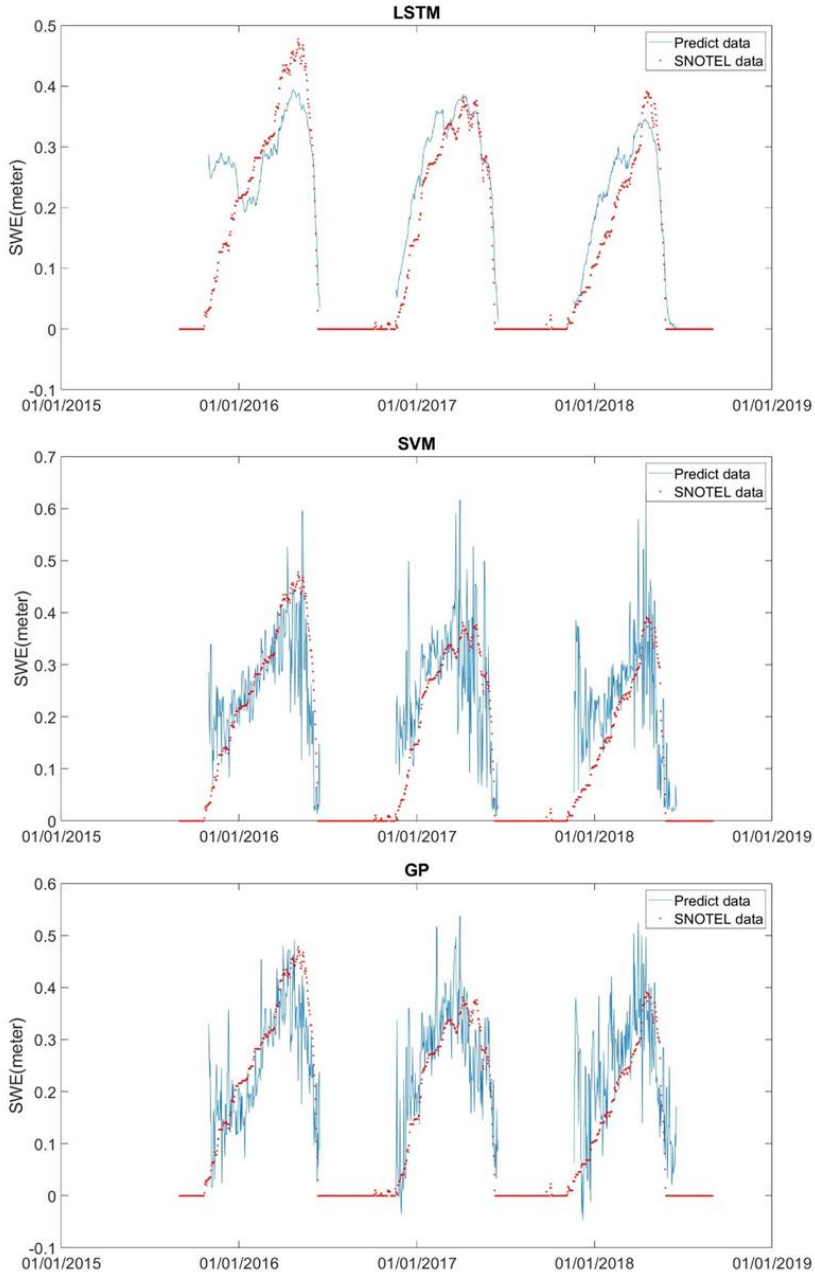


Figure 4.1-6 Time Series Plots of Station 111

4.1.2. SWE in the Regions with High Elevation and High Forest Cover

This study choose station #3 as the typical station that has a high elevation and

relatively dense forest cover. The basic characteristics of station #3 is shown below:

Table 4-2 Basic Information of Station 3

Station number	latitude	longitude	elevation	Land surface type
3	40.35 N	106.383 W	3340.6m	Forest

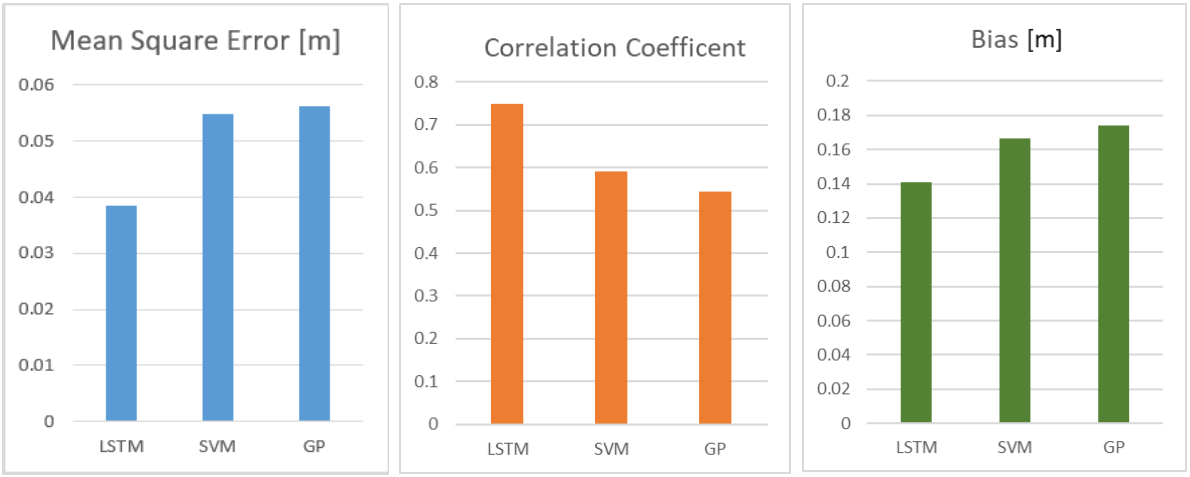


Figure 4.1-7 Statistical Parameters of Station 3 for All Three Algorithms

According to the analysis of the bar plots of statistical data (Figure 4.1-7), the LSTM algorithm still maintains the lowest MSE and the highest correlation coefficient among the three algorithms. Different from high-altitude areas with low vegetation coverage, SVM regression yields lower MSE and higher correlation coefficient than GP regression in this area. However, the MSE values of all three algorithms are about 10 times greater when compared with the results obtained at high altitudes with low to moderate forest cover. At the same time, their correlation coefficient drops by about 10 percent. The changes of the three statistical parameters show that the performance of all algorithms decreases to some extent in the forest area at high altitude. This is

consistent with the conclusion that vegetation has a significant influence on PMW snow remote sensing discussed in Chapter 2 of this study.

Next, the scatter plot of the predicted values of each algorithm relative to the

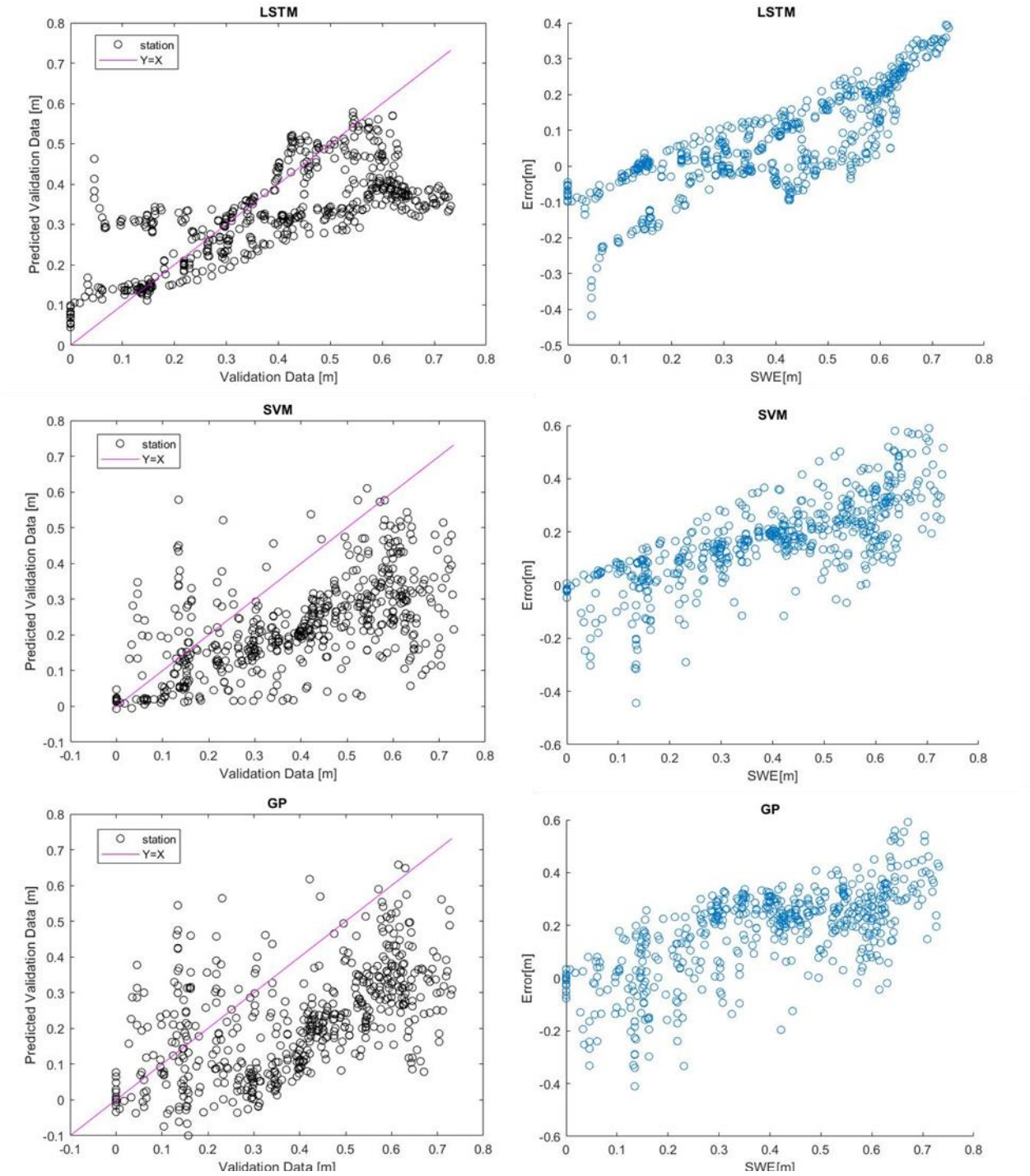


Figure 4.1-8 Prediction and Validation Plots and MSE Dot Plots for All Three Algorithms in Station 3

SNOTEL data and the scatter plot of the errors (bias) on the SWE are used for further analysis and comparison (Figure 4.1-8). As can be seen from the scatter plot of the predicted value of the algorithm with the SNOTEL measurements, the SWE predictions from all three algorithms are no longer uniformly concentrated on both sides of the line (predicted SWE equals to actual SNOTEL SWE) in the forest area at high altitude due to the influence of vegetation cover. The predicted SWE values are generally lower than the actual SNOTEL SWE measurements, and the deviation tends to increase gradually as the snow gets deeper. This is consistent with the results shown in the scatter plot of the errors (bias) on the SWE on the right. Nevertheless, the LSTM still has a better concentration than the other two algorithms, which is also consistent with its highest correlation coefficient. In addition, when SWE is less than 0.3m, SVM regression shows better concentration than the GP algorithm, but when SWE is higher, both of them show poor concentration.

In the time series plots (Figure 4.1-9), the effects of vegetation type and snow depth on PMW snow remote sensing and algorithm accuracy are more intuitively displayed. Compared with the results from high-altitude areas with low forest cover, the LSTM algorithm shows significantly more noise in the predicted values, and the agreement with the actual SNOTEL SWE measurements is also reduced. Moreover, when SWE exceeds 0.6m, the LSTM algorithm can no longer predict as accurately as the other two algorithms.

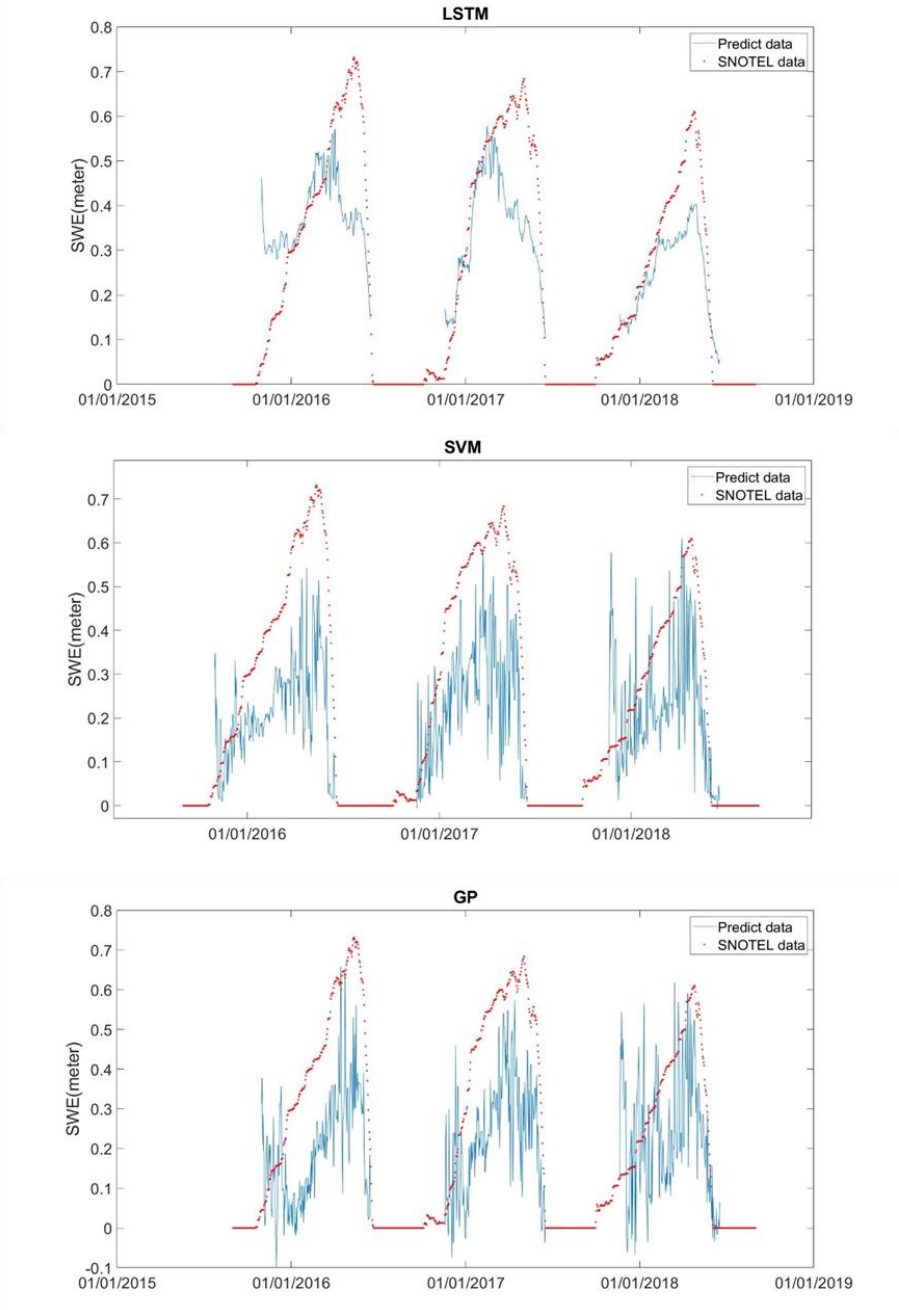


Figure 4.1-9 Time Series Plots of Station 3

This is because when the snow is too deep, the brightness temperature data generated by the PMW sensor no longer captures the assumed physics of volume scattering, which leads to the failure of the LSTM algorithm. This is because during the snowmelt season, the snow layer contains a large amount of liquid water, and the

different emissivity between wet snow and dry snow will degrade the quality of the SWE prediction.

By comparing three different algorithms, the LSTM still has the problem of low accuracy in predicting at the first half of the snow accumulation season during the first year, but it also maintains its advantages of less noise and high prediction accuracy in the last two years. Compared with GP regression, SVM regression has better performance in this region. In contrast to the results in high-altitude and low-vegetation areas, the SVM regression algorithm had less noise in this area than the GP regression algorithm, and the predicted values in the first year had a higher agreement with the actual SNOTEL SWE measurements. It is worth mentioning that the LSTM algorithm showed its ability to capture the interannual variability of SWE in the third year's prediction, while SVM and GP regression did not show such skill.

In general, the performance of all three machine learning algorithms in high-altitude forest areas is degraded due to the impact of vegetation and deep snow. Overall, the LSTM is still the best performing algorithm, while GP regression has the poorest performance. However, the LSTM's poor accuracy in the first half of the year persisted, and neither of them is able to provide a more accurate forecast of the snow melt season. In the snow accumulation season, LSTM can give a good prediction accuracy when SWE is less than 0.6 meters, while SVM and GP regression can only have a relatively good performance when SWE is less than 0.4 meters.

4.1.3. SWE in the Regions with Low Elevation and Low Forest Cover

This study choose station #8 as the typical station which has low elevation and low forest cover. The basic situation of station #8 is shown below:

Table 4-3 Basic Information of Station 8

Station number	latitude	longitude	elevation	Land surface type
8	39.6 N	106.52 W	2590.8m	Shrub

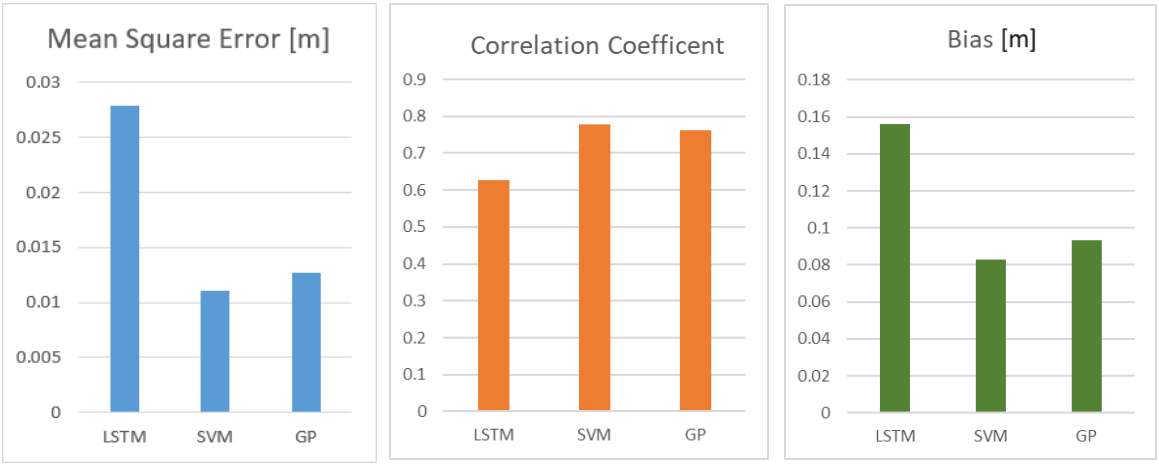


Figure 4.1-10 Statistical Parameters of Station 8 for All Three Algorithms

Figure 4.1-10 shows the statistical parameters of all three algorithms in low altitude and low forest cover regions. Different from the high-altitude regions, the LSTM algorithm has the highest MSE and the lowest correlation coefficient among the three algorithms in the low-altitude and low-forest cover areas. Compared with SVM regression, the MSE of LSTM is more than three times that of SVM, and the correlation coefficient is about 15% less. As the algorithm with the best performance of statistical parameters in this area, SVM regression significantly reduces MSE and improves the correlation coefficient compared to high-altitude forest areas, which indicates that

vegetation type has a significant impact on the results of SVM regression, and the GP regression algorithm also has the same characteristics. For LSTM, MSE and bias decreased to a certain extent compared with high-altitude forest areas, but the correlation coefficient was not improved. This shows that different from the characteristics of SVM and GP regression, which are more affected by vegetation and less affected by elevation, elevation and vegetation both have significant effects on the performance of LSTM algorithm.

Figure 4.1-11 shows that the predicted value of LSTM in the low-altitude areas with low vegetation cover has a worse performance than that in the high-altitude forest areas, and the predicted value has a large deviation from the actual SNOTEL SWE measurements, especially when the SWE value is lower than 0.2 meters. In contrast, the results of both SVM and GP regression are more concentrated near the line (predicted SWE equals to actual SNOTEL SWE), although the predicted SWE value is generally slightly higher than the actual SWE. According to the scatter plots of the errors (bias) on the SWE on the right, the deviations of all three algorithms have a trend of gradually decreasing with an increase in SWE. This indicates that the predicted SWE from the machine learning algorithms will be less accurate when the snow is shallow.

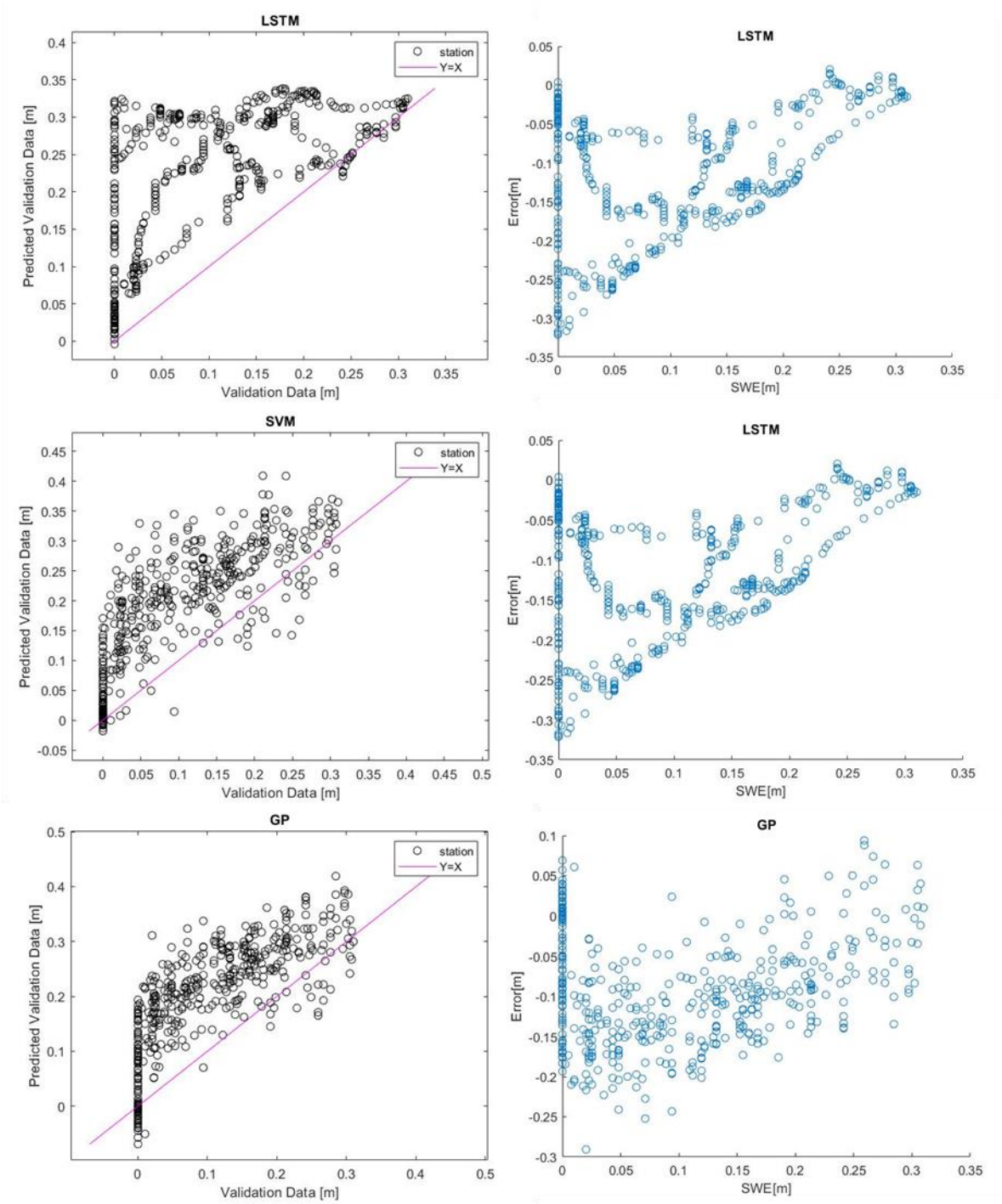


Figure 4.1-11 Prediction and Validation Plots and MSE Dot Plots for All Three Algorithms in Station 8

Different from the high-altitude areas, SVM regression and GP regression showed better prediction accuracy during the snow accumulation season than in the low-altitude area. In particular, the predicted SWE value using SVM regression during the second

snow season showed good agreement with the actual data, and a prediction accuracy comparable to that of high-altitude and low-vegetation areas. However, when it comes to the third (dry) year, as with the LSTM algorithm, SVM regression and GP regression cannot accurately predict SWE, and the predicted value given by the algorithm is much greater than the actual SWE. At the same time, the LSTM algorithm also lost its ability to capture the inter-annual variation of the SWE in this region.

In general, SVM regression has the best performance in low altitude areas with low vegetation cover while LSTM algorithm has the worst prediction accuracy. However, all three algorithms, in general, cannot make accurate predictions when SWE is less than 0.15 meters. This is because when the snow is shallow, volume scattering is minimal and the limitations of PMW remote sensing of snow are apparent. As a result, the machine learning algorithms have relatively little information content from the PMW observations to use during SWE prediction.

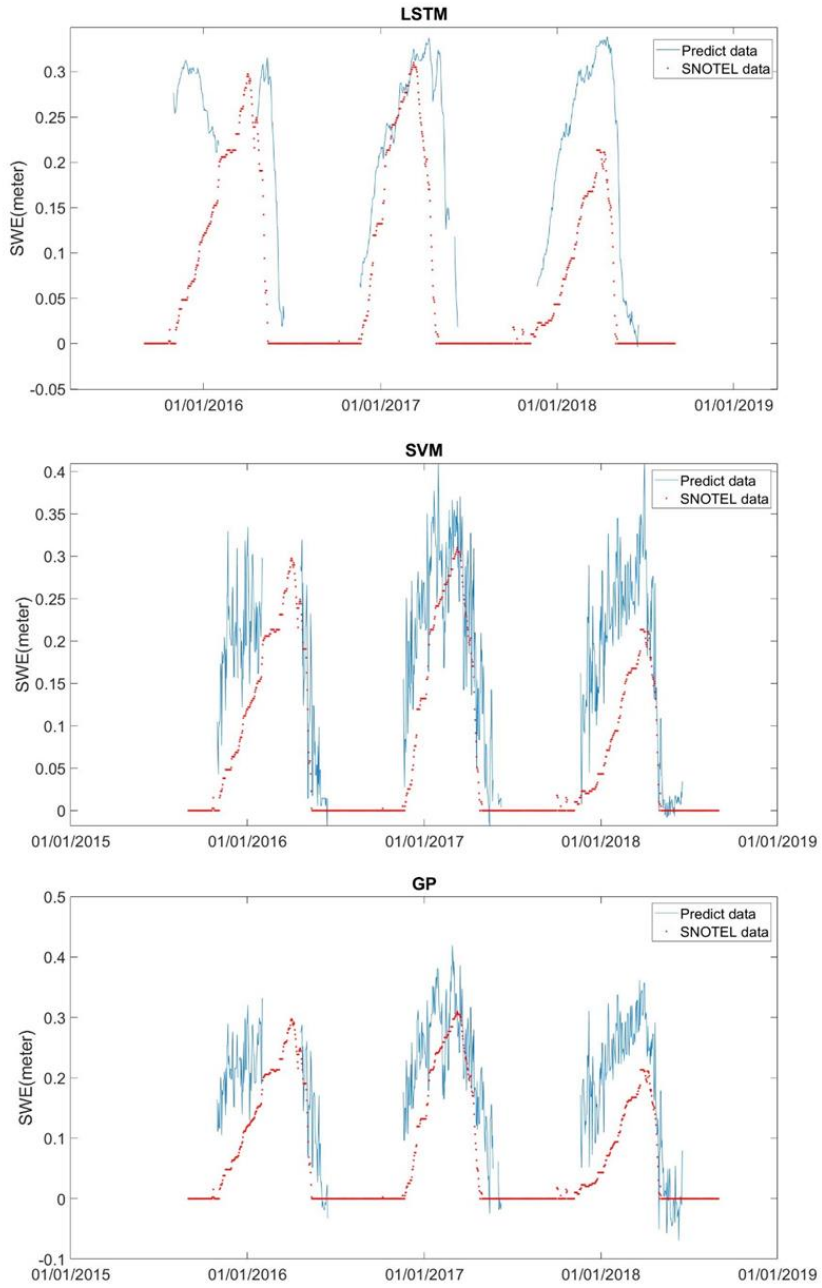


Figure 4.1-12 Time Series Plots of Station 8

4.1.4. SWE in the Regions with Low Elevation and High Forest Cover

In the overall study of the selected stations in the study area, the low-elevation forested areas have the worst prediction performance. This study choose station #88 as the typical station which has low elevation and low forest cover. The basic

characteristics of station #88 are shown below:

Table 4-4 Basic Information of Station 88

Station number	latitude	longitude	elevation	Land surface type
88	39.7164 N	106.15 W	2865.1m	forest

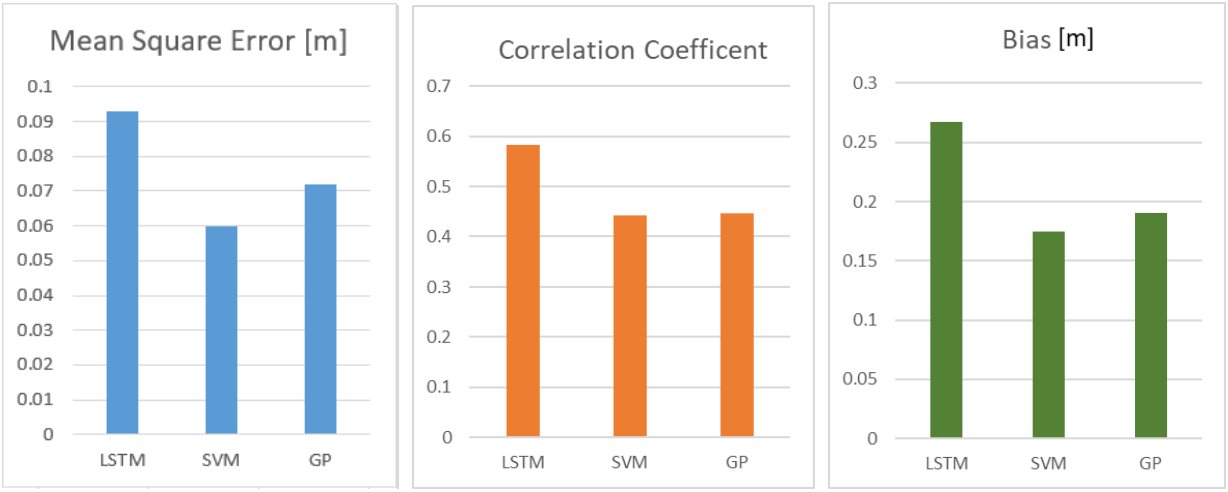


Figure 4.1-13 Statistical Parameters of Station 88 for All Three Algorithms

In the low-altitude forested area, the LSTM algorithm shows much higher MSE than the other two algorithms, which is similar to the scenario in the low-altitude area with low vegetation cover. However, the LSTM still has the best performance in terms of correlation coefficient. According to the comprehensive evaluation of MSE, correlation coefficient and bias, SVM has the lowest MSE and bias, and the correlation coefficient is similar to GP algorithm, so it can be considered the algorithm with the best

performance in the low-altitude forest area.

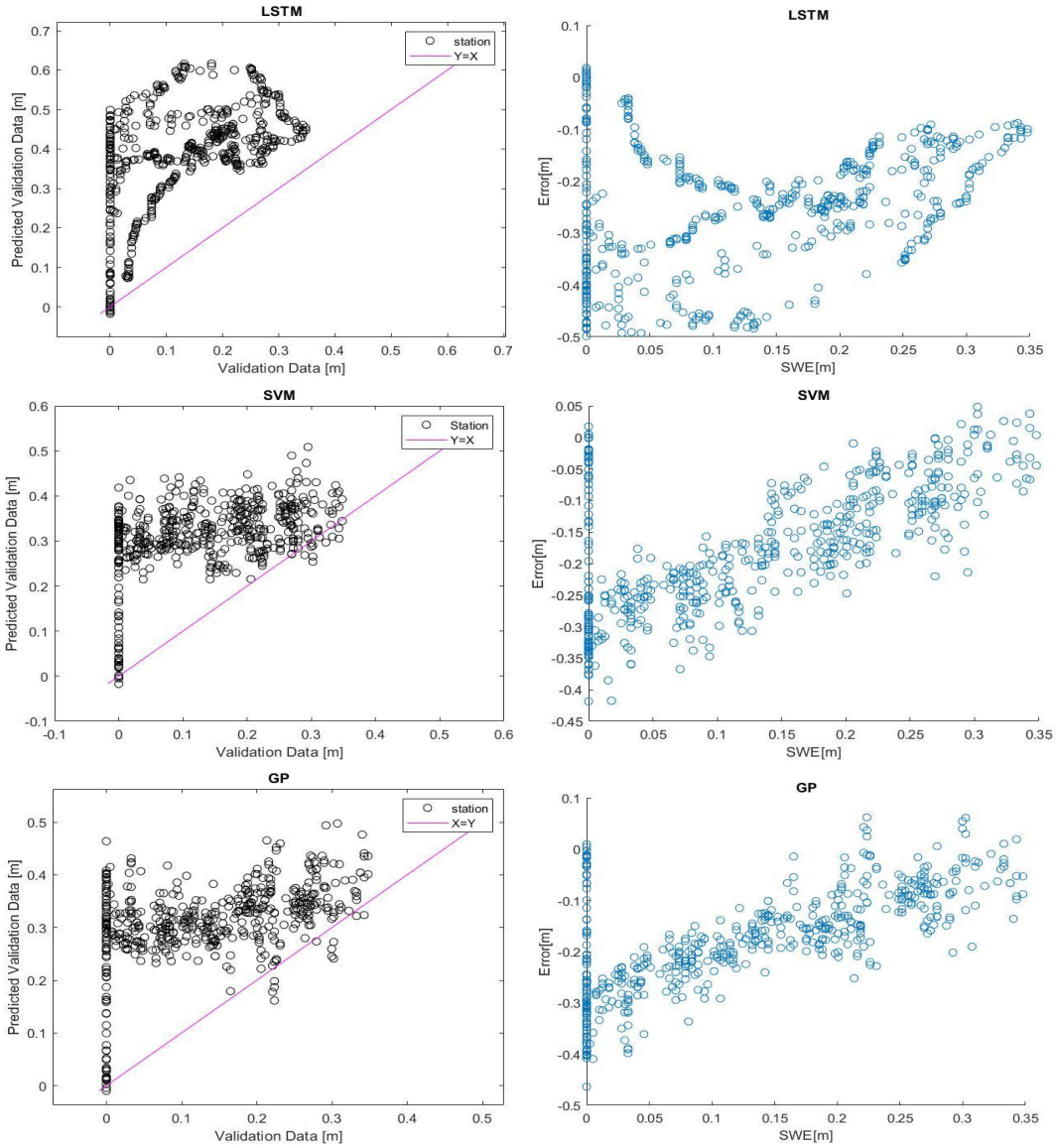


Figure 4.1-14 Prediction and Validation Plots and MSE Dot Plots for All Three Algorithms in Station 88

It can be seen from Figure 4.1-14 that all three algorithms have significant differences between the SNOTEL observations and high MSE in the shallow snow period when SWE is less than 0.2 meters. In addition, compared with the results of low-elevation areas with low vegetation cover, the MSE and bias in low-elevation forest

areas increased by about 30%. At the same time, the predicted values of all three

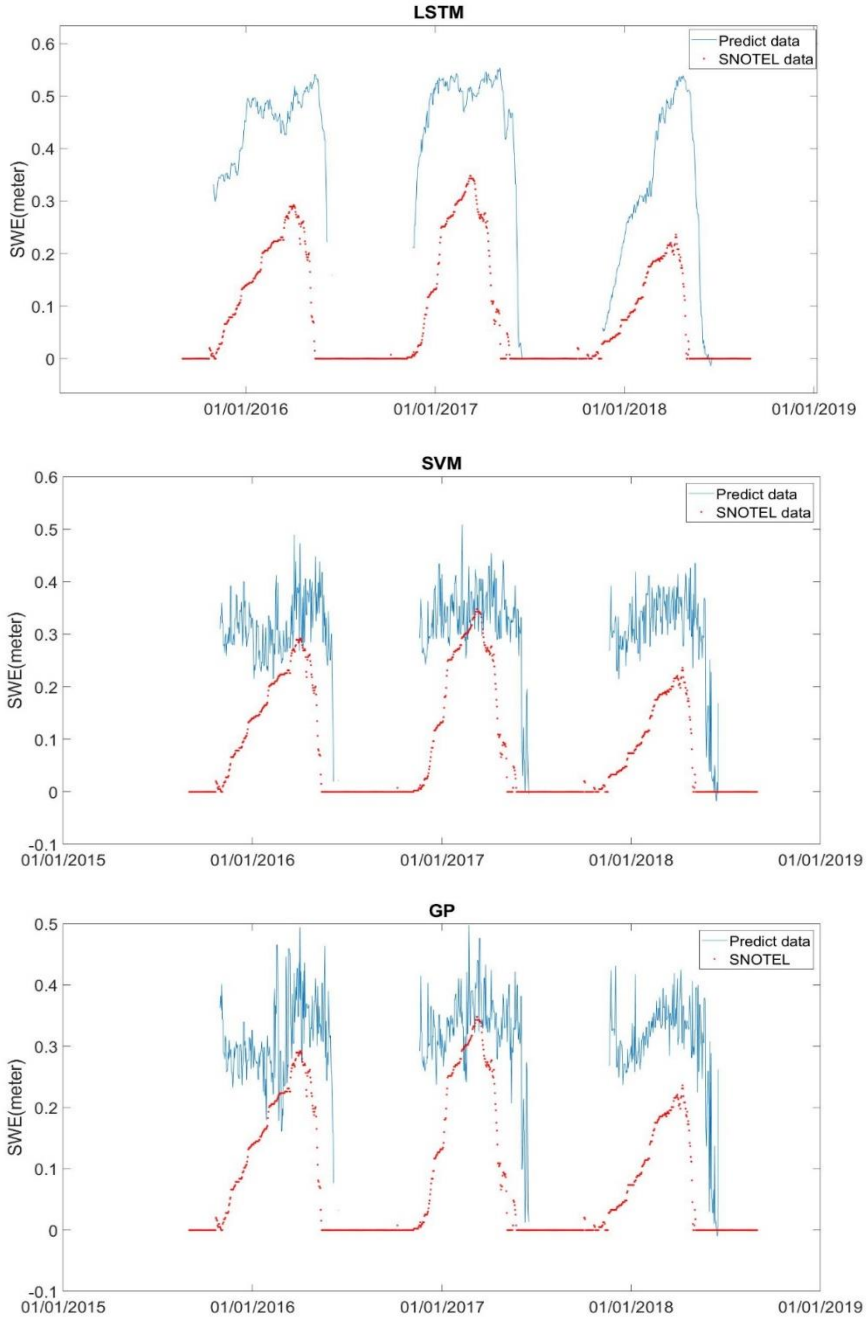


Figure 4.1-15 Time Series Plots of Station 88

algorithms are generally higher than the actual SNOTEL SWE.

From the time series diagram (Figure 4.1-15), it is more obvious that the predicted value of the LSTM algorithm for SWE can be twice that of the actual SNOTEL SWE.

However, SVM regression and GP regression algorithms can still give a better prediction value when SWE is greater than 0.2 meters, especially in the first and second year of the snow peak period. Although the LSTM algorithm had the least amount of variability (noise), its prediction accuracy is significantly lower than the other two algorithms, and it does not even yield the same general trend as the actual SWE during the first year. In terms of capturing the annual variation trend of snow cover, the GP regression algorithm performs the best, especially during the third year of snow accumulation season, which shows a prediction line that is more consistent with the actual SWE variation trend than the other two algorithms. Unfortunately, compared with the prediction performance of low vegetation covered areas, none of the three algorithms could predict the SWE of snow melt season well in forested areas.

In general, in low altitude forest areas, SVM regression has higher prediction accuracy, and GP regression algorithm has more realistic SWE change prediction. However, the LSTM algorithm, which has a relatively good performance in both high-altitude areas and low altitude areas with low vegetation coverage, shows a large prediction deviation relative to the SNOTEL measurements. At the same time, the three algorithms also show the limitations mentioned in the previous part that they cannot make relatively accurate prediction during the snowmelt season or during periods of shallow snow.

4.2. *Temporal Analysis*

In this paper, station #64 located in a grassland area with high altitude was selected as a typical station for time series analysis to explore the performance differences of LSTM during the accumulation season and snowmelt (ablation) season. At the same time, it also helps us to understand some limitations of PMW snow remote sensing.

Table 4-5 Basic Information of Station 64

Station number	latitude	longitude	elevation	Land surface type
64	39.1330 N	107.283W	3048m	Grass

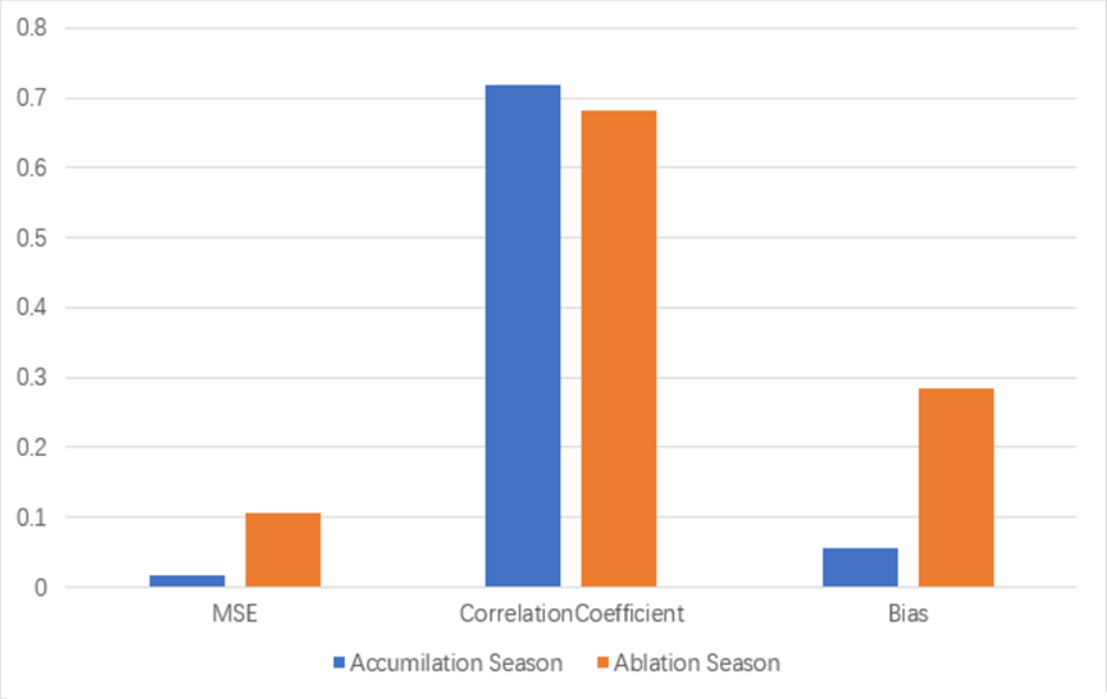


Figure 4.2-1 Statistical Parameters for Accumulation Season and Ablation Season using LSTM

As can be seen from Figure 4.2-1, the MSE and bias of the predicted value of the LSTM algorithm during the snow accumulation season are much smaller than those in the snow melting season. At the same time, the correlation coefficient during the

accumulation season is slightly higher than during the ablation season.

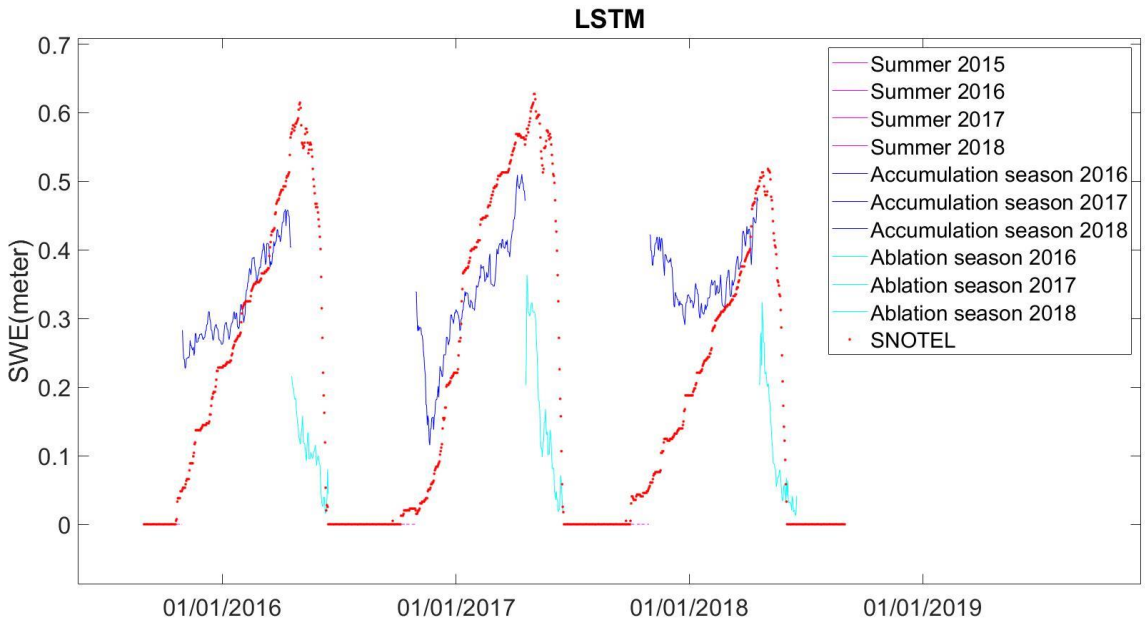


Figure 4.2-2 Time Series Plots of Station 63 using LSTM

From the time series diagram (Figure 4.2-2), it is obvious that in the snow accumulation season, the predicted value of the LSTM algorithm has a high degree of consistency with the actual SNOTEL SWE measurements and can yield good prediction accuracy during the period when SWE is greater than 0.3 meters. During the snowmelt season, although the predicted value can still show the variation trend of snow cover, the accuracy of the prediction is greatly reduced. This is consistent with the previously mentioned situation that the two have relatively large differences in MSE and bias but similar correlation coefficients.

The prediction performance of the LSTM in the snow accumulation season and snowmelt season has a big difference, which is determined by the different physical characteristics of the snow during the two different time periods. During the snow

accumulation season, snow melt is less common and the snow is mainly dry. During the snowmelt season, melting of the snow is more commonplace, and at night after the temperature decreased, the liquid water in the snow will refreeze, which leads to a more complicated and layered snow structure [Tsutsui, 2016]. This phenomenon, together with the liquid water produced by melting snow layers during the snowmelt season, has a great impact on the PMW brightness temperature data collected by PMW snow remote sensing. Due to the poor quality of the input brightness temperature data with relation to snow mass, the final prediction results of the machine learning algorithms will similarly suffer.

4.3. General Evaluation

According to the overall evaluation, the LSTM algorithm has the best performance in high-altitude areas with low vegetation coverage, while SVM and GP regression algorithms have better performance in low-altitude forest areas. In addition, the predicted value of the LSTM algorithm can better capture inter-annual variability and higher correlation with the observed SNOTEL data. This reason is that the LSTM includes a time series component that can put the information and characteristic of former time steps into the present time step compared to memoryless algorithms like SVM and GP regression [Zhang, Lindholm and Ratnaweera, 2018].

However, they still have a number of limitations. When the snow cover is shallow (SWE less than 0.2 m), the difference between the Tb data recorded by the PMW sensor

at different frequencies can be negligible [Armstrong, Chang, Rango and Josberger, 1993]. When the snow is deep (SWE greater than 0.5 m), the signal received by PMW sensor actually originates from the middle of the snowpack rather than from the soil under the snowpack, and the radiation signal emitted by the surface is scattered in the snow layer before exiting the snowpack [Mulders, 1987]. In addition, the liquid water in the wet snow emits more radiation itself than dry snow and thus increases the measured brightness temperature [Walker and Goodison, 1993]. The reason is that liquid water has a higher r dielectric constant compared to that of dry snow, which causes the dielectric of snowpack to increase. Treating water in wet snow regions as dry snow can lead to significant errors in snow retrievals [Tedesco and Narvekar, 2010]. As a result, the performance of all three algorithms during the ablation season is relatively poor.

Chapter5.Conclusion and Discussion

5.1.Conclusion

Based on the results and analysis in Chapter IV, this paper evaluates the performance of three different machine learning algorithms, and qualitatively analyzes the influence of various influencing factors on PMW snow remote sensing and algorithm performance. The main conclusions are summarized as follows:

(1) The long short term memory (LSTM) algorithm has relatively good performance in high altitude areas with low vegetation coverage. In these regions, the LSTM algorithm has less variability (noise), yields more accurate predictions, and best captures the trend of SWE variability across the entire snow season. At the same time, LSTM algorithm can also show part of the annual variation of snow cover.

(2) Support vector machine (SVM) regression has better performance in low altitude areas. Compared with the LSTM, SVM regression is more variable, but still has its advantages in low altitude areas and forest areas. In these areas, compared with LSTM, SVM regression yields a more accurate prediction with a higher degree of conformity to the general trend of snow mass throughout the entire snow season.

(3) Gaussian process (GP) regression algorithm has a slightly lower performance than SVM regression in most areas, but in the low-altitude forest area, GP regression algorithm has the highest degree of conformity to the general trend of snow mass throughout the entire snow season.

(4) The indirect influence of overlying vegetation on the algorithm is significant. The performance of the LSTM algorithm in forested areas is much worse than that in low vegetation areas. Compared with SVM regression and GP regression algorithms, the prediction accuracy of the LSTM algorithm is more sensitive to vegetation type. In forested areas, the prediction value of LSTM often yields the highest MSE and bias. In general, the performance of the three algorithms in forested areas is worse than that in the low vegetation cover area, and the influence of SVM regression and GP regression algorithm is less than that of the LSTM algorithm.

(5) The depth of the snow is another factor that has a significant effect on the performance of the algorithm. This is because when the snow is too shallow, the brightness temperature difference between different frequencies collected by the PMW sensor will be almost zero, while when the snow is too deep, the PMW sensor cannot receive the radiation signal emitted from the surface soil that is covered by the overlying snow. All three algorithms have the best performance under medium snow depth (SWE between 0.3-0.5 m). However, LSTM can still yield reliatvely accurate estimes in deep snow with SWE between 0.5 and 0.65 meters, but at the same time, the LSTM has the poorest performance when SWE is less than 0.3 meters. On the contrary, SVM regression and GP regression algorithms can still maintain good prediction accuracy when SWE is between 0.2-0.3 meters, but fails when SWE exceeds 0.5 meters.

(6) When the snow accumulation season and snow melt season are modeled separately, the snow accumulation season has a significantly greater accuracy. However, the snow melt season is affected by the liquid water and ice in the snow layer, and the accuracy of the prediction is significantly lower.

Appendix A

Table A-1 Statistical Value of Three algorithms for Stations with Low Elevation and Low Forest Cover

			LSTM		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
8	Shrubland	2590.8000	0.0279	0.1396	0.6994
69	Shrubland	2651.7600	0.0114	0.0550	0.6445
103	Shrubland	2816.3520	0.0197	0.0966	0.7137
7	Shrubland	2889.5040	0.0207	0.1034	0.6719
29	Grass	2895.6000	0.0240	0.0922	0.7163
98	Grass	2907.7920	0.0099	0.0540	0.8168
9	Shrubland	2913.8880	0.0426	0.1109	0.3416
52	Grass	2987.0400	0.0066	0.0099	0.8053
			SVM Regression		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
8	Shrubland	2590.8000	0.0110	0.0817	0.7876
69	Shrubland	2651.7600	0.0151	0.0923	0.6221
103	Shrubland	2816.3520	0.0169	0.0723	0.6577
7	Shrubland	2889.5040	0.0083	0.0119	0.7543
29	Grass	2895.6000	0.0372	0.0795	0.3003
98	Grass	2907.7920	0.0194	-0.0040	0.5527
9	Shrubland	2913.8880	0.0311	0.0268	0.2192
52	Grass	2987.0400	0.0099	0.0353	0.7446
			GP Regression		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
8	Shrubland	2590.8000	0.0127	0.0889	0.7623
69	Shrubland	2651.7600	0.0186	0.1015	0.4778
103	Shrubland	2816.3520	0.0208	0.1029	0.6911
7	Shrubland	2889.5040	0.0075	0.0264	0.7876
29	Grass	2895.6000	0.0406	0.0782	0.1619
98	Grass	2907.7920	0.0165	0.0088	0.5411
9	Shrubland	2913.8880	0.0349	0.0187	0.0642
52	Grass	2987.0400	0.0131	0.0321	0.7059

Table A-2 Statistical Value of Three algorithms for Stations with Low Elevation and High Forest Cover

Station number	Surface type	Elevation	LSTM		
			MSE	Bias	Correlation coefficient
40	Forest	2621.2800	0.0898	0.2719	0.2947
86	Forest	2657.8560	0.0718	0.2464	0.5063
2	Forest	2724.9120	0.0870	0.2763	0.8072
75	Forest	2752.3440	0.0962	0.2933	0.3973
88	Forest	2865.1200	0.0928	0.2784	0.4101
57	Forest	2956.5600	0.0741	0.2412	0.3130
Station number	Surface type	Elevation	SVM Regression		
			MSE	Bias	Correlation coefficient
40	Forest	2621.2800	0.0670	0.2444	0.1963
86	Forest	2657.8560	0.0678	0.2328	0.3110
2	Forest	2724.9120	0.0608	0.2073	0.5486
75	Forest	2752.3440	0.0447	0.1845	0.4177
88	Forest	2865.1200	0.0599	0.2229	0.5200
57	Forest	2956.5600	0.0390	0.1586	0.2731
Station number	Surface type	Elevation	GP Regression		
			MSE	Bias	Correlation coefficient
40	Forest	2621.2800	0.0590	0.2192	0.2830
86	Forest	2657.8560	0.0600	0.2212	0.3071
2	Forest	2724.9120	0.0930	0.2786	0.4976
75	Forest	2752.3440	0.0465	0.1916	0.4390
88	Forest	2865.1200	0.0720	0.2487	0.4840
57	Forest	2956.5600	0.0348	0.1329	0.0696

Table A-3 Statistical Value of Three algorithms for Stations with High Elevation and Low Forest Cover

			LSTM		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
67	Grass	3060.1920	0.0086	-0.0067	0.7494
16	Grass	3081.5280	0.0114	-0.0301	0.6534
4	Grass	3133.3440	0.0134	0.0140	0.7556
92	Grass	3139.4400	0.0133	0.0188	0.7014
111	Grass	3169.9200	0.0079	0.0441	0.8287
41	Grass	3215.6400	0.0055	0.0409	0.8810
31	Shrubland	3444.2400	0.0045	0.0032	0.9291
51	Grass	3474.7200	0.0069	-0.0177	0.7980
62	Shrubland	3474.7200	0.0184	-0.0805	0.7734
30	Grass	3535.6800	0.0170	-0.0308	0.7952
			SVM Regression		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
67	Grass	3060.1920	0.0106	0.0290	0.7267
16	Grass	3081.5280	0.0081	-0.0134	0.7062
4	Grass	3133.3440	0.0277	-0.0278	0.3944
92	Grass	3139.4400	0.0286	0.0178	0.1220
111	Grass	3169.9200	0.0119	0.0136	0.6115
41	Grass	3215.6400	0.0145	0.0632	0.6298
31	Shrubland	3444.2400	0.0166	-0.0367	0.6356
51	Grass	3474.7200	0.0127	0.0324	0.5929
62	Shrubland	3474.7200	0.0206	-0.0305	0.5706
30	Grass	3535.6800	0.0366	0.0421	0.4260
			GP Regression		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
67	Grass	3060.1920	0.0130	0.0277	0.7004
16	Grass	3081.5280	0.0087	-0.0044	0.6674
4	Grass	3133.3440	0.0258	-0.0265	0.4352
92	Grass	3139.4400	0.0408	-0.0363	-0.1522

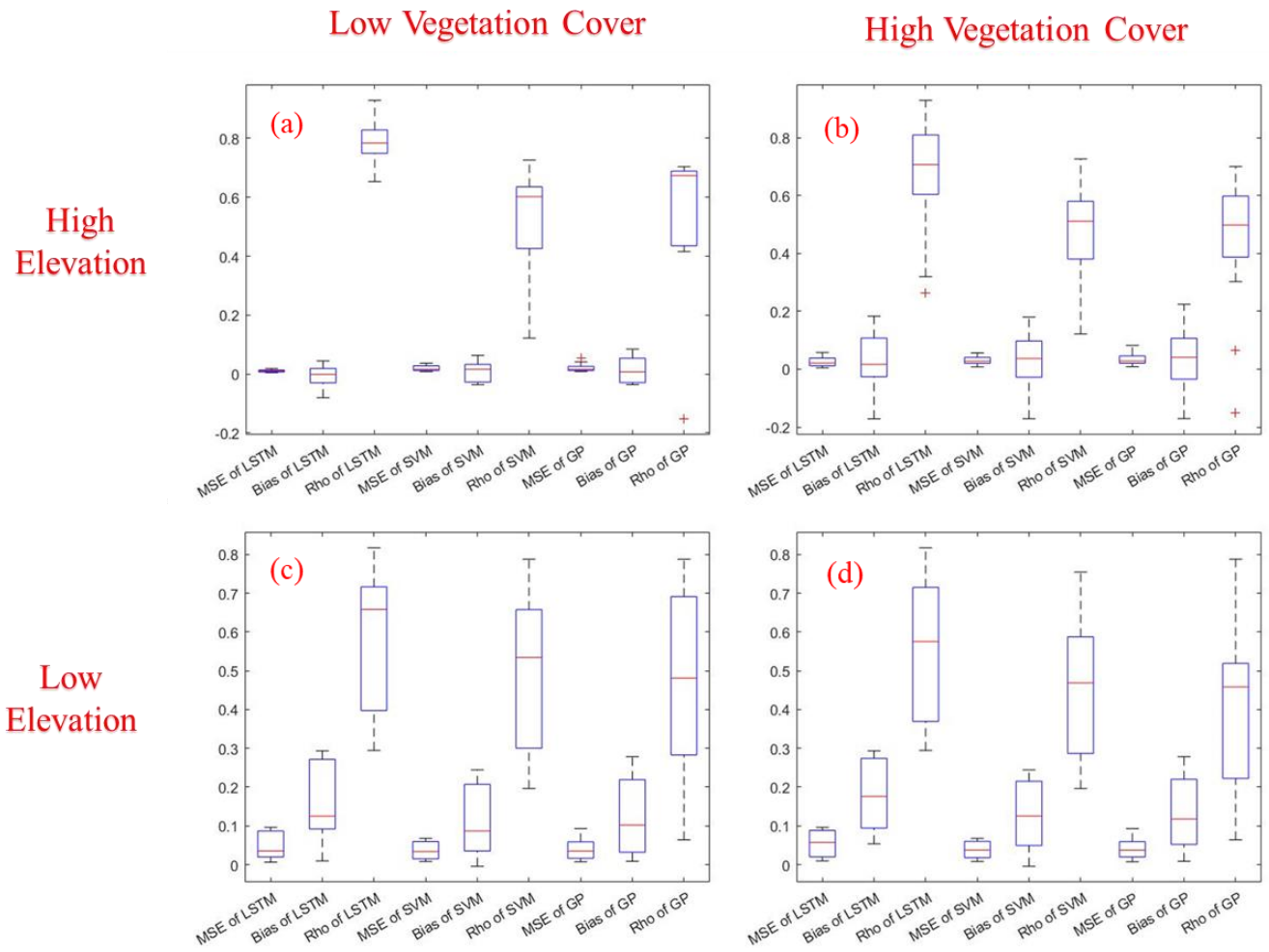
111	Grass	3169.9200	0.0099	0.0191	0.6836
41	Grass	3215.6400	0.0165	0.0847	0.6804
31	Shrubland	3444.2400	0.0144	-0.0347	0.6889
51	Grass	3474.7200	0.0119	0.0532	0.7040
62	Shrubland	3474.7200	0.0206	-0.0293	0.5810
30	Grass	3535.6800	0.0539	0.0555	0.4156

Table A-4 Statistical Value of Three algorithms for Stations with High Elevation and High Forest Cover

			LSTM		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
70	Forest	3020.5680	0.0361	0.1679	0.6692
95	Forest	3139.4400	0.0421	0.1829	0.8367
109	Forest	3157.7280	0.0254	0.1257	0.7127
54	Forest	3169.9200	0.0433	-0.1451	0.4107
68	Forest	3200.4000	0.0133	-0.0262	0.8096
18	Forest	3206.4960	0.0449	0.1554	0.6196
107	Forest	3230.8800	0.0163	0.1176	0.9118
96	Forest	3316.2240	0.0299	0.0791	0.2630
3	Forest	3340.6080	0.0384	-0.1458	0.7726
50	Forest	3383.2800	0.0485	0.1075	0.3198
27	Forest	3398.5200	0.0299	0.1056	0.6524
108	Forest	3398.5200	0.0119	-0.0122	0.3559
78	Forest	3413.7600	0.0575	-0.1716	0.5211
87	Forest	3413.7600	0.0171	0.0136	0.7124
49	Forest	3474.7200	0.0261	-0.1112	0.6041
			SVM Regression		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
70	Forest	3020.5680	0.0275	0.1136	0.2607
95	Forest	3139.4400	0.0417	0.1307	0.4410
109	Forest	3157.7280	0.0407	0.1802	0.5605
54	Forest	3169.9200	0.0228	-0.0506	0.5046
68	Forest	3200.4000	0.0406	0.0796	0.3838
18	Forest	3206.4960	0.0264	0.0444	0.5180
107	Forest	3230.8800	0.0330	0.1603	0.5633
96	Forest	3316.2240	0.0562	0.1699	0.3748
3	Forest	3340.6080	0.0548	-0.1711	0.5803
50	Forest	3383.2800	0.0209	-0.0154	0.4881
27	Forest	3398.5200	0.0252	0.0473	0.5409
108	Forest	3398.5200	0.0204	0.0877	0.3805
78	Forest	3413.7600	0.0387	-0.1149	0.5643

87	Forest	3413.7600	0.0505	0.0977	0.3058
49	Forest	3474.7200	0.0285	-0.0352	0.1757
			GP Regression		
Station number	Surface type	Elevation	MSE	Bias	Correlation coefficient
70	Forest	3020.5680	0.0243	0.0531	0.0636
95	Forest	3139.4400	0.0444	0.1406	0.4259
109	Forest	3157.7280	0.0452	0.1935	0.5451
54	Forest	3169.9200	0.0212	-0.0448	0.5328
68	Forest	3200.4000	0.0509	0.0872	0.3873
18	Forest	3206.4960	0.0299	0.0750	0.5142
107	Forest	3230.8800	0.0457	0.1935	0.5986
96	Forest	3316.2240	0.0820	0.2242	0.3125
3	Forest	3340.6080	0.0563	-0.1709	0.5772
50	Forest	3383.2800	0.0245	-0.0386	0.3872
27	Forest	3398.5200	0.0300	0.0881	0.5375
108	Forest	3398.5200	0.0207	0.1066	0.4821
78	Forest	3413.7600	0.0463	-0.1158	0.4528
87	Forest	3413.7600	0.0589	0.1473	0.3589
49	Forest	3474.7200	0.0222	-0.0151	0.3030

Figure A-1 Boxplots of Statistical Value for All Test Stations



Bibliography

Ahmad, J.A., Forman, B.A. and Kwon, Y. (2019) "Analyzing Machine Learning Predictions of Passive Microwave Brightness Temperature Spectral Difference Over Snow-Covered Terrain in High Mountain Asia," Frontiers in Earth Science, 7. Available at: https://doi.org/10.3389/feart.2019.00212.

Armstrong, R.L. and Brun, E. (2010) Snow and climate. Cambridge ... [etc.]: Cambridge University Press.

Armstrong, R.L. et al. (1993) "Snow depths and grain-size relationships with relevance for passive microwave studies," Annals of Glaciology, 17, pp. 171–176. Available at: https://doi.org/10.1017/s0260305500012799.

Atkinson, P.M. and Tatnall, A.R.L. (1997) "Introduction Neural networks in remote sensing," International Journal of Remote Sensing, 18(4), pp. 699–709. Available at: https://doi.org/10.1080/014311697218700.

Avanzi, F. et al. (2014) "A processing–modeling routine to use SNOTEL hourly data in snowpack dynamic models," Advances in Water Resources, 73, pp. 16–29. Available at: https://doi.org/10.1016/j.advwatres.2014.06.011.

Barnett, T.P. et al. (1989) "The Effect of Eurasian Snow Cover on Regional and Global Climate Variations," Journal of the Atmospheric Sciences, 46(5), pp. 661–686. Available at: https://doi.org/10.1175/1520-0469(1989)046<0661:teoesc>2.0.co;2.

Bechle, M.J., Millet, D.B. and Marshall, J.D. (2013) "Remote sensing of exposure to NO2: satellite versus ground based measurement in a large urban area," ISEE Conference Abstracts, 2013(1), p. 4380. Available at: https://doi.org/10.1289/isee.2013.o-1-35-01.

Ben-Hur, A. and Weston, J. (2010) Data mining techniques for the life sciences. Humana Press, pp. 223–239.

Campbell, J.B. (2002) "Introduction to remote sensing," CRC Press [Preprint].

Castle, G.H. (1969) "Operational Experience with Snow Pillows in California," Journal of the Irrigation and Drainage Division, 95(1), pp. 105–116. Available at: https://doi.org/10.1061/jrcea4.0000620.

Chang, A.T.C. et al. (1982) “Snow water equivalent estimation by microwave radiometry,” Cold Regions Science and Technology, 5(3), pp. 259–267. Available at: https://doi.org/10.1016/0165-232x(82)90019-2.

Chang, A.T.C., Foster, J.L. and Hall, D.K. (1987) “Nimbus-7 SMMR Derived Global Snow Cover Parameters,” Annals of Glaciology, 9, pp. 39–44. Available at: https://doi.org/10.1017/s0260305500000355.

CHANG, A.T.C., FOSTER, J.L. and HALL, D.K. (1996) “EFFECTS OF FOREST ON THE SNOW PARAMETERS DERIVED FROM MICROWAVE MEASUREMENTS DURING THE BOREAS WINTER FIELD CAMPAIGN,” Hydrological Processes, 10(12), pp. 1565–1574. Available at: https://doi.org/10.1002/(sici)1099-1085(199612)10:12<1565::aid-hyp501>3.0.co;2-5.

Chang, T.C. et al. (1976) “Microwave Emission From Snow and Glacier Ice,” Journal of Glaciology, 16(74), pp. 23–39. Available at: https://doi.org/10.1017/s0022143000031415.

Cherkassky, V. (1997) “The Nature Of Statistical Learning Theory~,” IEEE Transactions on Neural Networks, 8(6), pp. 1564–1564. Available at: https://doi.org/10.1109/tnn.1997.641482.

Cortes, C. and Vapnik, V. (1995) “Support-vector networks,” Machine Learning, 20(3), pp. 273–297. Available at: https://doi.org/10.1007/bf00994018.

Davis, D.T. et al. (1993) “Retrieval of snow parameters by iterative inversion of a neural network,” IEEE Transactions on Geoscience and Remote Sensing, 31(4), pp. 842–852. Available at: https://doi.org/10.1109/36.239907.

DERKSEN, C., WALKER, A. and GOODISON, B. (2005) “Evaluation of passive microwave snow water equivalent retrievals across the boreal forest/tundra transition of western Canada,” Remote Sensing of Environment, 96(3-4), pp. 315–327. Available at: https://doi.org/10.1016/j.rse.2005.02.014.

Domingos, P. (2012) “A few useful things to know about machine learning,” Communications of the ACM, 55(10), pp. 78–87. Available at: https://doi.org/10.1145/2347736.2347755.

Dozier, J. and Shi, J. (2000) “Estimation of snow water equivalence using SIR-C/X-SAR. I. Inferring snow density and subsurface properties,”

IEEE Transactions on Geoscience and Remote Sensing, 38(6), pp. 2465–2474. Available at: https://doi.org/10.1109/36.885195.

Durand, M. and Margulis, S.A. (2007) “Correcting first‐order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme,” Journal of Geophysical Research: Atmospheres, 112(D13). Available at: https://doi.org/10.1029/2006jd008067.

Durand, M., Kim, E.J. and Margulis, S.A. (2008) “Quantifying Uncertainty in Modeling Snow Microwave Radiance for a Mountain Snowpack at the Point-Scale, Including Stratigraphic Effects,” IEEE Transactions on Geoscience and Remote Sensing, 46(6), pp. 1753–1767. Available at: https://doi.org/10.1109/tgrs.2008.916221.

Dyer, J.L. and Mote, T.L. (2006) “Spatial variability and trends in observed snow depth over North America,” Geophysical Research Letters, 33(16). Available at: https://doi.org/10.1029/2006gl027258.

Eckerstorfer, M. et al. (2016) “Remote sensing of snow avalanches: Recent advances, potential, and limitations,” Cold Regions Science and Technology, 121, pp. 126–140. Available at: https://doi.org/10.1016/j.coldregions.2015.11.001.

Fearn, T. (2013) “Gaussian Process Regression,” NIR news, 24(6), pp. 23–24. Available at: https://doi.org/10.1255/nirn.1392.

Foppa, N., Stoffel, A. and Meister, R. (2007) “Synergy of in situ and space borne observation for snow depth mapping in the Swiss Alps,” International Journal of Applied Earth Observation and Geoinformation, 9(3), pp. 294–310. Available at: https://doi.org/10.1016/j.jag.2006.10.001.

Foster, J.L. et al. (2011) “A blended global snow product using visible, passive microwave and scatterometer satellite data,” International Journal of Remote Sensing, 32(5), pp. 1371–1395. Available at: https://doi.org/10.1080/01431160903548013.

Foster, J.L., Hall, D.K. and Chang, A.T.C. (1987) “Remote sensing of snow,” Eos, Transactions American Geophysical Union, 68(32), p. 682. Available at: https://doi.org/10.1029/eo068i032p00682-01.

Gers, F.A., Schmidhuber, J. and Cummins, F. (2000) “Learning to Forget: Continual Prediction with LSTM,” Neural Computation, 12(10), pp. 2451–2471. Available at: https://doi.org/10.1162/089976600300015015.

Goodison, B.E. and Walker, A.E. (1993) “Use of snow cover derived from satellite passive microwave data as an indicator of climate change,” Annals of Glaciology, 17, pp. 137–142. Available at: https://doi.org/10.1017/s0260305500012738.

Haasdonk, B. and Burkhardt, H. (2007) “Invariant kernel functions for pattern analysis and machine learning,” Machine Learning, 68(1), pp. 35–61. Available at: https://doi.org/10.1007/s10994-007-5009-7.

Hall, D.K. and Martinec, J. (1986) “Remote sensing of ice and snow,” Geocarto International, 1(1), pp. 64–64. Available at: https://doi.org/10.1080/10106048609354032.

Hallikainen, M., Ulaby, F. and Abdelrazik, M. (1986) “Dielectric properties of snow in the 3 to 37 GHz range,” IEEE Transactions on Antennas and Propagation, 34(11), pp. 1329–1340. Available at: https://doi.org/10.1109/tap.1986.1143757.

IV, J.F.E. et al. (1996) “Machine Learning, Neural, and Statistical Classification.,” Journal of the American Statistical Association, 91(433), p. 436. Available at: https://doi.org/10.2307/2291432.

Karney, C.F.F. (2012) “Algorithms for geodesics,” Journal of Geodesy, 87(1), pp. 43–55. Available at: https://doi.org/10.1007/s00190-012-0578-z.

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

Kelly, R.E. et al. (2003) “A prototype AMSR-E global snow area and snow depth algorithm,” IEEE Transactions on Geoscience and Remote Sensing, 41(2), pp. 230–242. Available at: https://doi.org/10.1109/tgrs.2003.809118.

Kiedron, P.W. et al. (1999) “Comparison of spectral irradiance standards used to calibrate shortwave radiometers and spectroradiometers,” Applied Optics, 38(12), p. 2432. Available at: https://doi.org/10.1364/ao.38.002432.

Kolassa, J. et al. (2017) “Soil moisture retrieval from AMSR-E and ASCAT microwave observation synergy. Part 2: Product evaluation,” Remote Sensing of Environment, 195, pp. 202–217. Available at: https://doi.org/10.1016/j.rse.2017.04.020.

Kratzert, F. et al. (2019) “Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample

datasets," Hydrology and Earth System Sciences, 23(12), pp. 5089–5110. Available at: https://doi.org/10.5194/hess-23-5089-2019.

Langlois, A. et al. (2011) "Improved Corrections of Forest Effects on Passive Microwave Satellite Remote Sensing of Snow Over Boreal and Subarctic Regions," IEEE Transactions on Geoscience and Remote Sensing, 49(10), pp. 3824–3837. Available at: https://doi.org/10.1109/tgrs.2011.2138145.

Larson, K.M. et al. (2009) "Can we measure snow depth with GPS receivers?," Geophysical Research Letters, 36(17). Available at: https://doi.org/10.1029/2009gl039430.

Lathi, B.P. and Dao, P.V. (1989) Modern digital and analog communication systems. Philadelphia: Saunders College Publ.

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

Luce, C.H., Tarboton, D.G. and Cooley, K.R. (1998) "The influence of the spatial distribution of snow on basin-averaged snowmelt," Hydrological Processes, 12(10-11), pp. 1671–1683. Available at: https://doi.org/10.1002/(sici)1099-1085(199808/09)12:10/11<1671::aid-hyp688>3.0.co;2-n.

McLaughlin, D. (2002) "An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering," Advances in Water Resources, 25(8-12), pp. 1275–1286. Available at: https://doi.org/10.1016/s0309-1708(02)00055-6.

Molotch, N.P. and Bales, R.C. (2006) "SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence," Hydrological Processes, 20(4), pp. 723–739. Available at: https://doi.org/10.1002/hyp.6128.

Mott, H. (1992) Antennas for radar and communications. New York: Wiley.

Mulders, M.A. (1987) Remote Sensing in Soil Science. Burlington: Elsevier.

Pulliainen, J.T., Grandell, J. and Hallikainen, M.T. (1999) "HUT snow emission model and its applicability to snow water equivalent retrieval," IEEE Transactions on Geoscience and Remote Sensing, 37(3), pp. 1378–1390. Available at: https://doi.org/10.1109/36.763302.

Raissi, M., Babaee, H. and Karniadakis, G.E. (2019) “Parametric Gaussian process regression for big data,” Computational Mechanics, 64(2), pp. 409–416. Available at: https://doi.org/10.1007/s00466-019-01711-5.

RANGO, A. (1996) “Spaceborne remote sensing for snow hydrology applications,” Hydrological Sciences Journal, 41(4), pp. 477–494. Available at: https://doi.org/10.1080/02626669609491521.

Reichle, R.H. (2008) “Data assimilation methods in the Earth sciences,” Advances in Water Resources, 31(11), pp. 1411–1418. Available at: https://doi.org/10.1016/j.advwatres.2008.01.001.

Rice, S.A. (1975) “Solid Water: Ice Physics . Peter V. Hobbs. Clarendon (Oxford University Press), New York, 1975. xviii, 838 pp., illus. $85.,” Science, 189(4208), pp. 1082–1082. Available at: https://doi.org/10.1126/science.189.4208.1082.a.

Robinson, D.A., Dewey, K.F. and Heim, R.R. (1993) “Global Snow Cover Monitoring: An Update,” Bulletin of the American Meteorological Society, 74(9), pp. 1689–1696. Available at: https://doi.org/10.1175/1520-0477(1993)074<1689:gscmau>2.0.co;2.

Ruppert, D. (2004) “The Elements of Statistical Learning: Data Mining, Inference, and Prediction,” Journal of the American Statistical Association, 99(466), pp. 567–567. Available at: https://doi.org/10.1198/jasa.2004.s339.

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

Serreze, M.C. et al. (1999) “Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data,” Water Resources Research, 35(7), pp. 2145–2160. Available at: https://doi.org/10.1029/1999wr900090.

Serreze, M.C. et al. (1999) “Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data,” Water Resources Research, 35(7), pp. 2145–2160. Available at: https://doi.org/10.1029/1999wr900090.

Shoesmith, E., Vapnik, V. and Kotz, S. (1984) “Estimation of Dependences Based on Empirical Data.,” The Statistician, 33(3), p. 324. Available at: https://doi.org/10.2307/2988246.

Shuttleworth, W.J. (2012) Terrestrial hydrometeorology. Hoboken (N. J.): Wiley-Blackwell.

Smola, A.J. and Schölkopf, B. (2004) "A tutorial on support vector regression," Statistics and Computing, 14(3), pp. 199–222. Available at: https://doi.org/10.1023/b:stco.0000035301.49549.88.

"Special issue on microwave radiometry and remote sensing applications" (2004) IEEE Transactions on Geoscience and Remote Sensing, 42(3), pp. 686–686. Available at: https://doi.org/10.1109/tgrs.2004.826834.

Takala, M. et al. (2011) "Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements," Remote Sensing of Environment, 115(12), pp. 3517–3529. Available at: https://doi.org/10.1016/j.rse.2011.08.014.

Tedesco, M. and Jeyaratnam, J. (2016) "A New Operational Snow Retrieval Algorithm Applied to Historical AMSR-E Brightness Temperatures," Remote Sensing, 8(12), p. 1037. Available at: https://doi.org/10.3390/rs8121037.

Tedesco, M. and Narvekar, P.S. (2010) "Assessment of the NASA AMSR-E SWE Product," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3(1), pp. 141–159. Available at: https://doi.org/10.1109/jstars.2010.2040462.

Tedesco, M. and Wang, J.R. (2006) "Atmospheric Correction of AMSR-E Brightness Temperatures for Dry Snow Cover Mapping," IEEE Geoscience and Remote Sensing Letters, 3(3), pp. 320–324. Available at: https://doi.org/10.1109/lgrs.2006.871744.

Tedesco, M. et al. (2004) "Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data," Remote Sensing of Environment, 90(1), pp. 76–85. Available at: https://doi.org/10.1016/j.rse.2003.12.002.

Thornes, J.E. and Lydolph, P.E. (1987) "Weather and Climate," The Geographical Journal, 153(2), p. 267. Available at: https://doi.org/10.2307/634891.

Tsang, L. et al. (1992) "Inversion of snow parameters from passive microwave remote sensing measurements by a neural network trained with a multiple scattering model," IEEE Transactions on Geoscience

and Remote Sensing, 30(5), pp. 1015–1024. Available at: https://doi.org/10.1109/36.175336.

Tsutsui, H. (2016) “Development of snow retrieval algorithm based on the microwave radiative transfer model for multiple snow layers and various land surface parameters,” 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) [Preprint]. Available at: https://doi.org/10.1109/igarss.2016.7730841.

Ulaby, F.T. and Stiles, W.H. (1980) “The active and passive microwave response to snow parameters: 2. Water equivalent of dry snow,” Journal of Geophysical Research, 85(C2), p. 1045. Available at: https://doi.org/10.1029/jc085ic02p01045.

Vapnik, V.N. (1999) “An overview of statistical learning theory,” IEEE Transactions on Neural Networks, 10(5), pp. 988–999. Available at: https://doi.org/10.1109/72.788640.

Verrelst, J. et al. (2016) “Spectral band selection for vegetation properties retrieval using Gaussian processes regression,” International Journal of Applied Earth Observation and Geoinformation, 52, pp. 554–567. Available at: https://doi.org/10.1016/j.jag.2016.07.016.

Walker, A.E. and Goodison, B.E. (1993) “Discrimination of a wet snow cover using passive microwave satellite data,” Annals of Glaciology, 17, pp. 307–311. Available at: https://doi.org/10.3189/s026030550001301x.

Waters, D.F., Pratt, L.M. and Cadou, C.P. (2021) “Gas turbine/solid oxide fuel cell hybrids for aircraft propulsion and power,” Journal of Propulsion and Power, 37(3), pp. 349–361. Available at: https://doi.org/10.2514/1.b38026.

Wu, R. and Wang, B. (2018) “Gaussian process regression method for forecasting of mortality rates,” Neurocomputing, 316, pp. 232–239. Available at: https://doi.org/10.1016/j.neucom.2018.08.001.

Wu, R. and Wang, B. (2018) “Gaussian process regression method for forecasting of mortality rates,” Neurocomputing, 316, pp. 232–239. Available at: https://doi.org/10.1016/j.neucom.2018.08.001.

Xue, Y. and Forman, B.A. (2015) “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,

170, pp. 153–165. Available at: https://doi.org/10.1016/j.rse.2015.09.009.

Yin, H. et al. (2022) “Rainfall-runoff modeling using long short-term memory based step-sequence framework,” Journal of Hydrology, 610, p. 127901. Available at: https://doi.org/10.1016/j.jhydrol.2022.127901.

Zhang, D., Lindholm, G. and Ratnaweera, H. (2018) “Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring,” Journal of Hydrology, 556, pp. 409–418. Available at: https://doi.org/10.1016/j.jhydrol.2017.11.018.

Zwally, H.J. and Gloersen, P. (1977) “Passive microwave images of the polar regions and research applications,” Polar Record, 18(116), pp. 431–450. Available at: https://doi.org/10.1017/s0032247400000930.