MACHINE LEARNING AND SNOW SCIENCE

Ibrahim Olalekan Alabi
Boise State University
Boise, ID 83725
ibrahimolalekana@u.boisestate.edu

October 21, 2021

ABSTRACT

Almost everything on earth benefits from the services rendered by snow. Snow directly influences the balance between the energy absorbed by the Earth and the energy reflected back into the atmosphere due to its high radiation reflection ability. About 80% to 90% of the total sunlight hitting its surface is reflected back into the atmosphere, thereby cooling off the planet. Above that, snow is the primary water source for many parts of the world. About one-sixth of the world depends on water from snowmelt to support life. Snow is often located in remote regions, and we often rely on remote sensing technologies to measure snow properties. Understanding snow properties is challenging because on the one hand, the historical pattern of snow accumulation and snowmelt has been altered due to global warming. On the other hand, remote sensing technologies do not always provide accurate estimates of snow properties due to the high variability in the snow both spatially and temporally. The need for precise prediction motivates the use of Machine Learning techniques for understanding snow properties. In this study, we survey articles that have applied Machine Learning techniques to predict Fractional Snow Cover, Snow Depth, and Snow Water Equivalent. The result of our survey indicates that Neural Networks, Random Forest, and Support Vector Machines are the most used techniques for estimating snow properties. Also, current gaps in application of Machine Learning to snow science are discussed and some open questions are presented.

Keywords: *Fractional Snow Cover, Snow Depth, Snow Water Equivalent, Machine Learning, Neural Networks, Random Forest, Support Vector Machine*.

1 Introduction

Snow is an essential component of the Earth's climate system because it directly influences the balance between the energy absorbed by the Earth and the energy reflected back into the atmosphere (the Earth's energy balance) due to its high radiation reflection ability (high albedo) and its thermal properties. Snow reflects about 80% to 90% of the total sunlight hitting its surface [3] and as a result, it regulates the planet's temperature. Also, snow acts as an insulating blanket to the ground beneath it by preventing heat and moisture from escaping into the atmosphere and it reduces the risk of wildfire.

Snow plays a crucial role in the water cycle. When the snow melts, water from the melt winds up in reservoirs, streams, rivers, lakes, and eventually oceans in many regions across the world. Snowmelt serves as the major source of water

for many regions across the world. About one-sixth of the world's population (1.2 billion people) relies on water from snowmelt for survival [43]. Snowmelt is the water source for some regions within northern latitudes and higher elevations across Europe and North America [8, 27, 47]. The Sierra Nevada seasonal snowpack in the state of California provides an additional 70% of water storage to the existing artificial reservoir system [23].

Almost everyone on the planet benefits from the climatic advantages of snow [43]. However, the increase in greenhouse gases emission has altered the historical pattern of snow accumulation and snowmelt, thereby threatening natural water sources [48]. Immense and uninterrupted snow can also be the cause of natural disasters such as avalanches and flooding [8, 16, 47]. In 2017, the Oroville lake of California experienced an intense inflow which led to the damage of the reservoir's spillway [23]. The spillway damage resulted in the forced evacuation of 188000 people and a $1 billion infrastructure repair cost [5, 23]. Subsequent research showed that the surge of runoff resulted from heavy snowmelt, and the heavy snowmelt resulted from the unusual warming of the atmosphere [5]. More of these catastrophic events are estimated to occur due to climate warming [5].

Due to the warmer atmosphere and the complex structure of snow data sets, traditional approaches for forecasting snow properties have been proven difficult [7] and biased [27]. Passive Microwave (PM) sensors produce real-time estimates of snow water equivalence (SWE) in mountainous regions but suffer from loss of signal in wet snow [34], saturation in deep snow [21, 26, 45], and SWE overestimation when the grain size is large [19]. Also, the Snow Data Assimilation System (SNODAS) which is a snow product operating on 1 km spatial resolution [11] produces biased daily SWE estimates when compared with the ground truth [27]. Since extreme snow events are disastrous - little or no snow may lead to drought, and intense snowfall may lead to floods, forecasting snow properties with minimal uncertainties is paramount. Also, a precise depiction of snow-covered area (SCA) and snow water equivalent (SWE) in mountainous terrain is required to improve streamflow timing [46].

Several approaches have been used to study and predict snow properties by combing remote sensing data and in-situ observations [47]. These approaches include both physics-based models such as energy-balance models and statistical models such as linear regression. Both approaches have their setbacks. Physics-based models have lots of unknown parameters which are extremely difficult to tune and linear regression does not fully capture the nonlinear patterns present in the remote sensing data and in-situ observations. For example, snow depth is often underestimated when linear models are used to study the relationship between snow depth and brightness temperature [47].

The poor performance of energy-balance models and linear regression drove the need for new approaches for estimating snow properties to better inform water resource management[22]. For this reason, recent researches have used the tools offered by Machine Learning (ML) to predict snow properties. Machine Learning approaches have been applied to many parts of the earth's system, such as land, ocean, and atmosphere [31]. They have been proven helpful for both regression and classification tasks and supervised and unsupervised tasks [31].

Machine Learning techniques commonly found in the literature for snow-related studies include linear regression, bagged regression, decision trees, support vector machines, random forest, and artificial neural networks [31]. For example, artificial neural networks have been used to predict snow depth by combining brightness temperatures and in-situ observations [10, 42]. Artificial neural networks, however, have also been combined with model simulations for predicting snow depth [47]. In addition, the random forest learning machine has been used for regional bias correction of snow water equivalent estimates [27]. Artificial neural networks have started to dominate snow-related research within a few years of incorporation.

With this study, our goal is to survey and review previous articles which have applied machine learning for estimating snow properties. Specifically, we categorize snow properties into;

1. Fractional snow cover

2. Snow depth
3. Snow water equivalent

To achieve our goal, we conducted a systematic study in which we reviewed research papers that applied machine learning to estimating snow properties. To that end, we surveyed research publications from Google Scholar and Web of Science for corresponding research works, and we identified 5 seed papers. Details of the seed papers are presented in Table1.

Table 1: Details of Seed Papers

Name	**Journal**	**Snow Property**	**Region of Focus**
A physiographic approach to downscaling fractional snow cover data in mountainous regions	Remote Sensing of Environment	Fractional Snow Cover	Southwestern Idaho
Application of machine learning techniques for regional bias correction of snow water equivalent estimates in Ontario, Canada.	Hydrology and Earth System Sciences	Snow Water Equivalent	Ontario
Direct Insertion of NASA Airborne Snow Observatory-Derived Snow Depth Time Series Into the iSnobal Energy Balance Snow Model	Water Resources Research	Snow Water Equivalent	California Sierra Nevada
Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach	Journal of Hydrology	Snow Depth	Alaska
Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan	The Cryosphere	Snow Water Equivalent	Afghanistan

The organization of the rest of the study is as follows: In Section Two, a review of the application of Machine Learning to retrieving snow products is presented, and in Section Three, current gaps in ML application and future paths are discussed.

2 Machine Learning Applications

In this section, we present a review of previous works that have applied Machine Learning techniques to tackle geosciences and remote sensing problems related to fractional snow cover (*FSC*), snow depth (*SD*), and snow water equivalent (*SWE*). Our review is focused on accurately describing the ML techniques used in the papers and not necessarily the remote sensing concepts. However, all remote sensing concepts required for understanding the ML techniques are described.

2.1 Fractional Snow Cover

The spatial extent of snow (snow cover area) is crucial in the management of hydrological regimes, forest fire risk assessment, and recreational demands [9]. Snow cover is variable both spatially and temporally, and it is often concentrated in remote regions [6]. Due to the complex topography of remote regions, it is difficult to monitor snow cover change using manual field-based techniques; making space-borne remote sensing the most feasible approach to

measure and monitor variation in snow cover [6, 36]. In general, snow cover estimates are represented by two mapping approaches; the *binary map approach* (snow/no snow) and the sub-pixel *fractional snow cover map approach* [29]. Fractional snow cover is the percentage of a pixel covered by snow, and as a result, it takes values between 0, 1 with both boundaries included. In this study, our review focuses on the sub-pixel fractional snow cover mapping approach.

The two instruments commonly used for capturing snow-covered area estimates are the space-borne Landsat remote sensing system and the Moderate Resolution Imaging Spectroradiometer (MODIS). The Landsat program is a joint program between the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS). The goal of the Landsat program is to acquire satellite imagery of the earth's surface [32]. The Landsat program launched its first satellite (Earth Resources Technology Satellite) in 1972 [1]. In 1975, the satellite was renamed Landsat 1 and subsequent satellites are named in sequential order. For example, Landsat 7 is the seventh satellite of the Landsat program. Presently, Landsat 8 is the latest satellite that provides earth imagery data. The ninth satellite of the Landsat program (Landsat 9) was successfully launched into space on September 27, 2021 [2]. Its data will be publicly available in early 2022 [2]. Every Landsat satellite launched has a remote sensing instrument aboard it. For example, the Thematic Mapper (TM) is the remote sensing instrument aboard Landsat 4 - 5, Enhanced Thematic Mapper Plus (ETM+) is the remote sensing instrument aboard Landsat 7, and Operational Land Imager (OLI) is the remote sensing instrument aboard Landsat 8. The Landsat fractional snow-covered area data is available for the Western U.S. and Alaska for Landsat 4 through 8 [6].

The Landsat system provides snow-covered area estimates processed to 30-meter spatial resolution with a 16-day return time (temporal resolution) [46]. By 30-meter spatial resolution, we mean one pixel on the Landsat image corresponds to a square of 30 by 30 meters on the ground, and by a 16-day temporal resolution, we mean the same image is taken every 16 days. On the other hand, MODIS, an instrument aboard NASA's Aqua and Tera satellites maps snow cover at a much higher temporal resolution but a lower spatial resolution. MODIS provides daily snow cover estimates processed to 500-meter spatial resolution [46]. Both Landsat and MODIS are able to provide data/pixel from which sub-pixel fractional snow cover can be extracted.

Both systems have spatial or temporal hindrances. The Landsat system has a high spatial resolution of 30 meters but a low temporal resolution of 16 days. The MODIS system has a lower spatial resolution of 500 meters with a daily return time. Because of its high temporal resolution, MODIS is the frequently employed instrument in snow cover mapping [29]. However, many applications require snow cover estimates at a much higher resolution than 500 meters because snow cover varies extensively within 500 meters. For example, hydrologic models in mountainous terrain often employ 30 meters of spatial resolution data provided by digital elevation models [46]. The Landsat system on the other hand provides snow cover estimates at a desirable spatial resolution but its temporal resolution is insufficient because snow cover can also vary extensively within 16 days. Because of the high variability in snow cover, fractional snow cover estimates are often desired at Landsat temporal resolution and MODIS spatial resolution.

In an attempt to get the best of both worlds, previous works have downscaled the fractional snow cover data obtained from the MODIS system. Walters et al. [46] developed an efficient algorithm for downscaling melt-season fractional snow-covered area data from the MODIS Terra system to a higher spatial resolution of 30 meters in southwestern Idaho, USA. The developed algorithm was based on physiographic characteristics (elevation, slope, and aspect) derived from digital elevation models (DEMs). The algorithm is based on the idea that the distribution of snow-covered areas in a partially snow-covered region is non-random (snow distribution patterns are relatively constant every year) and can be predicted using terrain physiographic features. The algorithm uses 13 Landsat binary snow cover maps (a combination of Landsat 5 and Landsat 7 images) for a region in southwestern Idaho for calibration and was validated against two other independent regions each in the same scene, approximately 100 km away. The calibration images were compiled over snowmelt seasons (January to May) between 2000 and 2011 and the validation was done for a date not used in calibration. The performance of the proposed model was assessed via the F1 measure (since they did a

binary classification of snow/no snow). The proposed model outperformed the baseline model which was an ensemble of randomly generated snow cover maps.

Traditional techniques for deriving fractional snow cover from MODIS and Landsat systems include Normalized Difference Snow Index (NDSI) and Normalized Difference Vegetation Index (NVDI) [40]. NDSI takes maximum advantage of the difference in snow reflectance in the visible band and the shortwave infrared band. It is computed as the difference in snow reflectances observed in the visible band and the shortwave infrared band divided by the total of the two reflectances [41]. In keeping with the convenient notation introduced in Walters et al. [46], NSDI is written mathematically as in Equation (1) with R_{vis} and R_{swir} representing the reflectance in a visible band and the reflectance in a short-wave infrared band respectively:

$$\text{NSDI} = \frac{R_{vis} - R_{swir}}{R_{vis} + R_{swir}}. \tag{1}$$

The Normalized Difference Vegetation Index (NDVI) on the other hand uses the same principle as the NDSI to compute vegetation properties. Using the notation introduced in Assefa et al. [37], NDVI is written mathematically as in Equation (2) with *NIR* and *RED* representing near-infrared and red reflectances respectively:

$$\text{NDVI} = \frac{NIR - RED}{NIR + RED}. \tag{2}$$

In recent publications, Machine Learning techniques have been increasingly used for fractional snow cover mapping [14, 16, 24, 38]. This is an attempt to improve FSC estimates due to the high variability that could occur in snow cover estimates spatially and temporally. The most used Machine Learning algorithm for mapping fractional snow cover so far is Artificial Neural Network [29]. Other commonly used ML techniques include support vector machines and random forests [31].

Dobreva and Klein [16] applied artificial neural networks (ANNs) for deriving FSC from remote sensing products with the aim of improving FSC mapping in forests. To implement the study, a one hidden layer feed-forward neural network was trained using MODIS surface reflectance band 1 through 7 (the first seven bands at 500 m resolution in the visible and near-infrared portion of the electromagnetic spectrum), NDSI, NDVI, and land cover as inputs. The land cover, generated by the *International Geosphere-Biosphere Programme* (IGBP) is a categorical variable with 17 levels. However, the the 17 levels can be reclassified into 10 levels as shown in Figure 1. The trained model was validated with FSC maps derived from Landsat Enhanced Thematic Mapper Plus (ETM+) scenes within North America. In choosing the training and test scenes, the study only considered flat terrain with no inclusion of complex mountainous terrain. In order to minimize land cover or snow cover fraction training bias, sample snow maps were selected through stratified random sampling. A total of 9 Landsat snow maps (Figure 1) representative of the various land cover characteristic of snow covered area in Northern America were used to create the training and validation sets. The testing set constitutes of four Landsat Enhanced Thematic Mapper Plus scenes which were not included in the training scenes (Figure 1).

The final training data consists of 27,647 observations (generated by combining the pixels from all of the training images). These observations were randomly split into three disjoint sets (1/2, 1/4, 1/4). The first 1/2 observations served as the training set, the second 1/4 observations served as the validation set, and the last 1/4 observations served as a withheld test set. To arrive at the final architecture of the neural network, the number of hidden neurons, the hidden activation function, and the normalization method for pre-processing the training data were determined via trial and error. The neural network was trained with several combinations of these hyperparameters and the final model has the following characteristics;

- Normalization approach: each continuous input variable is scaled to take values in the range [-1, 1].
- Architecture: 10 input neurons, 20 hidden neurons, and 1 output neuron

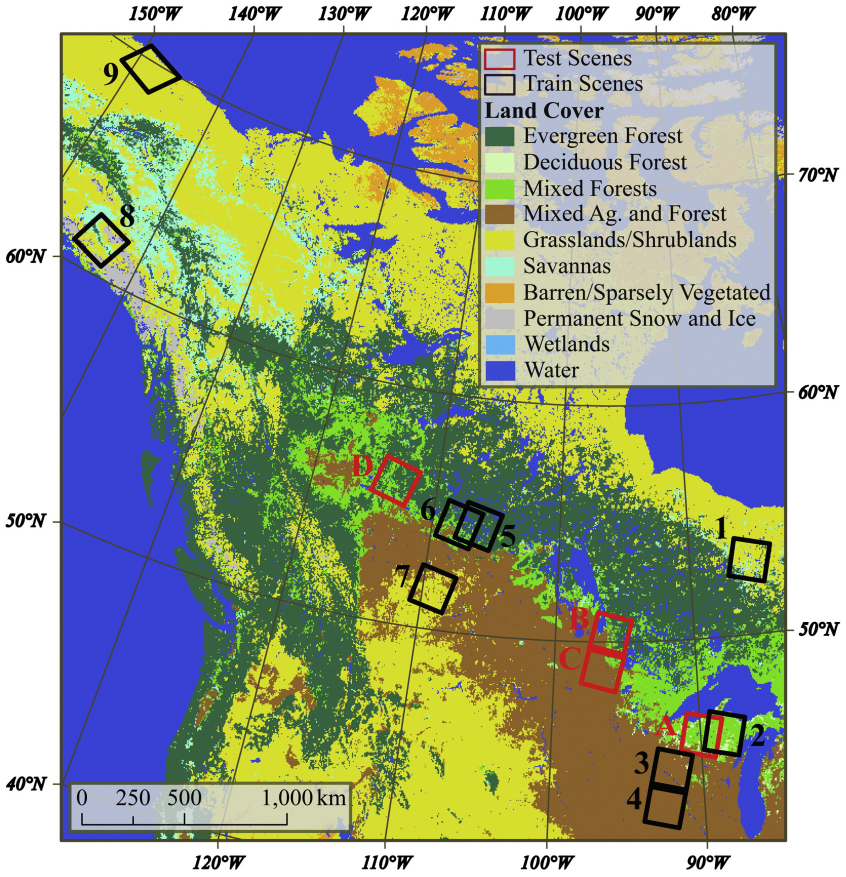

Figure 1: Training and testing scenes were selected to represent different land covers (Dobreva and Klein [16])

- Hidden activation: hyperbolic tangent
- Loss function: mean squared error
- Loss optimization method: gradient descent with momentum
- Back propagation algorithm: Levenberg–Marquardt backpropagation

In addition, the final architecture was trained for two different scenarios of the land cover variable. In the first scenario, the land cover was used as one categorical variable, and in the second scenario, it was dummy coded. The performance of the trained neural networks was evaluated on six different test sets using the root mean square error (RMSE) and coefficient of determination (R^2) metrics. The first test set is the withheld 1/4 observations from the training set and the second through fifth test sets are the four Landsat Enhanced Thematic Mapper Plus scenes from Figure1. The sixth test set is the total of all observations of the four Landsat ETM+ scenes. The neural networks returned test R^2 values range from 0.7 to 0.9 and RMSE values ranging from 0.12 to 0.22.

Kuter [29] employed the support vector regression (SVR) and the random forest regression (RFR) for estimating FSC from MODIS Terra data and compared the performance of these algorithms with previously proposed artificial neural networks (ANNs) and multivariate adaptive regression splines (MARS) by Kuter et al. [30] over a complex and heterogeneous alpine environment. The two algorithms were chosen because they are ubiquitously used in remote sensing applications. In order to make accurate results comparison, the same image data and the same experimental design was used for both studies. To implement the study, 20 MODIS - Landsat 8 image pairs obtained over European Alps were used as dataset. Of the 20 image pairs, 15 scenes served as training set and the remaining 5 scenes served as

independent test set. Also, an independent test set was formed by adding the observations of the five independent test sets. Image pairs were selected for a duration of five months (December to April) because these months correspond to the entire snow season. The training set consists of three images from each month and the test set consist of one image from each month. Just as the work of Dobreva and Klein [16], the study uses MODIS reflectance values of bands 1–7, NDSI, NDVI, and land cover class as the independent variables and FSC values as the dependent variable.

The training dataset consists of a total of 921,175 observations. In order to evaluate the dependency of the model on the data size, nine different training sets were generated from the total observations via simple random sampling and stratified random sampling. Before training the support vector regression and random forest regression, all independent variables were projected to the interval [0, 1] to improve generalization. The random forest was trained using 500 regression trees in the forest and the minimum number of observations per tree leaf node was set to 5. As for the support vector regression, both the polynomial kernel and the radial basis kernel were used. The performance of the models was evaluated using two metrics; the root-mean-square error (RMSE) and Pearson's correlation coefficient (R). The random forest model outperformed other models with a correlation value of 0.9 and an RMSE value of 0.13.

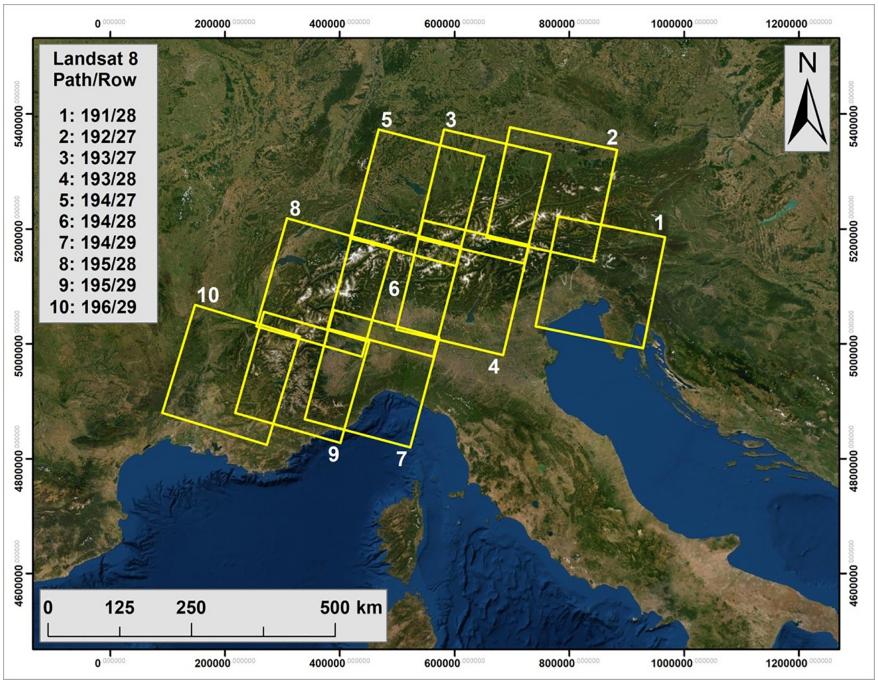

Figure 2: Locations of the Landsat 8 tiles over European Alps. (Kuter [29])

2.2 Snow Depth

Knowing the spatial extent of snow (i.e. snow cover area) is important, however, knowing the depth of snow is much more important [25]. For climate studies and applications such as weather forecast, disaster management, water delay planning, and hydrology in general, knowing snow depth is crucial [18, 35]. Despite the significance of snow depth, a quantitative estimate of the amount of snow stored in the mountains across the earth is still lacking [17]. Manually measuring snow depth is expensive, challenging, time-consuming, and potentially dangerous to field crews [15, 39]. Spatial intervals over which snow depth can be manually measured are limited to spacing which is likely not to capture snow variability, presenting a clear need for remote-sensing technologies [15]. Two remote sensing methods that are widely used for capturing snow depths are Passive Microwave (PM) and airborne Light Detection and Ranging (Lidar) technologies.

The Earth's surface emits microwave radiation and when satellite sensors detect such microwave radiation, we call it *the passive microwave* [4]. The extent of microwave radiation depends largely on the characteristics of the emitting surface. Snowpacks are able to emit microwave radiation at both the surface and deeper layers. Clouds do not emit microwave radiation, as such, microwave radiation can penetrate clouds and be used to detect snow depth during the day and night, regardless of cloud cover [4]. In snow-covered regions, satellite passive microwave radiometers measure microwave radiation in form of brightness temperature (T_b) [47]. Lidar on the other hand is a remote sensing technology that uses laser light to determine the depth of snow. While both methods have their shortcomings, the Lidar system produces more accurate snow depth estimates compared to the PM sensors, and in practice, Lidar data often serve as validation data [35]. One the one hand, Lidar is very expensive and on the other hand, PM suffer from loss of signal in wet snow.

Estimates of snow depths retrieved from passive microwave remote sensing technology are often used to establish the empirical relationship between brightness temperature (T_b) and snow depth [47, 49]. The most widely used approach for retrieving snow depths from PM remote sensing technology is the Chang algorithm [12]. The Chang algorithm uses the difference between the horizontally polarized brightness temperature frequencies of 18 and 37 GHz to estimate snow depth. Using the convenient notation introduced by Yang et al. [50], the Chang algorithm can be represented as in Equation (3), where SD is the snow depth, Tb_{18H} and Tb_{18H} are horizontally polarized brightness temperatures at 18 and 37 GHz channels respectively, and 1.59 is a constant obtained by assuming a grain size and snow density of 0.30 mm and 0.30 bcm^3 respectively.

$$\text{SD} = 1.59 \times (\text{Tb}_{18H} - \text{Tb}_{37H}). \tag{3}$$

The Chang algorithm poses two major challenges; it is a linear algorithm, and it is not always reliable in all regions due to the fixed empirical constant [13, 47, 50]. More sophisticated algorithms have been proposed but they are all computationally expensive and rely on complex auxiliary data to provide predictions with less error [49]. Since the Chang algorithm is linear, it does not fully capture the nonlinear relationship between snow depth and brightness temperature, snow depth tends to be underestimated [20]. The poor performance of traditional methods drove the need for a better approach for estimating snow depth, hence Machine Learning.

Wang et al. [47] established a deep learning based approach to estimate snow depth by combining satellite data and onsite measurements over Alaska. They chose deep learning based approach because it has the capability capture nonlinear relationships and because it has rarely been used for estimating snow depth. To obtain the satellite data set, two satellites were used; the Ground-based global Navigation Satellite System Reflectometry (GNSS-R) and the Special Sensor Microwave Imager/Sounder (SSMIS). Daily horizontally and vertically polarized brightness temperatures from 2008 to 2017 were obtained from SSMIS and onsite snow depth data from 2008 to 2017 for 155 sites were obtained from the Global Historical Climate Network (GHCN). Snow depth data obtained from GHCN include *latitude, longitude, measurement time, and snow depth*. The study also obtained snow depth observations from (GNSS-R) and used it as the the reference (ground truth) because of its high accuracy for snow depth retrieval and its large spatial scale compared to onsite data. Forest cover fraction and elevation data were taken from Terra MODIS Vegetation Continuous Fields (VCF) product and ETOPO1 Global Relief Model respectively. Finally, data from all sources were consolidated into a dataset of nine predictors and the outcome variable snow depth.

The study compared five methods for snow depth estimation; Deep Belief Network (DBN), Chang Algorithm (Equation (3)), Multiple Linear Regression (MLR), Back Propagation Neural Network (BPNN), and Generalized Regression Neural Network (GRNN). For all models except the Chang Algorithm, snow depth was estimated using the functional form

$$\text{SD} = g(\text{Tb}_{19H}, \text{Tb}_{19V}, \text{Tb}_{37H}, \text{Tb}_{37V}, \texttt{lat}, \texttt{lon}, \texttt{elevation}, \texttt{f}, \texttt{time}). \tag{4}$$

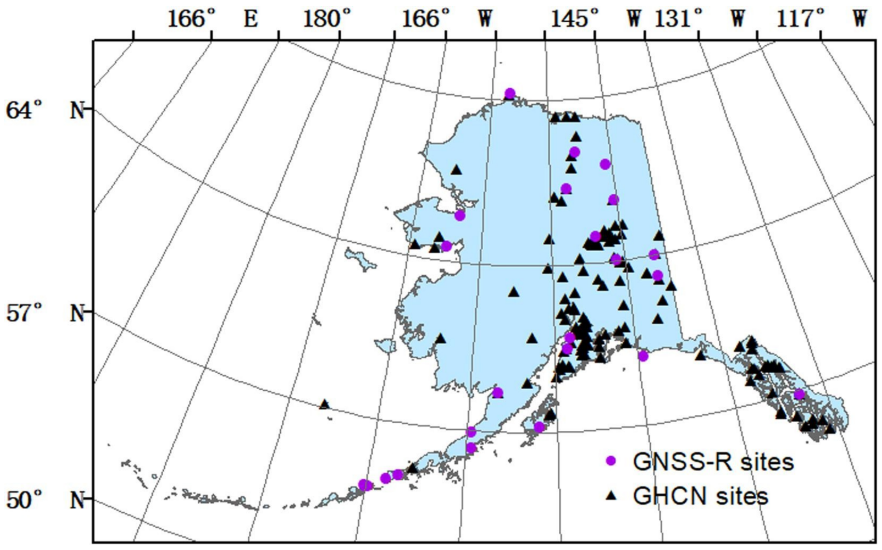

Figure 3: Study area (Wang et al. [47])

where $g(\cdot)$ is the learning machine, `f` is the forest cover fraction, `lat` is the latitude, and `lon` is the longitude.

The architecture of the DBN model consists of one input layer, two restricted Boltzmann machines (RBM) layers with 16 neurons each, one back propagation (BP) layer, and one output layer as shown in Figure 4

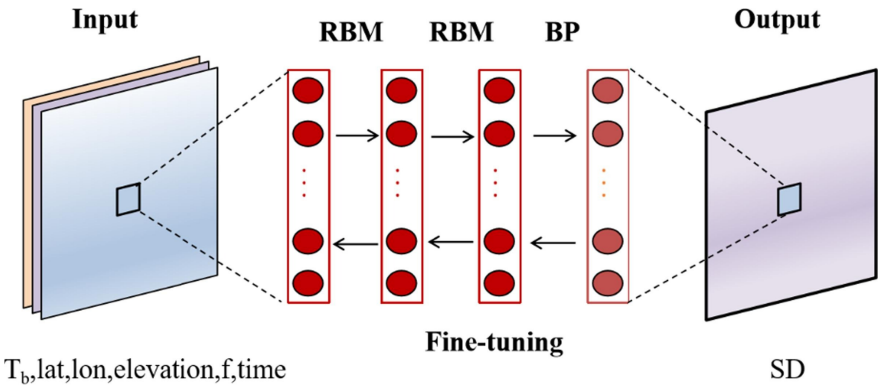

Figure 4: Structure of the Deep Belief Network (Wang et al. [47])

The DBN was trained in three major steps; the first step was to perform an unsupervised training on the predictor variables to obtain the initial weights, the second step involves training the model using gradient descent and wight adjustment using the backpropagation algorithm, and the third step was to perform 10-fold cross-validation and apply the trained network to estimate snow depth for places where ground stations do not exist. The training steps are depicted in Figure 5. The five models were trained in two phases, in the first phase, the onsite snow measurement served as the dependent variable (GNSS-R estimates serving as the independent test data) and for the second phase, both onsite estimates of snow depth and GNSS-R derived snow depth were combined and averaged to serve as the dependent variable.

To evaluate the model performance, three metrics were used; the correlation coefficient (R), the mean absolute error (MAE in cm), and the root-mean-square error (RMSE in cm). Results on the test set show that linear models (Chang and MLR) performed poorly compared to the nonlinear methods (DBN, BPNN, GRNN). In general, the DBM model

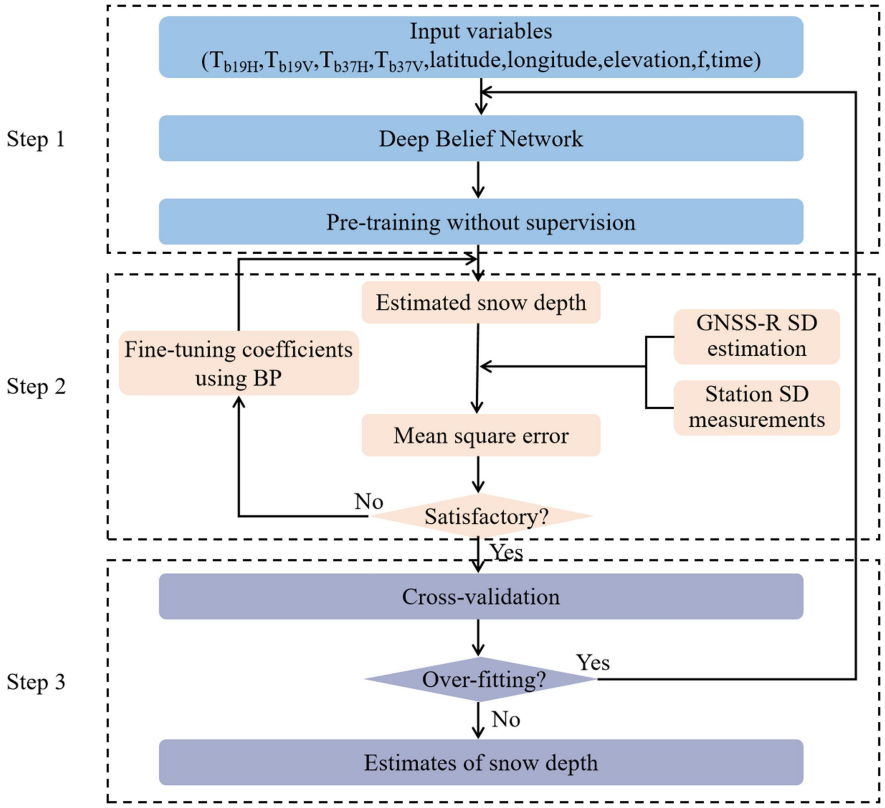

Figure 5: Flowchart of the Deep Belief Network (Wang et al. [47])

was found to be superior returning R, MAE, and RMSE values of 0.83, 10.16, and 17.18 respectively for the first phase training and 0.87, 8.80, and 14.39 respectively for the second phase training.

Yang et al. [49] investigated the abilitys of the random forest (RF) learning machine to estimate snow depth. To implement the study, both in-situ observations and satellite imagery data were combined. Vertically polarized passive microwave brightness temperatures from 1987 to 2018 were taken from Special Sensor Microwave/Imager (SSM/I) and Special Sensor Microwave Imager Sounder (SSMIS) instruments. In-situ data from 1987 to 2018 were obtained from China Meteorology Administration. The in-situ data consist of seven variables; *site name, observation time, latitude, longitude, altitude (m), near-surface soil temperature, and snow depth*. Additionally, land cover fraction data derived from Thematic Mapper (TM) imagery classification was also provided. Four random forest learning machines were built using different combination of the predictors. Details of the four random forest models can be found on Table 2.

Table 2: Details of the Four Random Forest Models (Yang et al. [49])

Name	**Predictors**	**Outcome**
RF1	Tb_{19v}, Tb_{37v}	Snow Depth
RF2	Tb_{19v}, Tb_{37v}, latitude, longitude	
RF3	Tb_{19v}, Tb_{37v}, latitude, longitude, elevation	
RF4	Tb_{19v}, Tb_{37v}, latitude, longitude, elevation, land cover fraction	

The four RF models were made of 1000 regression trees and the number of random variables used for splitting at each node was set to one-third of the predictors. Observations from 2012 to 2014 were used to train the RF models and

the performance of the four models was evaluated using three independent test sets; out-of-bag samples, temporal subset, and Spatio-temporal subset. The RF algorithm only uses about two-thirds of the training sample during training, the remaining observations not used for training is called the out-of-bag sample. The temporal subset consists of independent reference data from 2015–2018 and the Spatio-temporal subset consists of the spatially independent dataset from 2015 - 2018. The details of the train-test subset are given in Table

Table 3: Train-Test Subset Details (Yang et al. [49])

Name	**Test 1 (OOB)**	**Test 2 (temporal subset)**	**Test 3 (spatio-temporal subset)**
Training	date: 2012-2014 (28602 samples)	date: 2012-2014 (28602 samples)	date: 2012-2014 (28602 samples)
Testing	date: 2012-2014 (14301 samples)	date: 2015-2018 (34684 samples)	date: 2015-2018 (25879 samples)

The study employed three metrics for evaluating the performance of the models; root mean squared error (RMSE), bias, and correlation coefficient (R). Results from the fitted models indicated that the inclusion of auxiliary information improved the model performance. In general, RF1 performed poorly compared to the three other models while the performance of RF2, RF3, and RF4 was found not to differ significantly.

2.3 Snow Water Equivalent

Measuring the amount of water that will be produced by completely melting a snowpack (snow water equivalent) has posed a long-standing challenge [7]. In many practical applications where snow depth is known, snow water equivalent (SWE) is needed as well, however, measuring SWE is much more time-consuming and it is estimated to take approximately 20 times as long as the time for measuring snow depths [44]. Using remote sensing technologies for estimating SWE has also been proven difficult [28, 33] leaving ground-based observations the primary means for quantifying SWE [27]. Traditionally, SWE is defined in terms of snow depth and mass per unit volume (bulk density). Traditionally, at any given point, SWE can be estimated as follows;

$$\text{SWE} = h_s \cdot \frac{\rho_s}{\rho_w}. \tag{5}$$

where h_s is the snow depth (often measured in centimeters), ρ_s is the snow density (measured in grams per centimeters^3), and ρ_w is the density of water (1 gram per centimeters cubed). SWE estimates retrieved using Equation (5) are inaccurate because they do not always capture the full spatial variability of snow depth, driving the for a better estimation approach. Recent publications have explored the use of Machine learning approaches to improve the SWE estimates.

Bair et al. [7] developed two machine learning models (bagged regression trees and feed-forward neural networks) for predicting SWE in Afghanistan's watershed. The study employed reconstructed SWE as the outcome variable and a mix of physiographic and dynamic variables as predictors. All variables used in the study were sampled between 2003 and 2011. They were either computed or resampled to 3.125 km resolution using the Gaussian pyramid reduction or expansion approach. For example, vertically polarized passive microwave brightness temperature of 18 GHz was available at 6.25 km but was resampled to 3.125 km resolution. The choice of 3.125 km resolution was because it is the finest available resolution for 36 GHz vertically polarized passive microwave brightness temperature.

During data collection, only night time brightness temperature collected between 12:30 is - 1:30 am Afghanistan time were included. To predict reconstructed SWE, seven physiographic variables (day of the year, elevation, latitude, longitude, Northwest/west/southwest barrier difference, West/southwest distance to ocean, and southness) and four dynamic variables ($\text{Tb}_{18V} - \text{Tb}_{36V}$, $\text{Tb}_{10V} - \text{Tb}_{18V}$, fractional snow-covered area, and mean daily SWE reconstructed over all years except the target year) were used as the independent variables and the reconstructed SWE served as the dependent variable. Before fitting the ML models, fractional snow-covered area values of zero were removed because

a fractional snow-covered area value of zero directly implies an SWE value of zero. The study experimented with six learning machines (stepwise multiple linear regression, support vector machines, cross-validated regression trees, least-squares (LS) boosted regression trees, bagged regression trees, and feed-forward neural networks) on a randomly selected subset of the data and found bagged trees and feed-forward neural network to consistently outperform other models in terms of RMSE.

To obtain the final model, the method (bagged or Least Square boosted), number of learning cycles, the maximum number of splits, the minimum number of observations in the leaf node, and the number of predictors to randomly sample were optimized via cross-validation. The architecture of the feed-forward neural network was determined by testing different combinations of hidden neurons. The training was done each year from 2003 - 2011 separately and for each year, 90000 randomly selected observations were used with 80% of the 90000 observations serving as the training set and the remaining 20% serving as the test set. The trained models were evaluated using the bias and RMSE metrics. For the bagged regression tree, the RMSE ranged between 6 mm - 45 mm and for the feed-forward neural network, it ranged from 9 mm - 58 mm. Finally, predictor importance was determined using the bagged regression.

King et al. [27] employed machine learning techniques to bias-correct estimates of daily snow water equivalent from the Snow Data Assimilation System (SNODAS) in Ontario, Canada. Estimates of SWE from SNODAS were found to have a mean bias of 50% (16 mm SWE) and three different techniques to bias-correct SWE estimates from SNODAS using the absolute mean bias and RMSE as evaluation metrics. The four techniques considered in the study were mean bias subtraction (MBS), linear regression (LR), decision trees (DT), and random forest (RF). In keeping with the notation introduced in the study, Mean bias can be computed using Equation (6) where x_i and z_i are daily SNODAS and onsite estimates of SWE:

$$\text{MB} = \frac{1}{n}\sum_{i=1}^{n}(x_i - z_i). \tag{6}$$

The study performed both simple and multiple linear regressions. Simple linear regression was performed for all stations independently with SNODAS SWE estimates being the predictor and onsite SWE estimates being the outcome. For the multiple linear regression, decision tree, and random forest models, all predictors (elevation, month of observation, year of observation, 2m air temperature, total precipitation) were used to predict onsite SWE estimates. For both RF and DT a 10-fold cross-validation resampling was used to select the optimal combinations of hyperparameters. For both RF and DT, a maximum depth size of 15 was used and the RF model was made of 100 decision trees. The train-test split for all models was 75%-25% with the exception of MBS that doesn't require that the data is split into training and test sets. The result of the analysis indicated that RF was the most effective in reducing SNODAS bias amongst the compared method. The RF model in the study reduced the average bias across the region of study to $<$1 mm. The major conclusion of the study was that RF bias-corrected model is able to improve daily SWE estimates.

3 Gaps in Machine Learning Applications and Future Paths

In this study, we surveyed and reviewed a number of past publications that have used Machine Learning techniques to predict Fractional Snow Cover, Snow Depth, and Snow Water Equivalent. While our analysis is limited to the publications reviewed in our study, two major gaps in machine learning applications have been identified;

- *Lack of details for reproducible results*: while all authors showed keen interest in applying machine learning techniques to improve snow products, the lack of detailed explanation would make most of the results difficult to reproduce. Most of the works reviewed used neural networks, random forest, and support vector machines but failed to report how they arrived at the final model and even the hyperparameter values of the final

models. For example, Bair et al [7] mentioned that the number of hidden layers and the number of hidden neurons in each hidden layer they used for their feed-forward neural network were selected by testing different combinations on randomly selected subsets of the data but failed to report the number of hidden layers and number of neuron in each hidden layer of the final model. This is not to say the authors have not done a great job, we are only encouraging subsequent publications to add enough details to enable their works reproducible.

- *Lack of baseline models*: we have also noticed that only a handful of publications compared machine learning techniques to a simple baseline model. For example, Yang et al. [49]) compared four random forest models for estimating snow depth without using any simple traditional model as a baseline. Without simple baseline models, it would be difficult to define what true progress is. We know we are making progress when our machine learning techniques clearly outperform traditional baselines but how do we know when we don't make such a comparison?

3.1 Open Questions

While reviewing the publications for this study, we did notice that most of the data set were collected index to time (e.g. daily), however, none of the publications have considered analyzing the time series data using established time series techniques. Hence, our first open question is; would time series techniques (e.g. ARIMA and Recurrent Neural Network) improve our ability to predict snow products?

Also, most studies got their data using satellite imagery but did not use machine learning techniques known for image analysis. Hence our second open question is; would a convolutional neural network produce more accurate predictions of snow products since is used principally for image analysis?

References

[1] Landsat Missions. Landsat 1. https://www.usgs.gov/core-science-systems/nli/landsat/landsat-1?qt-science_support_page_related_con=0#qt-science_support_page_related_con, 2018. [Online; accessed 10-October-2021].

[2] National Aeronautics and Space Administration. Landsat 9. https://landsat.gsfc.nasa.gov/landsat-9, 2021. [Online; accessed 10-October-2021].

[3] National Snow and Ice Data Center. All about snow. https://nsidc.org/cryosphere/snow/climate.html, 2020. [Online; accessed 05-October-2021].

[4] National Snow and Ice Data Center. Remote sensing: Passive microwave. https://nsidc.org/cryosphere/seaice/study/passive_remote_sensing.html, 2020. [Online; accessed 14-October-2021].

[5] Scripps Institution of Oceanograpy. Heavy snowmelt delivered a surge of runoff to feather river and lake oroville. https://yubanet.com/regional/researchers-identify-factor-behind-2017-oroville-dam-spillways-incident/, 2021. [Online; accessed 27-September-2021].

[6] United States Geological Survey. Landsat fractional snow covered area. https://www.usgs.gov/core-science-systems/nli/landsat/landsat-fractional-snow-covered-area?qt-science_support_page_related_con=0#qt-science_support_page_related_con, 2020. [Online; accessed 09-October-2021].

[7] Edward H Bair, Andre Abreu Calfa, Karl Rittger, and Jeff Dozier. Using machine learning for real-time estimates of snow water equivalent in the watersheds of afghanistan. *The Cryosphere*, 12(5):1579–1594, 2018.

[8] Wouter R Berghuijs, Shaun Harrigan, Peter Molnar, Louise J Slater, and James W Kirchner. The relative importance of different flood-generating mechanisms across europe. *Water Resources Research*, 55(6):4582–4593, 2019.

[9] Ethan E Berman, Douglas K Bolton, Nicholas C Coops, Zoltan K Mityok, Gordon B Stenhouse, and RD Dan Moore. Daily estimates of landsat fractional snow cover driven by modis and dynamic time-warping. *Remote Sensing of Environment*, 216:635–646, 2018.

[10] Yungang Cao, Xiuchun Yang, and Xiaohua Zhu. Retrieval snow depth by artificial neural network methodology from integrated amsr-e and in-situ data—a case study in qinghai-tibet plateau. *Chinese Geographical Science*, 18(4):356–360, 2008.

[11] Tom Carroll, Don Cline, Greg Fall, Anders Nilsson, Long Li, and Andy Rost. Nohrsc operations and the simulation of snow cover properties for the coterminous us. In *Proc. 69th Annual Meeting of the Western Snow Conf*, pages 1–14. Citeseer, 2001.

[12] Alfred TC Chang, James L Foster, and Dorothy K Hall. Nimbus-7 smmr derived global snow cover parameters. *Annals of glaciology*, 9:39–44, 1987.

[13] Tao Che, Liyun Dai, Xingming Zheng, Xiaofeng Li, and Kai Zhao. Estimation of snow depth from passive microwave brightness temperature data in forest regions of northeast china. *Remote Sensing of Environment*, 183:334–349, 2016.

[14] Elzbieta H Czyzowska-Wisniewski, Willem JD van Leeuwen, Katherine K Hirschboeck, Stuart E Marsh, and Wit T Wisniewski. Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. *Remote Sensing of Environment*, 156:403–417, 2015.

[15] Jeffrey S Deems, Thomas H Painter, and David C Finnegan. Lidar measurement of snow depth: a review. *Journal of Glaciology*, 59(215):467–479, 2013.

[16] Iliyana D Dobreva and Andrew G Klein. Fractional snow cover mapping through artificial neural network analysis of modis surface reflectance. *Remote Sensing of Environment*, 115(12):3355–3366, 2011.

[17] Jeff Dozier, Edward H Bair, and Robert E Davis. Estimating the spatial distribution of snow water equivalent in the world's mountains. *Wiley Interdisciplinary Reviews: Water*, 3(3):461–474, 2016.

[18] KA Dressler, GH Leavesley, RC Bales, and SR Fassnacht. Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model. *Hydrological Processes: An International Journal*, 20(4):673–688, 2006.

[19] Michael Durand, Edward J Kim, Steven A Margulis, and Noah P Molotch. A first-order characterization of errors from neglecting stratigraphy in forward and inverse passive microwave modeling of snow. *IEEE Geoscience and Remote Sensing Letters*, 8(4):730–734, 2011.

[20] Thian Yew Gan, Roger G Barry, Mesgana Gizaw, Adam Gobena, and Rajagopalan Balaji. Changes in north american snowpacks for 1979–2007 detected from the snow water equivalent data of smmr and ssm/i passive microwave and related climatic factors. *Journal of Geophysical Research: Atmospheres*, 118(14):7682–7697, 2013.

[21] Steven Hancock, Robert Baxter, Jonathan Evans, and Brian Huntley. Evaluating global snow water equivalent products for testing land surface models. *Remote sensing of environment*, 128:107–117, 2013.

[22] Andrew R Hedrick, Danny Marks, Scott Havens, Mark Robertson, Micah Johnson, Micah Sandusky, Hans-Peter Marshall, Patrick R Kormos, Kat J Bormann, and Thomas H Painter. Direct insertion of nasa airborne snow observatory-derived snow depth time series into the isnobal energy balance snow model. *Water Resources Research*, 54(10):8045–8063, 2018.

[23] Brian Henn, Keith N Musselman, Leanne Lestak, F Martin Ralph, and Noah P Molotch. Extreme runoff generation from atmospheric river driven snowmelt during the 2017 oroville dam spillways incident. *Geophysical Research Letters*, 47(14):e2020GL088189, 2020.

[24] Jinliang Hou, Chunlin Huang, Ying Zhang, and Jifu Guo. On the value of available modis and landsat8 oli image pairs for modis fractional snow cover mapping based on an artificial neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4319–4334, 2020.

[25] Yanxing Hu, Tao Che, Liyun Dai, and Lin Xiao. Snow depth fusion based on machine learning methods for the northern hemisphere. *Remote Sensing*, 13(7):1250, 2021.

[26] Richard E Kelly, Alfred T Chang, Leung Tsang, and James L Foster. A prototype amsr-e global snow area and snow depth algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2):230–242, 2003.

[27] Fraser King, Andre R Erler, Steven K Frey, and Christopher G Fletcher. Application of machine learning techniques for regional bias correction of snow water equivalent estimates in ontario, canada. *Hydrology and Earth System Sciences*, 24(10):4887–4902, 2020.

[28] Max König, Jan-Gunnar Winther, and Elisabeth Isaksson. Measuring snow and glacier ice properties from satellite. *Reviews of Geophysics*, 39(1):1–27, 2001.

[29] Semih Kuter. Completing the machine learning saga in fractional snow cover estimation from modis terra reflectance data: Random forests versus support vector regression. *Remote Sensing of Environment*, 255:112294, 2021.

[30] Semih Kuter, Zuhal Akyurek, and Gerhard-Wilhelm Weber. Retrieval of fractional snow covered area from modis data by multivariate adaptive regression splines. *Remote Sensing of Environment*, 205:236–252, 2018.

[31] David J Lary, Amir H Alavi, Amir H Gandomi, and Annette L Walker. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10, 2016.

[32] Donald T Lauer, Stanley A Morain, and Vincent V Salomonson. The landsat program: Its origins, evolution, and impacts. *Photogrammetric Engineering and Remote Sensing*, 63(7):831–838, 1997.

[33] Dennis P Lettenmaier, Doug Alsdorf, Jeff Dozier, George J Huffman, Ming Pan, and Eric F Wood. Inroads of remote sensing into hydrologic science during the wrr era. *Water Resources Research*, 51(9):7309–7342, 2015.

[34] Zhong-Xin Li. Modelling the passive microwave remote sensing of wet snow. *Progress In Electromagnetics Research*, 62:143–164, 2006.

[35] Hans Lievens, Matthias Demuzere, Hans-Peter Marshall, Rolf H Reichle, Ludovic Brucker, Isis Brangers, Patricia de Rosnay, Marie Dumont, Manuela Girotto, Walter W Immerzeel, et al. Snow depth variability in the northern hemisphere mountains observed from space. *Nature communications*, 10(1):1–12, 2019.

[36] Changyu Liu, Xiaodong Huang, Xubing Li, and Tiangang Liang. Modis fractional snow cover mapping using machine learning technology in a mountainous area. *Remote Sensing*, 12(6):962, 2020.

[37] Assefa M Melesse, Wossenu Abtew, and Gabriel Senay. *Extreme Hydrology and Climate Variability: Monitoring, Modelling, Adaptation and Mitigation*. Elsevier, 2019.

[38] Vahid Moosavi, Hossein Malekinezhad, and Bagher Shirmohammadi. Fractional snow cover mapping from modis data using wavelet-artificial intelligence hybrid models. *Journal of Hydrology*, 511:160–170, 2014.

[39] Roy Rasmussen, Bruce Baker, John Kochendorfer, Tilden Meyers, Scott Landolt, Alexandre P Fischer, Jenny Black, Julie M Thériault, Paul Kucera, David Gochis, et al. How well are we measuring snow: The noaa/faa/ncar winter precipitation test bed. *Bulletin of the American Meteorological Society*, 93(6):811–829, 2012.

[40] GA Riggs, DK Hall, and MO Roman. Modis snow products collection 6 user guide (version 1.0). *NASA Goddard Space Flight Center: Greenbelt, MD, USA*, 2016.

[41] Vincent V Salomonson and I Appel. Estimating fractional snow cover from modis using the normalized difference snow index. *Remote sensing of environment*, 89(3):351–360, 2004.

[42] E Santi, S Pettinato, S Paloscia, P Pampaloni, G Fontanelli, A Crepaz, and M Valt. Monitoring of alpine snow using satellite radiometers and artificial neural networks. *Remote sensing of environment*, 144:179–186, 2014.

[43] Matthew Sturm, Michael A Goldstein, and Charles Parr. Water and life from snow: A trillion dollar science question. *Water Resources Research*, 53(5):3534–3544, 2017.

[44] Matthew Sturm, Brian Taras, Glen E Liston, Chris Derksen, Tobias Jonas, and Jon Lea. Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology*, 11(6):1380–1394, 2010.

[45] Matias Takala, Kari Luojus, Jouni Pulliainen, Chris Derksen, Juha Lemmetyinen, Juha-Petri Kärnä, Jarkko Koskinen, and Bojan Bojkov. Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment*, 115(12):3517–3529, 2011.

[46] Reggie D Walters, Katelyn A Watson, Hans-Peter Marshall, James P McNamara, and Alejandro N Flores. A physiographic approach to downscaling fractional snow cover data in mountainous regions. *Remote sensing of environment*, 152:413–425, 2014.

[47] Jiwen Wang, Qiangqiang Yuan, Huanfeng Shen, Tingting Liu, Tongwen Li, Linwei Yue, Xiaogang Shi, and Liangpei Zhang. Estimating snow depth by combining satellite data and ground-based observations over alaska: A deep learning approach. *Journal of Hydrology*, 585:124828, 2020.

[48] Xuejiao Wu, Tao Che, Xin Li, Ninglian Wang, and Xiaofan Yang. Slower snowmelt in spring along with climate warming across the northern hemisphere. *Geophysical Research Letters*, 45(22):12–331, 2018.

[49] Jianwei Yang, Lingmei Jiang, Kari Luojus, Jinmei Pan, Juha Lemmetyinen, Matias Takala, and Shengli Wu. Snow depth estimation and historical data reconstruction over china based on a random forest machine learning approach. *The Cryosphere*, 14(6):1763–1778, 2020.

[50] Jianwei Yang, Lingmei Jiang, Shengli Wu, Gongxue Wang, Jian Wang, and Xiaojing Liu. Development of a snow depth estimation algorithm over china for the fy-3d/mwri. *Remote Sensing*, 11(8):977, 2019.