Check for updates

Hydrological Processes

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Machine Learning Approach for Snow Depth Estimation From Temperature Sensors

Madison Gunn[1] | James S. Mills[2] | Michael Mahoney[2] | Colin Beier[2] | Tao Wen[1] | Samuel E. Tuttle[1]

[1]Syracuse University, Syracuse, New York, USA | [2]SUNY College of Environmental Science and Forestry, Syracuse, New York, USA

Correspondence: Madison Gunn (mwoodley@syr.edu)

Received: 22 January 2025 | **Revised:** 4 September 2025 | **Accepted:** 11 September 2025

Funding: This work was supported by the NY State Energy Research & Development Authority (NYSERDA) Environmental Research Program, which fully underwrites long-term ecosystem monitoring programs at Huntington Wildlife Forest (SUNY ESF Newcomb Campus) via agreements #122681 (2018–2022) and #199214 (2023–2027). We would also like to acknowledge partial funding from a Syracuse University CUSE grant for some of the instrumentation used in this study and the Syracuse University Department of Earth & Environmental Sciences Prucha Fund for partially funding site visits to collect data.

Keywords: machine learning | random forest | snow depth | snow temperature | temperature profile

ABSTRACT

Snow is an effective, natural insulator and the differences in its internal temperature dynamics compared to soil and atmosphere allow for estimation of snow depth from snow temperature measurements. We use temperature sensor profiles to estimate snow depth for monitoring multiple winter seasons in a remote 1.3 km^2 (130 ha) forested watershed in the Adirondack Mountains, New York, United States. Vertical temperature sensor profiles were installed in a grid pattern in 2019 to monitor snow energy state and soil microclimate. Each profile consists of iButton temperature sensors enclosed in PVC pipe at 20 cm vertical spacing, of which eight profiles were paired with trail cameras and snow stakes for daily snow depth estimation starting in November 2021. An additional four temperature profiles with sensors exposed directly to the snow at 10 cm vertical sensor spacing were added in November 2022. We use photographs paired with temperature profiles to train random forest (RF) machine learning models to estimate snow depth from snow temperature profiles and landscape properties. Comparison of our RF model predictions versus camera-derived snow depths shows that we can accurately infer snow depth with a root mean squared error (RMSE) between 1.8 and 6.5 cm, which is lower than or comparable to existing methods. Our random forest method demonstrated effectiveness in an area with a shallow snowpack and frequent midwinter melt events, and showed little sensitivity to sensor mounting method, vertical sensor spacing, or time of day.

1 | Introduction

Snow cover distribution and snow depth are important variables in energy balance models and are often used to determine shifts in regional and global climate and soil microclimates (Bojinski et al. 2014; Lawrence and Slater 2010). With the global cryosphere changing and rain becoming more common during the winter months in some regions (e.g., northern and western United States of America (U.S.) and the Canadian Arctic), changes in snow depth have major implications for hydrological regimes such as spring streamflow prediction and negative impacts on local ecology (Bintanja and Andry 2017; Berghuijs et al. 2014; Contosta et al. 2019; Dudley et al. 2017; Siirila-Woodburn et al. 2021). Other important impacts from the decrease in seasonal snow cover on regional hydrology and ecology include the timing and length of springtime snow melt, loss of snow water resources, increased chance of soil freezing due to snowpack loss, and disturbance to tree health in hardwood forests (Berghuijs et al. 2014; Comerford et al. 2013; Contosta et al. 2019; Dudley et al. 2017; Kapnick and Hall 2012; Musselman et al. 2021; Reinmann et al. 2019). Changes in snow cover and snow depth are especially important in forest-dominated regions as snow

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Hydrological Processes* published by John Wiley & Sons Ltd.

can persist or melt faster depending on canopy influences on snowpack (Dickerson-Lange et al. 2017; Lundquist et al. 2013; Revuelto et al., 2015; Varhola et al. 2010). Within the northern hemisphere, over 40% of snow-covered areas are in a forest (either montane or boreal) environment (Sturm and Liston 2021).

In forest environments, snowpacks are strongly affected by canopy cover type, canopy density, and the influence of canopy cover on shortwave and longwave radiation, which causes differential accumulation and melting depending on canopy type (Gleason et al. 2021; Golding and Swanson 1986; Hardy et al. 2004; Lundquist et al. 2013; Storck et al. 2002; Varhola et al. 2010). Forested snowpacks, in the U.S. and other midlatitude regions, have experienced warmer winters in the last few decades with more frequent high temperature anomalies, which often lead to more snow melting events throughout the winter snow season (Diaz et al. 2003; Gleason et al. 2021; Hu and Nolin 2020; Musselman et al. 2021; Serreze et al. 1999; Wieder et al. 2022; Wilson et al. 2024). With the increase of midwinter snowmelt events, there is often a reduction in snow cover which depletes water resources in the spring and summer seasons for some communities (Burakowski et al. 2008; Contosta et al. 2019; Diaz et al. 2003; Siirila-Woodburn et al. 2021). Additionally, the increased chance of rain during the winter, usually in the form of rain-on-snow events, often leads to higher rates of snow melting and flooding, sometimes over multiple days (Jones and Perkins 2010; Serreze et al. 1999).

To better understand and monitor multi-seasonal changes in forested snowpacks, vertical profiles of temperature sensors (i.e., a series of sensors arranged vertically through the expected depth over which a snowpack accumulates) can be used to infer snow depth and soil microclimate, with the timing of snow melt events as a potential added value (Dafflon et al. 2022; Lewkowicz 2008; Reusser and Zehe 2011). Other established methods exist to monitor snow depth, but there are advantages and disadvantages to each. Ultrasonic and laser snow depth sensors are more expensive than an array of temperature sensors and require supporting infrastructure (e.g., external power supply), but are highly accurate (Ryan et al. 2008). Timelapse cameras paired with snow stakes (e.g., Breen et al. 2024) can provide snow depth information at an equal or lesser cost than temperature sensor profiles and at good accuracy (error within a few centimetres). However, cameras are not infallible, as data can be lost due to cameras being installed at a bad angle or from snow, fog, or condensation obscuring the lens. These issues do not affect temperature sensors. Temperature sensors also tend to have longer battery life (up to many years), while cameras that do not have a consistent power source often need battery recharging or replacement in less than 1 year. While developing a sensor network requires front-loaded cost, as individual low-cost temperature sensors can range from US $50 to US $150 per unit (Dafflon et al. 2022), temperature sensors require comparatively minimal maintenance costs. Therefore, both temperature sensor arrays and timelapse cameras are cost-effective options for multi-site snow depth observation networks that require long deployments, for example, in remote areas.

Temperature profiles also provide information about the snow energy state and cold content of a snowpack (i.e., the energy required to raise a snowpack to 0°C) that other snow depth monitoring methods (e.g., ultrasonic sensors, timelapse cameras paired with snow stakes) do not, which can be useful for snow modelling and snowmelt prediction (e.g., Raleigh et al. 2013). Snow temperature can be measured manually (e.g., in an excavated snow pit), but temperature sensors provide many advantages because they can be programmed to run for multiple years even in remote settings such as the Canadian Arctic (e.g., Tutton and Way 2021) with minimal maintenance, and do not require disturbance of the snowpack.

Previous authors have recognised that vertical profiles of temperature sensors provide information that can be used to infer snow depth. These methods, like ours, take advantage of the fact that snow acts as an insulator, so temperature variability within the snowpack is more muted than in the overlying atmosphere. Within the snow, the most temperature variability occurs near the snow surface, where it interacts most directly with the overlying atmosphere. The least variability occurs at the base of the snowpack, where it is in contact with the ground surface (which experiences little temperature variability). Lundquist and Lott (2008) used sensors buried 2–20 cm beneath the ground surface to determine if the area was snow-covered or bare during the winter. Lundquist and Lott (2008) found that when the snowpack was shallow or patchy, soil temperature oscillations were dampened compared to air temperatures.

Other work utilising temperature sensors to infer snow depth has been mostly focused on statistical methods for snow depth prediction. For example, Reusser and Zehe (2011) inferred snow depth from temperature data without the need for climatic data or snow models. They used sensors spaced every 15 cm from 0 to 120 cm above the ground and the reduction of the diurnal temperature variation from snow insulating the sensors to infer snow height. They showed a mean absolute error of 6 cm for their snow heights compared against a reference measurement. A more recent study by Tutton and Way (2021) used a new method that combined light and temperature methods with change point analysis (e.g., detecting the timing of statistically significant changes in the magnitude or trend of a variable) to determine snow depth. They found strong correlations among their four methods (e.g., raw light, interpolated light, interpolated light threshold or mean, and raw temperature change) with better performance in the combined light and temperature methods compared to temperature-only methods, yielding a new cost-effective method that worked relatively well in a remote study region of the Canadian Arctic. Another recent study, Dafflon et al. (2022), used a low-cost distributed temperature profiling (DTP) system across several locations in the East River watershed in Colorado for snow depth estimation. To generate results, they calculated the temperature gradient between consecutive sensors and used a 24-h moving window to identify the highest temperature difference between a pair of sensors at a given time and then selected the lowest height of the identified sensor pair for their snow depth estimate. Dafflon et al. (2022) DTP results showed strong agreement with a nearby Butte SNOTEL station, despite their DTP setup being at a lower elevation than the SNOTEL site and with lower temporal and spatial resolution.

Although our study is not the first to estimate snow depth from vertical profiles of temperature sensors, past studies have not

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

demonstrated snow depth accuracy that reaches or improves what can be achieved using timelapse cameras or ultrasonic sensors, so there is still room for improvement. Questions also remain about the optimal approach for deploying these installations as part of snow monitoring efforts. Our study aims to identify a method that can fully realise the potential of snow temperature measurements for more comprehensive (both snow depth and temperature), accurate, and cost-effective measurements of snow properties. For this work, we use a new approach of training machine learning models to estimate snowpack depth from a network of temperature profiles deployed in a forested watershed in northeastern New York, U.S. We use a multi-season analysis of vertical profiles of air/snow/soil temperature to establish the effectiveness of this methodology. We execute a series of supplementary machine learning models to address various outstanding questions about the use of temperature profiles for snow depth inference and factors affecting random forest model accuracy, such as:

- Does the type of temperature sensor installation (e.g., shielding sensors from ambient radiation within PVC pipe versus sensors directly exposed to the air/snow/soil) have a notable influence on snow depth prediction accuracy?
- How does the vertical resolution of temperature sensors (e.g., 10 vs. 20 cm spacing) impact snow depth prediction accuracy?
- Does snow depth prediction accuracy depend on the time of day (e.g., due to changing characteristics of the vertical temperature profile throughout the diurnal cycle)?
- Which variables are the most impactful for snow depth prediction (e.g., temperature at specific heights above the ground surface, variability in temperature over time at specific heights, or site-specific landscape characteristics)?

Our work aims to demonstrate an effective machine learning framework that could be applied in a wide variety of environments and snow conditions. Our results may be useful in informing the design of resilient, cost-effective temperature profiles and providing recommendations for the use of temperature profiles in snowpack monitoring and prediction efforts going forward.

2 | Data and Methods

2.1 | Study Site Description

This study was located within the Archer Creek Watershed, a long-term ecological monitoring site in the Adirondack Mountains, New York, USA (see Figure 1). The watershed is part of the Huntington Wildlife Forest, a 60 km^2 (6000 ha) research, teaching, and demonstration campus of the State University of New York, College of Environmental Science and Forestry. The climate in the Huntington Wildlife Forest is cool and humid, with a mean annual temperature of 5.3°C and a mean annual precipitation of 1.11 m (Beier et al. 2021). Archer Creek Watershed (1.35 km^2 in area) is the northern subset of the larger Arbutus Lake Watershed (3.5 km^2 in area) surrounding Arbutus Lake (see Figure 1b).

During the winter, the snowpack often reaches its maximum depth in early March, with an average annual maximum of 47 cm since 2009 based on meteorological measurements from within the Archer Creek Watershed (Beier et al. 2021). The elevation of the Archer Creek Watershed ranges from 521 to 743 m, with an average slope of 12° (21%) and a total relief of 222 m (Beier et al. 2021; Gomez et al. 2016; USGS 1995). The forest within this watershed is composed of uneven-aged hardwood trees (e.g., American beeches, sugar maples, and red maples) and mixed conifers (e.g., eastern hemlocks, white pines, and spruces), with conifers prevalent in topographic low areas (Beier et al. 2021). Streamflow gauges measure the outlet of the Archer Creek Watershed (which is also the inlet of Arbutus Lake), as well as two 'v' notch weir gauges (Figure 1b) for the headwater streams in the Archer Creek sub-watershed.

2.2 | iButton Temperature Profiles and Automated Cameras

In this analysis, we include temperature profile data from a total of 12 sites within the Archer Creek Watershed. Eight of the sites have temperature sensors enclosed within PVC pipe at 20 cm vertical spacing, excluding the 80 cm height, while the remaining four sites have temperature sensors exposed directly to the air/snow/soil surface at 10 cm vertical spacing. A full summary of the temperature profile installations is provided in Table 1.

In 2019, we installed vertical profiles enclosed in PVC pipe of iButton (Maxim Scientific models DS1921G-F5# and DS1922L-F5#) temperature sensors at eight sites in the Archer Creek Watershed (see Figure 1c for spatial distribution; see Table S1 for complete site details) to measure soil microclimate, with particular interest in snow accumulation and timing of springtime melt (Beier et al. 2021). For each profile, the iButton sensors were enclosed in 3.8 cm diameter PVC pipe housing and placed at specific heights using a plastic fob secured with a screw to an internal PVC rod that vertically bisected the housing (see Figure 2b). The iButtons were positioned every 20 cm, from −20 cm below the ground to 100 cm above the ground surface, excluding the 80 cm height, with six total iButtons on each profile. iButton temperature sensors are nickel-sized, all-in-one temperature sensors and data loggers with a temperature range of −40°C to +85°C at US $80–$100 per unit (i.e., model DS1922L) and US $40 to $50 (i.e., model DS1921G) per unit. Both these models feature the same temperature range, but the DS1922L has a larger data storage capacity and an option for a higher temperature resolution. For this work, the DS1922L was only used in the below ground heights. We programmed the iButton sensors to collect temperature measurements at a 4 h interval at 0.5°C resolution. For each of the eight sites, a pair of iButton profiles (i.e., two vertical temperature profiles per site, located within 5 m apart) were installed for redundancy in case of data loss. These paired, PVC-enclosed iButton temperature profiles will be hereafter referred to as the 'PVC arrays' (Figure 2b).

In November 2021, a graduated snow stake and either a Crenova HC1000A 16 MP Hunting Trail Camera or Apeman

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

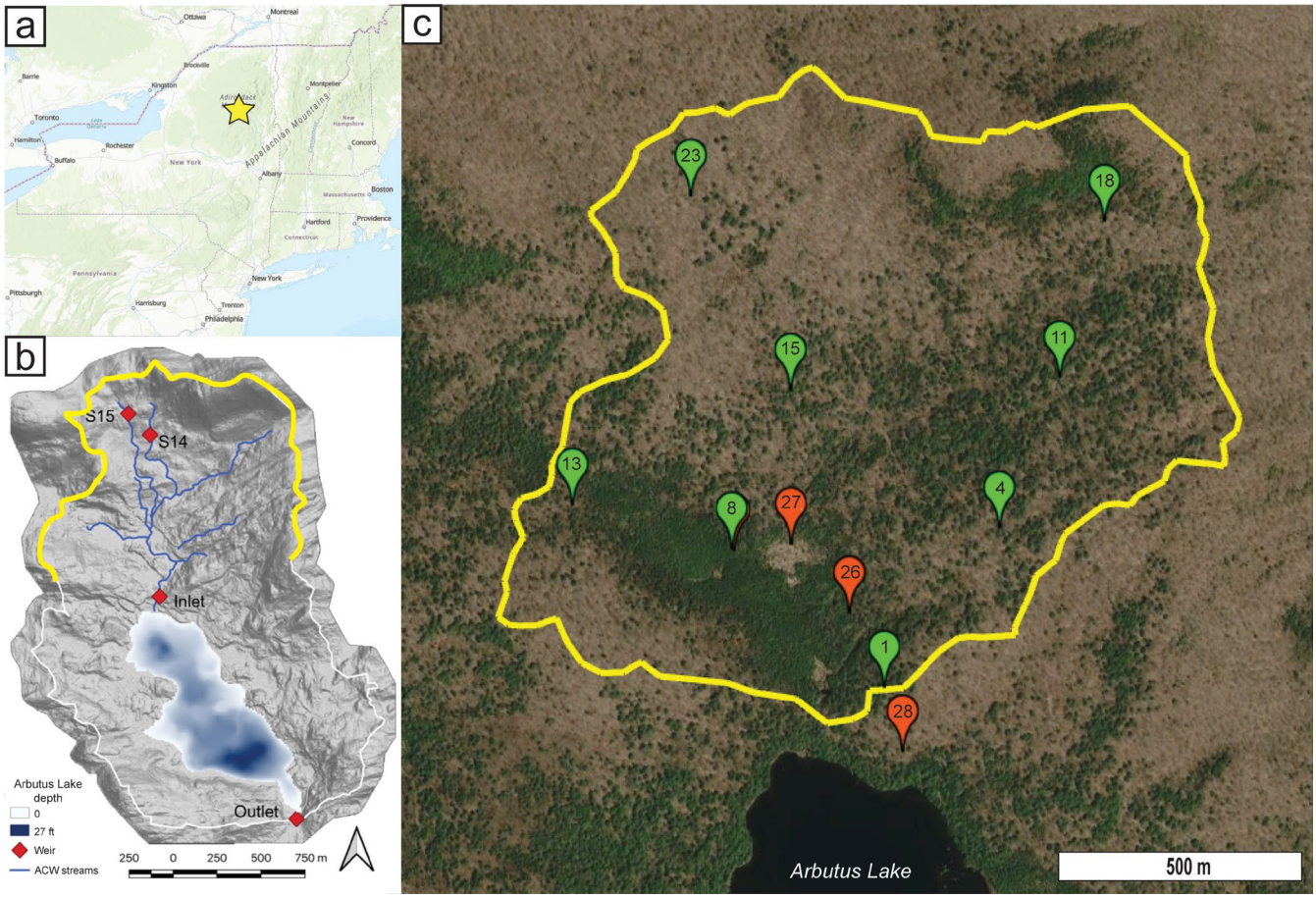

FIGURE 1 | (a) Inset topographic map of the northeastern United States with the study site of the Archer Creek watershed marked by a yellow star. (b) Inset map of the Arbutus Lake and its watershed with most of the Archer Creek sub-watershed border marked in yellow (original figure sourced from Beier et al. 2021). (c) Satellite image from May 2020 of the Archer Creek sub-watershed (yellow border) north of Arbutus Lake (partially shown at the bottom of image). Paired temperature profile arrays enclosed in PVC pipe and co-located with trail cameras and snow stakes are marked green. Single temperature profiles that are co-located with trail cameras and snow stakes but not enclosed in PVC are marked orange (Maps data: Google Earth, 2024 CNES/Airbus).

TABLE 1 | Summary information for iButton temperature sites with camera verification data in the Archer Creek Watershed in Newcomb, NY, USA.

Site ID	Camera data	Lat.	Long.	Elev. (m)	Aspect (°W)	Slope (°)	Canopy gap fraction
1	12/2021–05/2024	43.994	−74.243	540.9	270.0	13.1	0.25
4	12/2021–10/2022	43.997	−74.241	563.4	225.0	1.8	0.31
8	12/2021–05/2023	43.997	−74.247	525.7	195.5	2.7	0.19
11	12/2021–12/2022	44.000	−74.239	576.9	220.2	6.1	0.35
13	12/2021–07/2022	43.998	−74.252	533.8	77.8	5.4	0.30
15	12/2021–07/2022	44.000	−74.246	552.6	182.9	9.7	0.23
18	12/2021–05/2024	44.003	−74.238	626.8	186.5	14.9	0.32
23	04/2021–09/2022	44.003	−74.249	605.2	129.2	6.0	0.42
8e	11/2022–05/2024	43.997	−74.247	525.6	188.6	2.4	0.24
26	11/2023–05/2024	43.996	−74.244	521.3	188.1	0.5	0.26
27	01/2023–05/2024	43.997	−74.246	525.9	206.6	1.9	0.63
28	11/2022–05/2024	43.993	−74.243	520.9	202.2	8.3	0.43

Note: Sites numbered between 1 and 24 represent the eight paired PVC-enclosed temperature arrays, part of the larger snow monitoring network installed in 2019. Sites 8e and 26–28 are the four exposed temperature arrays, installed in mid-November 2022.

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

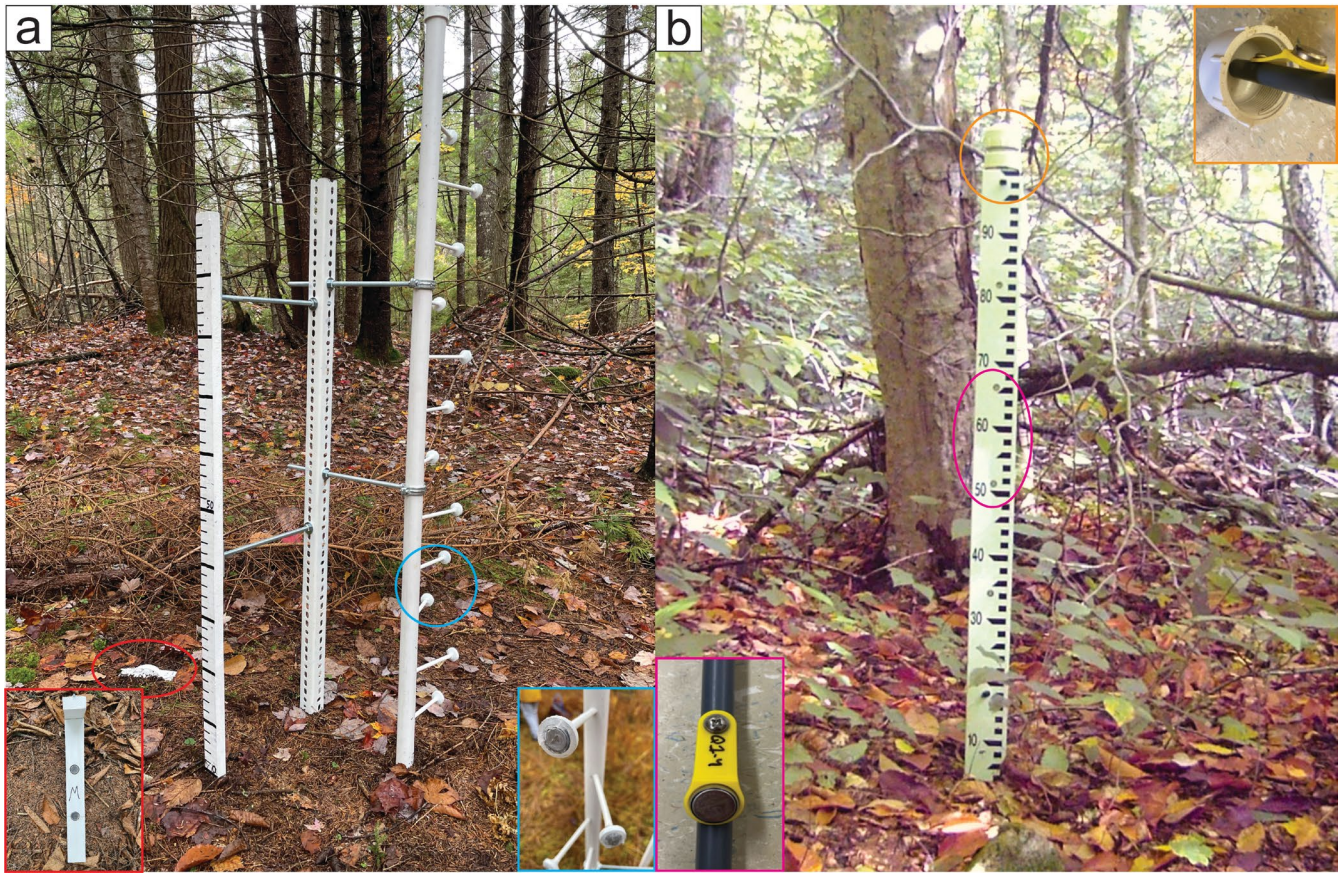

FIGURE 2 | Photos of the two vertical temperature profile types used in this study. (a) Image of an exposed iButton temperature profile (C2) with iButton temperature sensors facing north (right) and a marked snow stake (left) paired with a WingScapes Timelapse Camera (not pictured) for snow depth estimation. The bottom right inset image (outlined in blue) shows a close-up of aboveground temperature sensors, while the bottom left inset image (outlined in red) shows a close-up of below ground sensors from the exposed profile (outlined in red). (b) Image of a PVC iButton temperature profile (Site 23) with an attached snow depth stake. The bottom left inset image (outlined in pink) shows an internal view of an iButton temperature sensor within the PVC pipe, attached to a PVC rod that runs the length of the pipe. The top right inset image (outlined in orange) shows an internal view of the PVC rod attached to the top cap of the PVC pipe, along with an iButton sensor at the 1 m height.

H40 16 MP Trail Camera were added to each of the eight PVC arrays, with images collected once per day at noon local time for manual snow depth validation. We manually extracted snow depth from these timelapse images using visual inspection and the graduated snow stake as a reference. Since their initial installation, some cameras were found to be faulty or unresponsive or had to be replaced due to damage from large wildlife, such as black bears. This resulted in some years with more functional camera images than others, or some sites with camera images for only part of the winter. Full details on the PVC arrays are summarised in Table 1. These eight sites are part of a larger network of 24 PVC arrays throughout the watershed distributed in a grid pattern with approximately 225 m spacing (see Table S1), but we only include the eight sites that have camera imagery for at least one winter (between November 2021 and May 2023) that can be used for independent snow depth validation.

To assess the effect of enclosing the iButtons in PVC pipe, we installed a different iButton profile design in mid-November 2022 at three additional sites, with iButton sensors mounted on custom-built PVC seatings attached to horizontal 5 mm diameter carbon fibre rods embedded in a vertical PVC pipe. The carbon fibre rods were located every 10 cm in the vertical direction along the PVC pipe. For these three additional profiles, the iButtons were installed directly exposed to the air/snow/soil surface (see Figure 2a) versus being secured within PVC pipe housing. These profiles were co-located with snow depth stakes and WingScapes Timelapse Pro cameras. One profile (Site 8e) was placed in a heavily coniferous forest near one of the PVC arrays, Site 8 (see Table 1), while another was placed in a mostly deciduous section of the forest (Site 28), and a third profile was installed in an open grassy meadow (Site 27). These locations were chosen to represent the three end-member forest cover types within the Archer Creek Watershed. A fourth profile of the same type was installed in October 2023 at another site under a coniferous canopy (Site 26) to supplement a complimentary research project. Each of these new temperature profiles, hereafter referred to as the 'exposed' profiles, collects hourly temperature data and has iButtons spaced every 10 cm over the same −20 cm to +100 cm vertical range as the PVC profiles. The timelapse cameras for these exposed profile sites collect hourly images that correspond to the hourly iButton temperature data.

Additional data used in this analysis include upward-looking hemispherical canopy photos taken beneath the canopy at

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

each temperature profile site using a 180° fisheye lens on a cloudy day during late winter/early spring when there were no leaves on deciduous trees and no snow on tree branches. Images were processed using HemiView software from Delta-T Devices to calculate canopy gap fraction; a value between 0 and 1, where 0 represents a very closed canopy and 1 represents a completely open sky (see Table 1). Latitude and longitude of each temperature profile site were obtained from a handheld GPS unit (Garmin eTrex 20) and averaged using waypoint averaging for overall accuracy. The elevation of each temperature array site was extracted from a 10 m digital elevation model (DEM) of New York State produced by the U.S. Geological Survey (USGS 1995). Aspect and slope were calculated from the DEM in R statistical software using the 'terrain' function (version 3.6-26) of the 'raster' package (Hijmans et al. 2015).

2.3 | Random Forest Regression Model

We use random forest regression to predict snow depth from our vertical iButton temperature profiles. Random forest (RF) regression is a machine learning method that uses a simple ensemble of randomised decision trees that are averaged together to make a prediction (Breiman 2001). In RF models, users can fine-tune the model with hyperparameters, such as how many decision trees are used in an ensemble and the number of random variables at each node (Breiman 2001). Typically, 60%–70% of the samples in a RF dataset are randomly selected to train the decision trees while the remaining 30%–40% are set aside for evaluation (or 'testing') to see how well the RF model performed (Yang et al. 2020). Random forests automatically generate out-of-bag samples (OOB, i.e., samples that are set aside from a training dataset to evaluate model fit), and model prediction errors can also be generated from these OOB samples (Breiman 2001; Rodriguez-Galiano et al. 2015; Yang et al. 2020). If OOB values are not consistent with testing errors/validation datasets (e.g., increasing while testing values decrease), the random forest model is likely overfitting and not representative of the variable that is being predicted (Breiman 2001; Rodriguez-Galiano et al. 2015).

Here, we focus only on temperature profiles that had independent daily camera images of snow stakes to use as validation data for snow depth. Some PVC profiles were missing temperature data at one or more heights over the study period due to an iButton sensor reaching the end of its battery life and ceasing data collection. Because each PVC site had two identical temperature profiles, this allowed for some redundancy. If one of the profiles experienced iButton failure (i.e., missing data) at a given height, we used the profile that had the fewest data gaps overall during the time of interest (November 2021–May 2023). Then, we gap-filled the temperature data where needed using the second profile at the same site. After gap-filling using the second iButton profile at each given PVC site, six of our original eight PVC arrays (16.7% of the total 96 iButtons) still had a partial gap in data for some period of November 2021–May 2023 for at least one height (see Table S2 for specific dates). For these periods of missing temperature data, we gap-filled using multiple linear regression trained on other time periods where temperature data were not missing from that given height. The dependent variable in our regression was the temperature data from the missing height of the given profile, and the independent variables were temperature data from all other heights of the same profile. We fit the regression using the timesteps when temperature data were available for all heights in the profile and then used the fitted regression coefficients to predict the temperature for the height with missing data. For arrays that were missing more than one height, we first fit the regression to the height that was lowest to the ground. If the original data gaps overlapped, then we re-ran the regression again for the second height using the new gap-filled data. After these two gap-filling methods, no missing data remained in any of our temperature profiles.

To train our RF regression models, we used the 'ranger' R package (version 0.17.0; Wright and Ziegler 2017) with its default parameters to randomly select 70% of the timesteps of temperature data to use as a training set, while the remaining 30% was withheld for evaluation (or 'testing'). The target variable for our RF regression model was snow depth. We assigned the 'ground truth' for our target variable as the camera-derived snow depth values from the timelapse camera images, which we obtained manually from the timelapse camera imagery. The predictor variables in our RF regression models were iButton temperature (°C) for each height, a running 24-h standard deviation of iButton temperature at each height, elevation, aspect (degrees west), slope (degrees), and canopy gap fraction for the given temperature profile. We included the standard deviation of temperature at each height in the model predictors because sensors within the snow should show less diurnal temperature variance than sensors outside the snow because snow acts as an insulator.

We trained and evaluated one RF regression model that included 12 temperature profiles (eight PVC-enclosed and four exposed) to understand snow depth prediction accuracy using all available sites in our watershed. We ran six additional RF regression models to address random forest model accuracy specifically relating to the impacts of temperature profile design (PVC-enclosed vs. exposed), vertical sensor spacing (10 vs. 20 cm), and whether the time of day (morning vs. evening) affected predictive accuracy. For all models, we excluded the historically snow-free season (June through October) to prevent biasing the RF models by including extensive periods with no snow. The full details of each RF model, including which arrays, temperature sensor heights, and timesteps used in each RF model are noted in Table 2.

In addition to the standard testing statistics for RF regression methods (i.e., out-of-bag error, testing error), we conducted a set of leave-one-out (LOO) RF models for each model scenario. For each LOO RF model, we removed all of the data for one given site from the training dataset, re-fit the RF model without that site, and then evaluated how well the model could predict snow depth for the given site that was omitted. We repeated this procedure for every site used within a given RF model. This LOO analysis is analogous to applying a fitted RF model to a site in a similar environment with temperature profile data but without any validation data (such as a newly installed temperature profile site, or additional sites within the same network, for example, the 16 PVC temperature profile sites in the Archer Creek Watershed monitoring network that were not paired with cameras and snow stakes, see Table S1).

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Random forest regression models from this study with the name given to each model, the specific temperature array sites used, the time period included, and the temperature sensor heights used for each given model run.

Model experiment	RF model name	Sites used	Time period and timestep	Sensor heights used
All Sites	RF_Combined	PVC: 1, 4, 8, 11, 13, 15, 18, 23 Exposed: 8e, 26, 27, 28	11/2021–05/2024, 12 PM only	−0.2 to 1 m, every 20 cm excluding 0.8 m
Influence of Profile Design	RF_PVC	1, 4, 8, 11, 13, 15, 18, 23	11/2021–05/2023, 12 PM only	−0.2 to 1 m, every 20 cm excluding 0.8 m
	RF_Exposed	8e, 26, 27, 28	11/2022–05/2024, 12 PM only	−0.2 to 1 m, every 20 cm excluding 0.8 m
Influence of Sensor Spacing	RF_Exposed_20cm	8e, 26, 27, 28	11/2022–05/2024, hourly	−0.2 to 1 m, every 20 cm
	RF_Exposed_10cm	8e, 26, 27, 28	11/2022–05/2024, hourly	−0.2 to 1 m, every 10 cm
Influence of Time of Day	RF_Exposed_Morning	8e, 26, 27, 28	11/2022–05/2024, 8 AM only	−0.2 to 1 m, every 10 cm
	RF_Exposed_Evening	8e, 26, 27, 28	11/2022–05/2024, 8 PM only	−0.2 to 1 m, every 10 cm

For each of the seven RF models listed in Table 2, we ran 500 iterations, each one with a different random subset of 70% of the timesteps included in the training set, with the remaining 30% reserved for testing. We report the mean across these 500 iterations as our error statistics, which reduces potential bias from a random subset producing a training and testing set that was favourable or unfavourable to snow depth prediction accuracy simply due to random chance. For the LOO prediction error statistics, we report the mean across all 500 iterations at all sites (with error statistics for each site reported in Table S5). For all visualisations in our results and in the Supporting Information, we selected a RF model with error statistics that was the closest match to the mean prediction error statistics of the 500 iterations.

For each RF model, we evaluated each predictor variable's influence on the model predictions using Shapley value analysis from a combination of R packages, specifically 'shapr' (version 1.0.1) package on GitHub (Jullum et al. 2024), 'fastshap' (version 0.1.1; Greenwell 2024), and 'SHAPforxgboost' (version 0.1.3; Liu and Just 2020). Shapley value analysis represents the overall importance that a given variable has on the predictions from an RF model. Generally, the larger the spread of Shapley values for a given variable, the more influence that variable has on predictions from the RF model (Strumbelj and Kononenko 2010). We used these R packages to create 'beeswarm' plots of the Shapley values (Strumbelj and Kononenko 2010; see Figures S2–S8). Shapley values can be used to infer the directionality of each variable's influence on our RF model predictions. For example, within our RF models, higher temperatures at 0 m height were associated with negative Shapley values, which indicate that higher temperatures at 0 m height led to lower snow depth predictions from the RF model. Another example was the inverse relationship for low temperatures at 0 m height, which were associated with positive Shapley values, and thus higher snow depth predictions. This is consistent with the physical understanding of snowpack dynamics, as higher temperatures lead to a decreased snowpack.

3 | Results

Based on camera-derived snow depth observations from our period of interest (November 2021–May 2023), the snowpack in the Archer Creek Watershed reached its annual maximum depth in mid-February or in early March, depending on the year. The highest accumulation during our study period occurred after a large storm in March 2023, with snow depth reaching a maximum of 95 cm (at Site 27), near the upper limit measurable by our temperature profiles. The lowest annual maximum snow depth reached was 42 cm (at Site 15) on February 4th, 2022. Camera-derived snow depth observations and temperature data indicate that midwinter ablation periods were common occurrences and sometimes resulted in a complete melt-out of the snowpack, with the final spring snowmelt occurring by mid-April.

At one of our sites, both a PVC array (Site 8) and an exposed temperature profile (Site 8e) were installed, located approximately 5 m apart. We compared the temperature data from

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

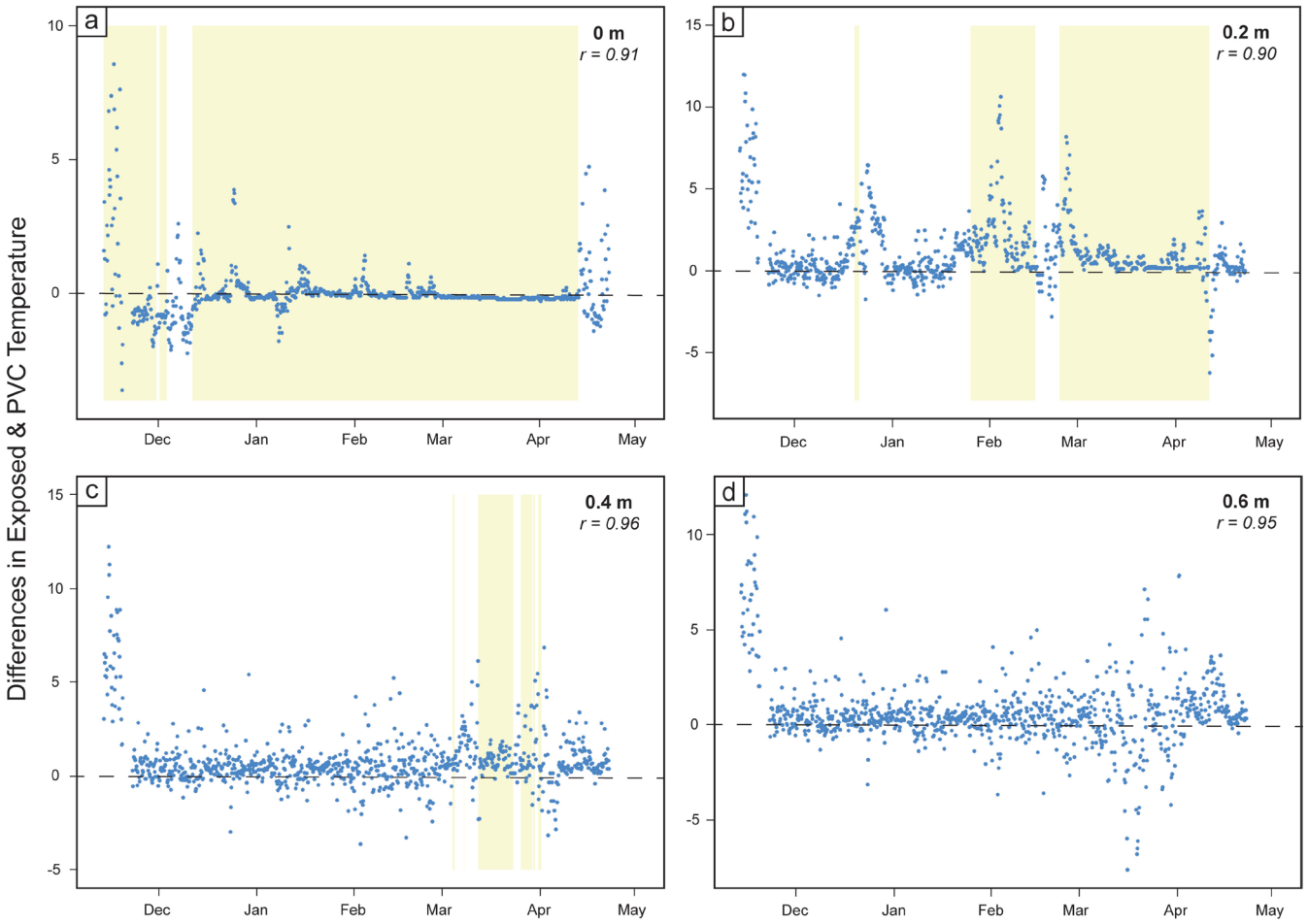

FIGURE 3 | Temperature differences between exposed sensors at Site 8e and PVC-enclosed iButton sensors at Site 8 (a coniferous canopy) for corresponding heights of (a) 0 m, (b) 0.2 m, (c) 0.4 m, and (d) 0.6 m, respectively. The two sensor arrays are approximately five meters apart. Each time series shown are at a 4-h timestep from November 2022–May 2023 with their respective Pearson's correlations. The yellow shaded region indicates when the snow depth was greater than or equal to the respective height.

these two temperature profile types and found an overall high correlation during the winter, with a Pearson's correlation (r) ≥ 0.90 for all heights at or above the ground surface (Figures 3 and S1). Exposed iButton sensors showed slightly warmer snow and air temperatures than sensors within the PVC pipe, especially during snow-free or low snow periods. The average temperature difference between the two temperature profile types (PVC enclosed vs. exposed) was 0.67°C ± 1.15°C during snow-free periods. When there was at least some snow on the ground (> 0 m), the temperature difference between the two profile types was less, averaging 0.03°C ± 0.93°C. Sensors below the ground at −0.2 m had the worst correlation of any measurement height between the two temperature profile types (Pearson's $r = 0.69$), due to the low variability of the soil temperature and a period early in the winter when the exposed iButton measured lower temperatures than the PVC-enclosed iButton (see Figure S1).

Note that perfect 1:1 correlation may be an unrealistic expectation due to small-scale variability in environmental conditions over the 5-m spacing between the two setups. Given the small differences in temperature measurements between the two temperature profile designs, we felt that it was reasonable to include the two different temperature profile setups within a single RF model (RF_Combined). However, we also separately evaluate two RF models, each restricted to a single profile design (RF_PVC and RF_Exposed, respectively), to further compare the snow depth prediction accuracy of each profile type.

3.1 | Random Forest Regression Models

To evaluate how accurately an RF model might predict snow depth across a diverse set of sites within a single watershed, we trained the RF_Combined model using camera-derived snow depth and temperature data at 12:00 PM local time at all sites (including both PVC and exposed profiles). The out-of-bag and testing errors show that exposed and PVC sensors spaced 20 cm apart can jointly predict snow depth with a reasonable accuracy of 6.2 and 6.1 cm RMSE (4.2 cm MAE), respectively. A summary table of the number of timesteps, out-of-bag, testing, and LOO average errors from each RF model run is included in Table 3.

The LOO average RMSE (analogous to applying a fitted RF model to a new site on which the model was not trained) for the RF_Combined model was higher than the OOB and testing errors, at 9.5 cm RMSE and 6.8 cm MAE. Site-specific RMSEs ranged

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Random forest regression models with number of timesteps, out-of-bag error, testing error, and leave-one-out error statistics. Error statistics are reported in units of cm.

Model experiment	RF model name	# of timesteps	Out-of-bag RMSE	Testing RMSE	Leave-one-out average RMSE
All Sites	RF_Combined	1882	6.2	6.1	9.5
Influence of Profile Design	RF_PVC	844[a]	6.1	6.1	10.1
	RF_Exposed	844	6.5	6.5	8.7
Influence of Sensor Spacing	RF_Exposed_20cm	20,105	2.2	2.2	8.8
	RF_Exposed_10cm	20,105	1.8	1.8	7.7
Influence of Time of Day	RF_Exposed_Morning	813	6.0	6.0	7.7
	RF_Exposed_Evening	860	5.8	5.8	7.1

[a]The full set of data available to the RF_PVC model is 1038 timesteps but was reduced to 844 and randomly sampled for direct comparison to RF_Exposed.

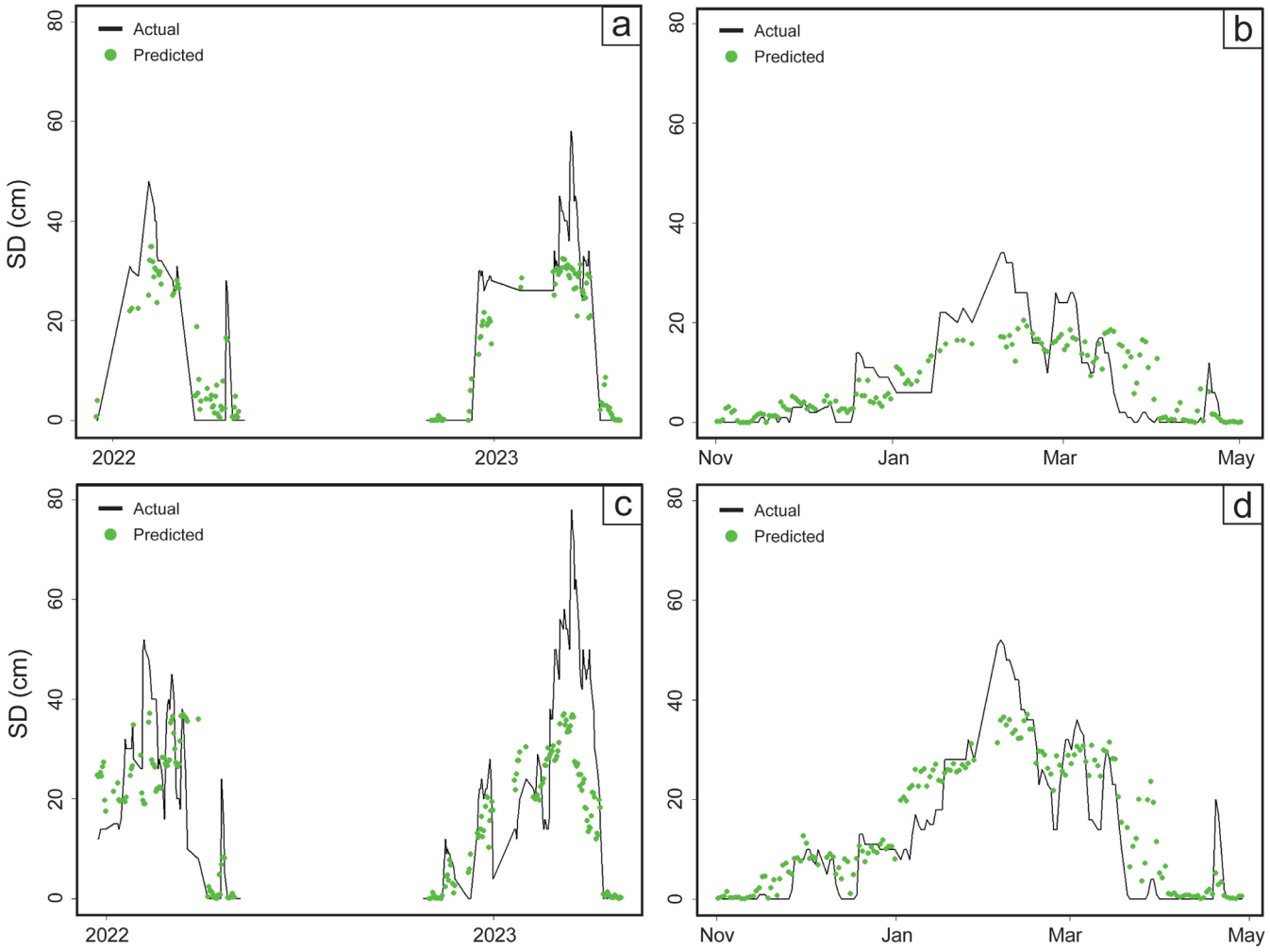

FIGURE 4 | Example plots showing snow depth predictions from RF_Combined compared to available camera data from November 1st, 2021, to May 1st, 2024, excluding the snow-free season (June through October). These models were trained on all sites except the array that is being tested (i.e., the array shown in each panel above). Each figure shows the camera-derived daily snow depths as 'Actual' with the black line and daily RF regression predicted snow depths at 12 pm as 'Predicted' as green points. The specific arrays shown are (a) Site 1, (b) Site 8, (c) Site 18, and (d) Site 23, respectively.

between 5.7 and 14.6 cm among the 12 sites (see Table S5), with four example sites shown in Figure 4. We observed lower RMSE for sites with higher canopy cover (i.e., lower gap fraction), although there was variability among sites and we do not have enough sites to be confident in this finding. These results suggest the RF model is more successful at predicting snow depth for sites at which it was trained, and extrapolating to apply the model to a new site leads to slightly higher errors.

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

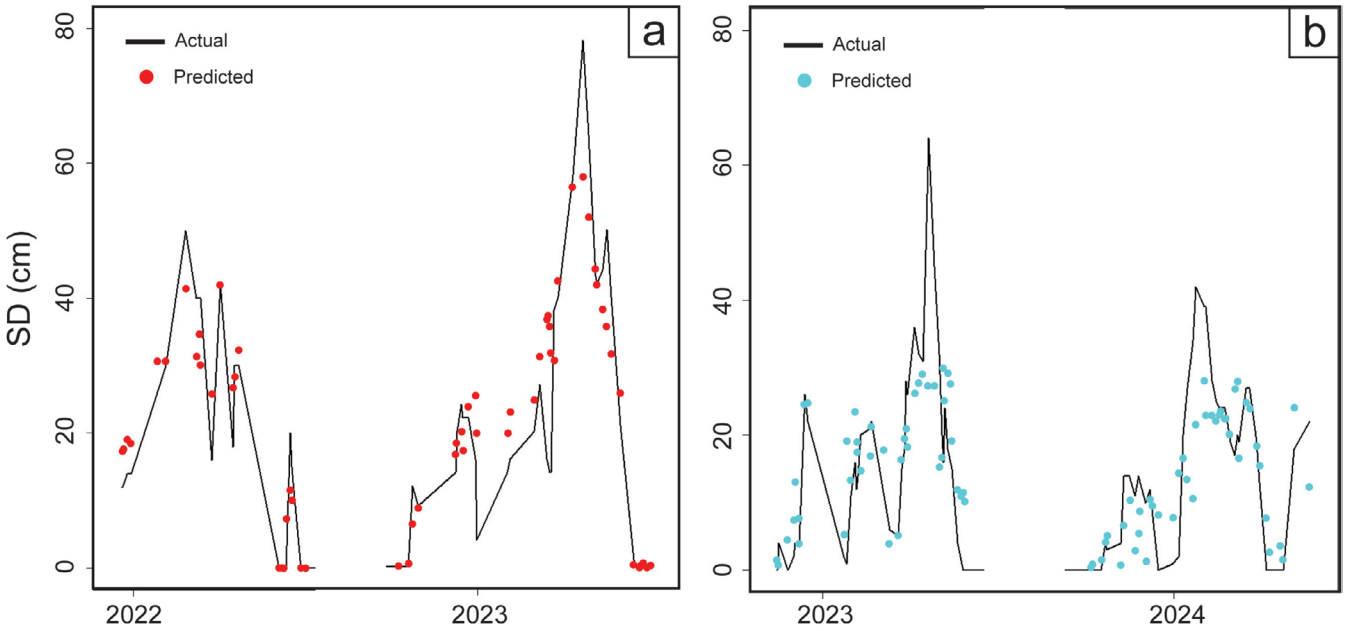

FIGURE 5 | Example plots of RF model predictions trained on all available sites compared to camera-derived snow depth (black). (a) Predicted snow depths from RF_PVC from November 1st, 2021 to May 1st, 2023 (red) versus camera data set aside for testing for Site 18. (b) Predicted snow depths from RF_Exposed versus camera data set aside for testing from November 1st, 2022 to May 1st, 2024 (blue) for Site 8e.

Next, we wanted to address whether thermal regulation from the PVC pipe of the PVC profiles had any effect on snow depth prediction accuracy, compared to the exposed profiles. We ran two models, RF_PVC (Figure 5a) and RF_Exposed (Figure 5b), and saw that they performed similarly to each other for all three error statistics (i.e., out-of-bag, testing, and LOO average errors) with less than 1 cm difference for out-of-bag and testing RMSE. The PVC profiles yielded slightly lower OOB and testing errors, but slightly higher LOO errors, compared to the exposed profiles. We note that the data included in the RF_PVC and RF_Exposed models were not from the same exact dates, due to lack of overlap of camera-derived validation data, but the sample size, sensor heights, and observation times were kept consistent between the two RF models. The errors for the RF_PVC and RF_Exposed models were similar to the RF_Combined model (see Tables 3 and S3), despite a smaller sample size for training. Thus, mixing exposed profiles with PVC profiles in the RF_Combined model does not yield increased snow depth prediction accuracy.

To determine whether vertical spacing has any effect on our predictions, we focused on only the exposed profiles, which have temperature sensors every 10 cm for four sites. We ran two models, RF_Exposed_10cm (Figure 6a) and RF_Exposed_20cm (Figure 6b), and found minimal difference between the two models. RF_Exposed_10cm performed slightly better than RF_Exposed_20cm in terms of out-of-bag and testing errors (< 0.5 cm difference in RMSE) and yielded a decrease of only 1 cm RMSE for the leave-one-out error. Based on these results, decreasing the sensor interval from 20 cm to 10 cm did not significantly improve (or worsen) snow depth estimation from RF models. The lower error for these two RF models, compared to all other RF models, is primarily due to the much larger sample size available for training (20,000+ timesteps versus < 1900 timesteps), given that all hours of the day were included in these two models. This was confirmed by a supplementary analysis where we ran RF_Exposed_20cm with multiple random subsets of the data and a range of sample sizes. We found that OOB and testing errors decreased exponentially with increasing sample size for training, while LOO errors decreased only minimally with increasing sample size. RF models trained on large datasets, such as RF_Exposed_20cm and RF_Exposed_10cm, perform best for snow depth prediction but are not easily generalised to new sites due to only four total sites within either model.

To evaluate whether time of day had any effect on our RF predictions, we focused on only the exposed profiles and separated them out by time of day (morning: 08:00 local time vs. evening: 20:00 local time). We ran two models, RF_Exposed_Morning (Figure 6c) and RF_Exposed_Evening (Figure 6d), and found that the time of day had only a small influence on RF accuracy. Using out-of-bag and testing errors, the evening model performed slightly better than the morning model, with RMSE values of 5.8 cm versus 6.0 cm, respectively. For the LOO analysis, the evening predictions performed slightly better than the predictions made in the morning, with an average RMSE of 7.1 versus 7.7 cm, respectively. Both models' errors were similar but slightly lower than for the RF_Exposed model, which had a similar sample size but different times of collection at 12 pm. Comparing all three models (RF_Exposed, RF_Exposed_Morning, and RF_Exposed_Evening) that use exposed iButtons profiles at one time per day timesteps shows a similar small range of errors. Temperature data for exposed RF models were somewhat insensitive to the time of day and thus differences in solar radiation and air temperature conditions.

Using Shapley value analyses (see Figures S2–S8 for beeswarm plots for each of the seven RF models), we found that temperature sensors near the ground surface (i.e., ≤ 0.4 m) generally had greater influence than landscape features such as elevation and aspect across all models, as indicated by the mean of the

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

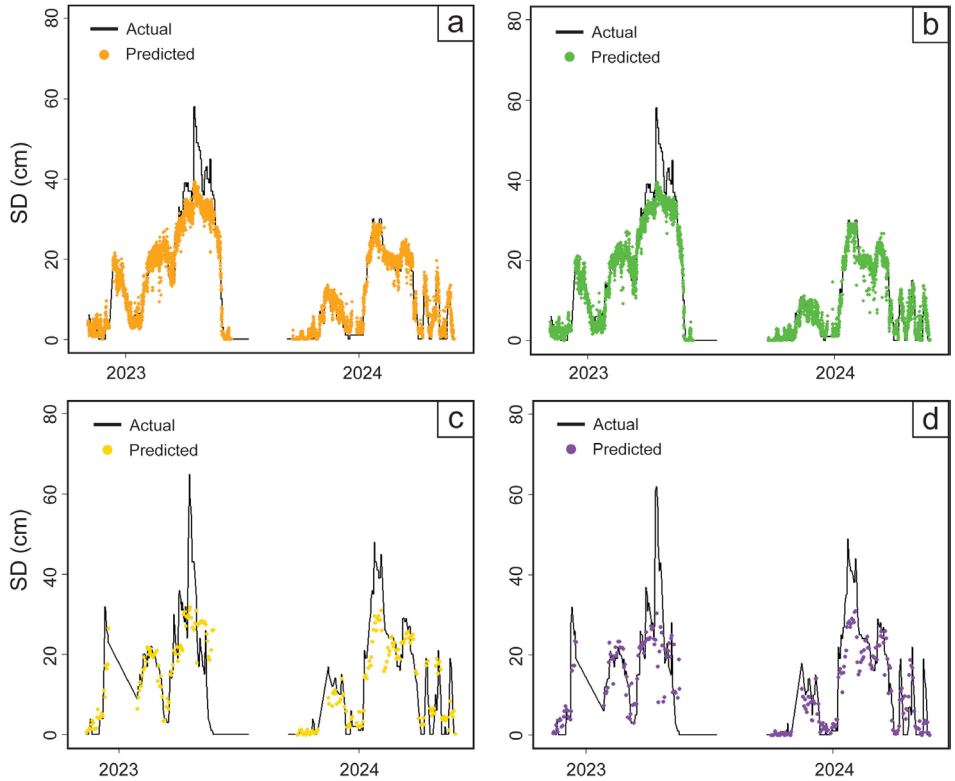

FIGURE 6 | Example plots of camera-derived snow depth (black, all plots) versus RF model predictions from all available sites. (a) RF predictions in orange for the exposed array at Site 8e at an hourly timestep with 10 cm vertical spacing (e.g., RF_Exposed_10cm) and (b) with 20 cm vertical spacing in blue (e.g., RF_Exposed_20cm). (c) RF predictions in yellow for Site 28 at 8 AM (e.g., RF_Exposed_Morning) and (d) RF predictions in purple for the same site at 8 PM (e.g., RF_Exposed_Evening).

absolute value of the Shapley values for each predictor variable in the training set. The temperature and standard deviation of temperature at or near the ground surface were important in all models, suggesting that the diurnal variation of temperature in the snowpack was useful for predicting snow depth, due to expected insulation within the snowpack. The same variables near the top of the vertical profiles held little importance in all models.

4 | Discussion

We used a novel approach of random forest regression combined with vertical profiles of temperature to estimate snow depth at multiple locations within a forest environment in the northeastern United States. Our work shows how vertical profiles combined with random forest regression can provide accurate snow depth estimates over multiple winters, establishing the effectiveness of combining machine learning methods with snow temperature data. Our RF models, with out-of-bag and testing RMSE ranging from 1.8 to 6.5 cm among our seven models, fell within the low end of the error range reported by previous studies.

Previous studies that used snow temperature to infer snow characteristics covered a wide range of study areas with varied site characteristics that were more heterogeneous than our watershed and used different techniques to estimate snow depth from temperature profiles. These other study regions include the upper Colorado River basin (with expected prediction errors between 5 and 10 cm; Dafflon et al. 2022), central Europe (Reusser and Zehe 2011), and daily estimates from the subarctic/Arctic Labrador in northeastern Canada (Tutton and Way 2021). Notably, the prediction errors for our most accurate models, RF_Exposed_20cm and RF_Exposed_10cm, were lower than any previous study of which we have knowledge. These two hourly RF models achieved out-of-bag and testing RMSEs between 2.2 cm for 20 cm sensor spacing and 1.8 for 10 cm spacing. The fact that the OOB and testing errors are the same suggests that the RF models are not overfitting and thus would generalise well to new data from the four sites included in the training dataset. These RMSEs are below the 6 cm minimum error that Reusser and Zehe (2011) calculated for a 15 cm sensor spacing, and at (or near) the limit of accuracy of the camera-derived validation data. These low errors suggest that RF models may be able to utilise information from snow/air temperature dynamics that are harder to represent in prescribed algorithms. Our other RF models with <1900 total timesteps achieved accuracy between 5.8 and 6.5 cm RMSE for OOB and testing errors. While this level of accuracy is poorer than what can be achieved using automated snow depth sensors (i.e., ultrasonic sensors that often cost $1000 or more), it comes at a much lower cost and with the added value of information about the energy state of the snowpack. Our study, while having a higher initial setup cost (i.e., iButtons range between $80 and $110 per unit) compared to a study deriving snow depths from timelapse cameras paired with snow stakes (i.e., timelapse cameras are between $50 and $200 per unit), provides more detail about a snowpack beyond snow

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

depth. Our snow temperature data detail how this snowpack might react to environmental changes in the future, allowing for improved accuracy in snowmelt forecasts. This interaction between the snow and air temperature and internal snow thermal dynamics is especially important in areas with frequent melt events like the northeastern U.S.

For this study, our target camera-derived snow depth data were limited to daily temporal resolution for the eight PVC arrays and any RF models that used PVC temperature data. More frequent camera imagery (e.g., at 4 or 6-h frequency) to match the 4-h temperature data or other nearby automated snow depth verification data would be useful to improve model prediction accuracy by increasing the size of the training set. For example, a supplementary analysis where we trained RF_Exposed_20cm using data with increased temporal frequency (i.e., once every hour instead of 4-hourly) increases prediction accuracy by as much as 40%. We also observed frequent shifts in snow depth from melting, densification, or snowfall increases in the camera images, sometimes over a few hours, so improved temporal resolution of snow depth data from daily to sub-daily (e.g., 6 h, or better) could improve its usability for hydrological, ecological, and snowpack modelling purposes.

Comparing the two temperature profile designs, we found that PVC shielding of iButtons temperature sensors had little influence on snow depth prediction accuracy, with RF_PVC and RF_Exposed having similar errors to the RF_Combined model. This was further supported by the minimal differences in temperature between setup types in Figure 3, suggesting that little to no predictive capability is lost due to thermal regulation by the PVC pipe. However, supplementary analyses that trained RF models only on data from the PVC arrays and then applied them to data from the exposed profiles (and vice versa) resulted in larger prediction errors. The larger prediction errors suggest that temperature dynamics within the PVC pipe versus sensors exposed directly to the snow/air surface were not completely identical. For sites that have high animal traffic, the PVC-enclosed profile is more robust to animal interference, which could potentially save on maintenance costs and limit data loss. For instance, animal interference from large mammals such as black bears and deer reduced the number of images we had available to extract snow depth for training and testing, limiting the overall extent of our verification data. Additionally, the PVC profiles were less influenced by sun exposure and thus showed less bias towards higher air temperatures on sunny days but can experience buildup of moisture within the tube during heavy rain periods and potential loss of data. However, our exposed profiles provide a more faithful representation of the actual snow temperature due to their direct physical contact with the snow. We recommend configurations like our exposed profiles for studies that aim to directly measure snow temperature.

Although our sites have similar land surface properties and meteorological forcing due to their location within the same watershed, we observed a range of snow depth across our sites. When any site with comparatively deep or shallow snow was removed from the RF training set, this had a greater influence on our predictions than removing a site with snow depth closer to the watershed average. This was especially prevalent in our LOO RMSEs and emphasises the need for the RF training set to incorporate the full range of conditions to effectively extrapolate snow depth predictions to a new location. We might also expect poorer accuracy if the sites had more drastic changes in elevation, steeper slopes, or less foliage. Additionally, using LOO analysis from any of our RF_Exposed models, the two sites with an open canopy (Sites 27 and 28) yielded nearly double the RMSE of the two closed canopy sites (Site 8e and Site 26) (see Table S5). Further analysis is required to determine if this finding is robust or a result specific to our site or training dataset. However, we note that most of the exposed models only included four sites, and a larger set of training sites with a broader set of environmental conditions might result in RF models that can be more easily applied to new sites. There were also indications that some of the iButton temperature data may have been impacted by temperature biases related to iButton batteries nearing their end of life, potentially resulting in high RMSEs for some sites (e.g., Site 18; see Table S5). Thus, our model errors may be conservative, at least for some sites, and might be slightly improved with more frequent and/or accurate temperature data or combining temperature data with light detection methods.

Using the exposed profiles to evaluate influence on time of day for RF model predictions, we found that our RF models performed similarly to each other when predicting snow depths only in the morning (i.e., 8:00 local time) versus only in the evening (i.e., 20:00 local time). This suggests the exposed temperature profiles perform similarly, regardless of solar radiation conditions, which might be expected to result in high temperature gradients. Using the exposed profiles to evaluate whether vertical resolution affected prediction accuracy, we found minimal improvement in predicted snow depth whether we used sensors at 10 cm or 20 cm vertical spacing, suggesting that a vertical resolution of 15–20 cm may be sufficient even in environments with shallow snowpacks and/or frequent melt events. We saw greater improvement in our RF models' prediction accuracy when using larger sample sizes for model training (i.e., 20,000+ timesteps vs. < 1900 timesteps). For our RF models, increasing the sample size decreased OOB and testing errors while the LOO errors only decreased minimally. Thus, it is possible to obtain very high snow depth accuracy using well-trained RF models that are applied to temperature data from existing sites (on which the models were trained). The RF models perform less well (7–10 cm RMSE, 5–7.5 cm MAE) when extrapolating to a 'new' site (i.e., a site with similar environmental characteristics but without any validation data). This level of accuracy may be sufficient for some applications, and errors can be minimised by including similar sites to a 'new' site in the model training dataset. Future studies might benefit from adopting 15 to 20 cm vertical spacing, reducing the cost of each individual temperature profile and thereby making this monitoring approach more feasible in areas with deeper snowpacks than the northeastern United States.

For our work, we found that the temperature and standard deviation of temperature near the top of a vertical profile ($\geq$ 0.6 m) held little importance in all models. This makes sense because these sensors were often outside the snowpack, and it was likely that the diurnal variation of temperature for temperature sensors exposed to the air was similar to each other and thus shared similar temperature dynamics. Variables other than temperature (e.g., site characteristics) showed little importance in all RF model results, aside from canopy gap fraction for some (but not

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

all) of the RF models (see Figures S2–S8). Supplementary analyses classified different predictor variables into three groups (temperature, standard deviation of temperature, or land surface characteristics) to determine influence on model prediction accuracy. Models with each permutation of the three groups that were omitted from a model run showed that including either of the temperature variable groups (i.e., the temperature or standard deviation of temperature variable groups) greatly improved model prediction accuracy, while the land surface characteristics contributed little to model prediction accuracy. Adding a second temperature group (e.g., adding the standard deviation of temperature group to a model that already included the temperature group) resulted in a slight increase in model prediction accuracy. This suggests that the temperature data contain sufficient information about snow depth to render the land surface variables unnecessary for accurate RF model prediction in watersheds with similar site characteristics. We expect that site characteristics might show higher importance in a more heterogeneous and/or more topographically variable environment.

5 | Conclusions

Our results showed that random forest models trained on vertical profiles of temperature sensors were accurate predictors of snow depth, even in shallow snowpacks that can experience rapid changes over short periods. This may be a cost-effective method for long-term monitoring of both snow energy state (i.e., snow temperature) and snow depth. Changes in temperature profile design (PVC-enclosed vs. exposed), vertical sensor spacing (10 vs. 20 cm), and time of day (morning vs. evening) had little influence on snow depth prediction accuracy from our RF models. The number of total observations in a training dataset and the range of snow conditions represented in the training dataset were important factors for RF model accuracy. Our results showed similar or better accuracy compared to other studies that estimated snow depth from temperature profiles using different methodologies, which supports the utility of machine learning methods for snow depth estimation from snow temperature data. Future work will apply our RF models to estimate snow depth across the entire Archer Creek watershed using the broader network of snow temperature profiles without timelapse camera imagery.

Acknowledgements

This research was supported by the NY State Energy Research & Development Authority (NYSERDA) Environmental Research Program, which fully underwrites long-term ecosystem monitoring programs at Huntington Wildlife Forest (SUNY ESF Newcomb Campus) via agreements #122681 (2018–2022) and #199214 (2023–2027). We would also like to acknowledge partial funding from a Syracuse University CUSE grant for some of the instrumentation used in this study and the Syracuse University Department of Earth & Environmental Sciences Prucha Fund for partially funding site visits to collect data. We would also like to thank fellow Syracuse University graduate student Heather Gunn for assistance with field data collection.

Data Availability Statement

Hydrological and meteorological data of Arbutus Lake are publicly available at the Adirondack Long-Term Monitoring at Huntington Forest website (https://adk-ltm.org/), managed by the Adirondack Ecological Center (AEC). The snow temperature to support this study's main findings and the R code used to create the random forest models to analyse our findings will be made available on Github.

References

Beier, C., J. Mills, P. McHale, C. T. Driscoll, and M. J. Mitchell. 2021. "Long-Term Ecosystem Monitoring at Huntington Forest: Integrating Hydrology, Biogeochemistry and Climatic Controls on Watershed Processes." *Hydrological Processes* 35, no. 8: e14328. https://doi.org/10.1002/hyp.14328.

Berghuijs, W. R., R. A. Woods, and M. Hrachowitz. 2014. "A Precipitation Shift From Snow Towards Rain Leads to a Decrease in Streamflow." *Nature Climate Change* 4, no. 7: 583–586. https://doi.org/10.1038/nclimate2246.

Bintanja, R., and O. Andry. 2017. "Towards a Rain-Dominated Arctic." *Nature Climate Change* 7, no. 4: 263–267. https://doi.org/10.1038/nclimate3240.

Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp. 2014. "The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy." *Bulletin of the American Meteorological Society* 95: 1431–1443. https://doi.org/10.1175/BAMS-D-13-00047.1.

Breen, C. M., W. R. Currier, C. Vuyovich, Z. Miao, and L. R. Prugh. 2024. "Snow Depth Extraction From Time-Lapse Imagery Using a Keypoint Deep Learning Model." *Water Resources Research* 60, no. 7: e2023WR036682. https://doi.org/10.1029/2023WR036682.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45, no. 1: 5–32. https://doi.org/10.1023/A:1010933404324.

Burakowski, E. A., C. P. Wake, B. Braswell, and D. P. Brown. 2008. "Trends in Wintertime Climate in the Northeastern United States: 1965–2005." *Journal of Geophysical Research: Atmospheres* 113: D20114. https://doi.org/10.1029/2008JD009870.

Comerford, D. P., P. G. Schaberg, P. H. Templer, A. M. Socci, J. L. Campbell, and K. F. Wallin. 2013. "Influence of Experimental Snow Removal on Root and Canopy Physiology of Sugar Maple Trees in a Northern Hardwood Forest." *Oecologia* 171, no. 1: 261–269. https://doi.org/10.1007/s00442-012-2393-x.

Contosta, A. R., N. J. Casson, S. Garlick, et al. 2019. "Northern Forest Winters Have Lost Cold, Snowy Conditions That Are Important for Ecosystems and Human Communities." *Ecological Applications* 29, no. 7: e01974. https://doi.org/10.1002/eap.1974.

Dafflon, B., S. Wielandt, J. Lamb, et al. 2022. "A Distributed Temperature Profiling System for Vertically and Laterally Dense Acquisition of Soil and Snow Temperature." *Cryosphere* 16, no. 2: 719–736. https://doi.org/10.5194/tc-16-719-2022.

Diaz, H. F., J. K. Eischeid, C. Duncan, and R. S. Bradley. 2003. "Variability of Freezing Levels, Melting Season Indicators, and Snow Cover for Selected High-Elevation and Continental Regions in the Last 50 Years." In *Climate Variability and Change in High Elevation Regions: Past, Present & Future*, edited by H. F. Diaz, 33–52. Springer Netherlands. https://doi.org/10.1007/978-94-015-1252-7_3.

Dickerson-Lange, S. E., R. F. Gersonde, J. A. Hubbart, et al. 2017. "Snow Disappearance Timing Is Dominated by Forest Effects on Snow Accumulation in Warm Winter Climates of the Pacific Northwest, United States." *Hydrological Processes* 31, no. 10: 1846–1862. https://doi.org/10.1002/hyp.11144.

Dudley, R. W., G. A. Hodgkins, M. R. McHale, M. J. Kolian, and B. Renard. 2017. "Trends in Snowmelt-Related Streamflow Timing in the Conterminous United States." *Journal of Hydrology* 547: 208–221. https://doi.org/10.1016/j.jhydrol.2017.01.051.

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gleason, K. E., J. B. Bradford, A. W. D'Amato, S. Fraver, B. J. Palik, and M. A. Battaglia. 2021. "Forest Density Intensifies the Importance of Snowpack to Growth in Water-Limited Pine Forests." *Ecological Applications* 31, no. 1: e02211. https://doi.org/10.1002/eap.2211.

Golding, D. L., and R. H. Swanson. 1986. "Snow Distribution Patterns in Clearings and Adjacent Forest." *Water Resources Research* 22, no. 13: 1931–1940. https://doi.org/10.1029/WR022i013p01931.

Gomez, J., P. Vidon, J. Gross, C. Beier, J. Caputo, and M. Mitchell. 2016. "Estimating Greenhouse Gas Emissions at the Soil–Atmosphere Interface in Forested Watersheds of the US Northeast." *Environmental Monitoring and Assessment* 188, no. 5: 295. https://doi.org/10.1007/s10661-016-5297-0.

Greenwell, B. 2024. "fastshap: Fast Approximate Shapley Values. R Package Version 0.1.1." https://bgreenwell.github.io/fastshap/. https://github.com/bgreenwell/fastshap.

Hardy, J. P., R. Melloh, G. Koenig, et al. 2004. "Solar Radiation Transmission Through Conifer Canopies." *Agricultural and Forest Meteorology* 126, no. 3: 257–270. https://doi.org/10.1016/j.agrformet.2004.06.012.

Hijmans, R. J., J. Van Etten, J. Cheng, et al. 2015. "Package 'Raster'." *R Package* 734: 473. http://www.R-project.org.

Hu, J. M., and A. W. Nolin. 2020. "Widespread Warming Trends in Storm Temperatures and Snowpack Fate Across the Western United States." *Environmental Research Letters* 15, no. 3: 034059. https://doi.org/10.1088/1748-9326/ab763f.

Jones, J. A., and R. M. Perkins. 2010. "Extreme Flood Sensitivity to Snow and Forest Harvest, Western Cascades, Oregon, United States." *Water Resources Research* 46: W12512. https://doi.org/10.1029/2009WR008632.

Jullum, M., L. H. B. Olsen, J. Lachmann, and A. Redelmeier. 2024. "shapr: Explaining Machine Learning Models With Conditional Shapley Values in R and Python." https://norskregnesentral.github.io/shapr/.

Kapnick, S., and A. Hall. 2012. "Causes of Recent Changes in Western North American Snowpack." *Climate Dynamics* 38, no. 9: 1885–1899. https://doi.org/10.1007/s00382-011-1089-y.

Lawrence, D. M., and A. G. Slater. 2010. "The Contribution of Snow Condition Trends to Future Ground Climate." *Climate Dynamics* 34, no. 7: 969–981. https://doi.org/10.1007/s00382-009-0537-4.

Lewkowicz, A. G. 2008. "Evaluation of Miniature Temperature-Loggers to Monitor Snowpack Evolution at Mountain Permafrost Sites, Northwestern Canada." *Permafrost and Periglacial Processes* 19: 323–331. https://doi.org/10.1002/ppp.625.

Liu, Y., and A. Just. 2020. "SHAPforxgboost: SHAP Plots for 'XGBoost'. R Package Version 0.1.3." https://github.com/liuyanguu/SHAPforxgboost/.

Lundquist, J. D., S. E. Dickerson-Lange, J. A. Lutz, and N. C. Cristea. 2013. "Lower Forest Density Enhances Snow Retention in Regions With Warmer Winters: A Global Framework Developed From Plot-Scale Observations and Modeling." *Water Resources Research* 49, no. 10: 6356–6370. https://doi.org/10.1002/wrcr.20504.

Lundquist, J. D., and F. Lott. 2008. "Using Inexpensive Temperature Sensors to Monitor the Duration and Heterogeneity of Snow-Covered Areas." *Water Resources Research* 44: 4. https://doi.org/10.1029/2008WR007035.

Musselman, K. N., N. Addor, J. A. Vano, and N. P. Molotch. 2021. "Winter Melt Trends Portend Widespread Declines in Snow Water Resources." *Nature Climate Change* 11, no. 5: 418–424. https://doi.org/10.1038/s41558-021-01014-9.

Raleigh, M. S., C. C. Landry, M. Hayashi, W. L. Quinton, and J. D. Lundquist. 2013. "Approximating Snow Surface Temperature From Standard Temperature and Humidity Data: New Possibilities for Snow Model and Remote Sensing Evaluation." *Water Resources Research* 49, no. 12: 8053–8069. https://doi.org/10.1002/2013WR013958.

Reinmann, A. B., J. R. Susser, E. M. C. Demaria, and P. H. Templer. 2019. "Declines in Northern Forest Tree Growth Following Snowpack Decline and Soil Freezing." *Global Change Biology* 25, no. 2: 420–430. https://doi.org/10.1111/gcb.14420.

Reusser, D. E., and E. Zehe. 2011. "Low-Cost Monitoring of Snow Height and Thermal Properties With Inexpensive Temperature Sensors." *Hydrological Processes* 25, no. 12: 1841–1852. https://doi.org/10.1002/hyp.7937.

Rodriguez-Galiano, V., M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas. 2015. "Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines." *Ore Geology Reviews* 71: 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001.

Ryan, W. A., N. J. Doesken, and S. R. Fassnacht. 2008. "Evaluation of Ultrasonic Snow Depth Sensors for U.S. Snow Measurements." *Journal of Atmospheric and Oceanic Technology* 25: 667–684. https://doi.org/10.1175/2007JTECHA947.1.

Serreze, M. C., M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty. 1999. "Characteristics of the Western United States Snowpack From Snowpack Telemetry (SNO^{TEL}) Data." *Water Resources Research* 35, no. 7: 2145–2160. https://doi.org/10.1029/1999WR900090.

Siirila-Woodburn, E. R., A. M. Rhoades, B. J. Hatchett, et al. 2021. "A Low-to-No Snow Future and Its Impacts on Water Resources in the Western United States." *Nature Reviews Earth and Environment* 2, no. 11: 800–819. https://doi.org/10.1038/s43017-021-00219-y.

Storck, P., D. P. Lettenmaier, and S. M. Bolton. 2002. "Measurement of Snow Interception and Canopy Effects on Snow Accumulation and Melt in a Mountainous Maritime Climate, Oregon, United States." *Water Resources Research* 38, no. 11: 5-1-5–5-1-16. https://doi.org/10.1029/2002WR001281.

Strumbelj, E., and I. Kononenko. 2010. "An Efficient Explanation of Individual Classifications Using Game Theory." *Journal of Machine Learning Research* 11: 1–18.

Sturm, M., and G. E. Liston. 2021. "Revisiting the Global Seasonal Snow Classification: An Updated Dataset for Earth System Applications." *Journal of Hydrometeorology* 22, no. 11: 2917–2938. https://doi.org/10.1175/JHM-D-21-0070.1.

Tutton, R. J., and R. G. Way. 2021. "A Low-Cost Method for Monitoring Snow Characteristics at Remote Field Sites." *Cryosphere* 15, no. 1: 1–15. https://doi.org/10.5194/tc-15-1-2021.

U.S. Geological Survey. 1995. "Index of Digital Elevation Models (DEM), New York." https://cugir.library.cornell.edu/catalog/cugir-008186?id=23.

Varhola, A., N. C. Coops, M. Weiler, and R. D. Moore. 2010. "Forest Canopy Effects on Snow Accumulation and Ablation: An Integrative Review of Empirical Results." *Journal of Hydrology* 392, no. 3: 219–233. https://doi.org/10.1016/j.jhydrol.2010.08.009.

Wieder, W. R., D. Kennedy, F. Lehner, et al. 2022. "Pervasive Alterations to Snow-Dominated Ecosystem Functions Under Climate Change." *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 30: e2202393119. https://doi.org/10.1073/pnas.2202393119.

Wilson, G., M. Green, J. Campbell, A. Contosta, N. Lany, and A. Bailey. 2024. "Long-Term Measurements of Seasonal Snowpacks Indicate Increases in Mid-Winter Snowmelt and Earlier Snowpack Disappearance in the Northeastern U.S." *PLOS Climate* 3, no. 12: e0000529. https://doi.org/10.1371/journal.pclm.0000529.

Wright, M. N., and A. Ziegler. 2017. "Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software* 77: 1–17. https://doi.org/10.18637/jss.v077.i01.

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Yang, J., L. Jiang, K. Luojus, et al. 2020. "Snow Depth Estimation and Historical Data Reconstruction Over China Based on a Random Forest Machine Learning Approach." *Cryosphere* 14, no. 6: 1763–1778. https://doi.org/10.5194/tc-14-1763-2020.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Summary information for PVC and exposed iButton temperature sites in the Archer Creek Watershed in Newcomb, NY, USA. Sites 1–24 are the original paired PVC enclosed temperature arrays as part of a large snow monitoring network, only eight of these had accompanying cameras for daily snow depth verification which were used for this study. Sites 8e and 26–28 are the exposed temperature arrays installed in mid-November 2022. **Table S2:** Summary of six PVC sites of the original eight that had missing camera imagery for a small gap from November 2021 to May 2023. The sites numbers, missing sensor heights, and dates of when there is no available camera data are listed below. **Table S3:** Random forest regression models with number of timesteps and different variance/error statistics. The specific error statistics shown below are the out-of-bag (OOB) error, testing error, and leave-one-out (LOO) error (which are analogous to applying a fitted RF model to a new site). Error statistics are reported in cm. **Table S4:** Random forest regression model names with number of timesteps and the number of timesteps from each site represented in a specific model. Temperature profile sites that are not included in a specific model are labelled with a dash. **Table S5:** Site-specific leave-one-out (LOO) RMSEs (which are analogous to applying a fitted RF model to a new site) organised by site and random forest regression model. Temperature profile sites that are not included in a given RF model are labelled with a dash. All RMSE values are reported in units of cm. **Figure S1:** Scatterplots of PVC-enclosed temperature data against exposed temperature data at Site 8e for all matching heights of (a) −0.2 m, (b) 0 m, (c) 0.2 m, (d) 0.4 m, (e) 0.6 m, and (f) 1 m. Each plot is colourized by time of year from November (green) to May (purple). Pearson's correlations for each comparison are added to the bottom right corner of each plot. **Figure S2:** Beeswarm plot where each point represents a Shapley value (phi) for a given observation and variable from the RF_Combined model. The overall spread of values for a variable shows the relative importance it has on RF model predictions. A negative phi value means that the variable lowered the snow depth prediction from the RF model while a positive phi value means it increased the snow depth prediction. Variables labelled with 'T', stand for temperature in°C, 'sdT' refers to the 24-h standard deviation of temperature at that height, and the −0.2 m sensor height (i.e., 0.2 m below ground level) is represented by '_.0.2'. **Figure S3:** Beeswarm plot where each point represents a Shapley value (phi) for a given observation and variable from the RF_PVC model in Figure 5a. The overall spread of values for a variable shows the relative importance it has on RF model predictions. A negative phi value means that the variable lowered the snow depth prediction from the RF model while a positive phi value means it increased the snow depth prediction. Variables labelled with 'T', stand for temperature in°C, 'sdT' refers to the 24-h standard deviation of temperature at that height, and the −0.2 m sensor height (i.e., 0.2 m below ground level) is represented by '_.0.2'. **Figure S4:** Beeswarm plot where each point represents a Shapley value (phi) for a given observation and variable from the RF_Exposed model in Figure 5b. The overall spread of values for a variable shows the relative importance it has on RF model predictions. A negative phi value means that the variable lowered the snow depth prediction from the RF model while a positive phi value means it increased the snow depth prediction. Variables labelled with 'T', stand for temperature in°C, 'sdT' refers to the 24-h standard deviation of temperature at that height, and the −0.2 m sensor height (i.e., 0.2 m below ground level) is represented by '_.0.2'. **Figure S5:** Beeswarm plot where each point represents a Shapley value (phi) for a given observation and variable from the RF_Exposed_10cm model in Figure 6a. The overall spread of values for a variable shows the relative importance it has on RF model predictions. A negative phi value means that the variable lowered the snow depth prediction from the RF model while a positive phi value means it increased the snow depth prediction. Variables labelled with 'T', stand for temperature in°C, 'sdT' refers to the 24-h standard deviation of temperature at that height, and the −0.2 m sensor height (i.e., 0.2 m below ground level) is represented by '_.0.2'. **Figure S6:** Beeswarm plot where each point represents a Shapley value (phi) for a given observation and variable from the RF_Exposed_20cm model in Figure 6b. The overall spread of values for a variable shows the relative importance it has on RF model predictions. A negative phi value means that the variable lowered the snow depth prediction from the RF model while a positive phi value means it increased the snow depth prediction. Variables labelled with 'T', stand for temperature in°C, 'sdT' refers to the 24-h standard deviation of temperature at that height, and the −0.2 m sensor height (i.e., 0.2 m below ground level) is represented by '_.0.2'. **Figure S7:** Beeswarm plot where each point represents a Shapley value (phi) for a given observation and variable from the RF_Exposed_Morning model in Figure 6c. The overall spread of values for a variable shows the relative importance it has on RF model predictions. A negative phi value means that the variable lowered the snow depth prediction from the RF model while a positive phi value means it increased the snow depth prediction. Variables labelled with 'T', stand for temperature in°C, 'sdT' refers to the 24-h standard deviation of temperature at that height, and the −0.2 m sensor height (i.e., 0.2 m below ground level) is represented by '_.0.2'. **Figure S8:** Beeswarm plot where each point represents a Shapley value (phi) for a given observation and variable from the RF_Exposed_Evening model in Figure 6d. The overall spread of values for a variable shows the relative importance it has on RF model predictions. A negative phi value means that the variable lowered the snow depth prediction from the RF model while a positive phi value means it increased the snow depth prediction. Variables labelled with 'T', stand for temperature in°C, 'sdT' refers to the 24-h standard deviation of temperature at that height, and the −0.2 m sensor height (i.e., 0.2 m below ground level) is represented by '_.0.2'. **Figure S9:** Median absolute deviation (MAD) between modelled and observed snow depth for the RF_Exposed_10cm model (i.e., the model with the most training data). The black points are the MAD calculated from the testing set, and the red points are the MAD calculated from the training set, each binned in 1-degree increments of snowpack temperature. The size of the points indicates the relative amount of data that falls within a given 1-degree bin.

10991085, 2025, 9, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/hyp.70273, Wiley Online Library on [24/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License