

ENHANCING GLOBAL OBSERVATION OF SOIL FREEZE/THAW STATE  
THROUGH MULTI-SCALE DATA FUSION

by

Jeremy Johnston  
A Dissertation  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
in Partial Fulfillment of  
The Requirements for the Degree  
of  
Doctor of Philosophy  
Civil, Environmental, and Infrastructure Engineering

Committee:

<u><i>Viviana Maggioni</i></u>	Dr. Viviana Maggioni, Dissertation Director
<u><i>Paul R. Houser</i></u>	Dr. Paul Houser, Co-Dissertation Director
<u><i>Celso Moller Ferreira</i></u>	Dr. Celso Ferreira, Committee Member
<u><i>Kirin Furst</i></u>	Dr. Kirin Furst, Committee Member
<u><i>Sam Salem</i></u>	Dr. Sam Salem, Department Chair

Date: April 21, 2022

Spring Semester 2022  
George Mason University  
Fairfax, VA

Enhancing Global Observation of Soil Freeze/Thaw State Through Multi-Scale Data  
Fusion

A Dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy at George Mason University

by

Jeremy M. Johnston  
Master of Science  
George Mason University, 2019  
Bachelor of Science  
George Mason University, 2017

Director: Viviana Maggioni, Associate Professor  
Sid and Reva Dewberry Department of Civil, Environmental, and Infrastructure  
Engineering

Spring Semester 2022  
George Mason University  
Fairfax, VA

Copyright 2022 Jeremy M. Johnston  
All Rights Reserved

## **DEDICATION**

I dedicate this to all those who pushed, helped, and put up with me throughout this 5-year process. Especially my mother.

I also dedicate this to the pursuit of new and exciting science. To discoverers, teachers, and explorers tackling new challenges. I share this accomplishment with all of you.

## ACKNOWLEDGEMENTS

Just a few years ago I would have never considered the possibility that I could one day call myself a scientist. Through the direct support and guidance of many, and through taking the time to listen and learn from others, I have come a long way towards achieving this goal. Most importantly, without the help of my advisors Dr. Viviana Maggioni and Dr. Paul Houser, I certainly would not have had the opportunity to explore research questions interesting to me. Their efforts to steer me down this path took me to places I hadn't dreamed of going, and I am forever grateful. For their guidance through the world of research and for pushing me to pursue new opportunities, I thank you both. I know I still have a long way to go.

I want to also thank all of my colleagues and fellow graduate students at George Mason and within Dr. Maggioni's Research Group. Collaborating and learning from you over the years has been a pleasure. Namely, Azbina Rahman, Tasnuva Rouf, Yiwen Mei, Ishrat Dollan, Yuan Xue, Soelem Bhuiyan, Gustavo Coelho, Maryam Zavareh, Andy Sachs, Felicio Cassalho, Tyler Meisse, and many others. I thank you all and hope to continue these collaborations into the future.

Moreover, I'd to thank my committee member Dr. Celso Ferreira, for being a wonderful teacher, supporter, and collaborator during my years at George Mason. And, to Dr. Kirin Furst, who graciously accepted my request to be a member of my committee and has always asked me challenging questions.

To all of those who helped me dive headfirst into field work, measuring snow, soil moisture, capturing imagery from new perspectives, as well as the many amazing scientists I have been able to learn from outside of the office, I thank you.

Critically, without the support of the NASA Terrestrial Hydrology Program (Grant #: NNX17AL38G), none of this work would be possible.

Last, but maybe most importantly, I want to thank my family and friends. Especially my parents, for supporting me at research events, expressing interest, and even poking fun at the number of times I have said the words 'freeze-thaw' over the last few years. I am beyond grateful to you all.

## TABLE OF CONTENTS

	Page
List of Tables .....	vii
List of Figures .....	viii
List of Equations .....	x
List of Abbreviations and Symbols.....	xi
Abstract.....	xii
1 Background and Project Motivation.....	1
1.1 Significance of Freeze/Thaw Processes .....	1
1.2 Current Observational Approaches .....	4
1.2.1 Physics of Microwave Remote Sensing and Freeze/Thaw Transitions.....	7
1.2.2 An Introduction to Existing Sensors and Available Freeze/Thaw Products ....	14
1.2.3 Known Limitations .....	21
1.3 Project Overview.....	22
1.3.1 Research Objectives .....	23
1.3.2 Project Tasks.....	23
2 Improving Characterization of Existing Freeze/Thaw Observation Products.....	25
2.1 Examining Near Surface Temperature Variables and their Relationship to the Microwave Derived Freeze/Thaw Observations.....	26
2.1.1 Background and Objectives.....	27
2.1.2 Comparison Methods.....	32
2.1.3 Results and Discussion .....	39
2.2 A Global Comparison of Satellite Freeze/Thaw Records .....	64
2.2.1 The Need for Global Scale Comparisons of Freeze/Thaw Records.....	64
2.2.2 Methods for Global Comparison .....	68
2.2.3 Global Comparison Results .....	78
2.3 Chapter Outcomes .....	96
3 Exploring Variables for Freeze/Thaw Estimation at Fine Scales.....	100
3.1 Improving Freeze/Thaw State Classification Using Sub-grid Temperature .....	101
3.1.1 Background and Motivations.....	101
3.1.2 Research Methods.....	104

3.1.3 Exploring Sub-grid Temperature and Freeze/Thaw .....	116
3.2 Exploring Field-scale Thermal, Soil Freezing, and Melt Processes .....	134
3.2.1 Study Methods.....	136
3.2.2 Results, Analysis, and Project Outcomes .....	142
3.3 Chapter Outcomes .....	154
4 GMU-Freeze/Thaw: Remote Sensing Data Fusion for Improved Freeze/Thaw Estimates .....	157
4.1 Motivations and Approaches for the Enhancement of Freeze/Thaw Observation	158
4.2 Deriving an Improved Classifier .....	162
4.2.1 Data Inputs.....	163
4.2.2 Decision Trees and Random Forests for Freeze/Thaw Classification.....	169
4.2.3 Feature Creation and Model Development.....	177
4.3 Product Performance and Results .....	181
4.3.1 Freeze/Thaw Modeling at Point Locations.....	181
4.3.2 Global Algorithm Training and Development.....	190
4.3.3 Global Application and Validation.....	196
4.4 Chapter Outcomes .....	205
5 Research Contributions, Future Opportunities, and Dissemination .....	207
Appendix.....	210
References.....	215

## LIST OF TABLES

Table	Page
Table 1 Selected satellite derived freeze/thaw products .....	29
Table 2 Selected temperature products .....	32
Table 3 Classification accuracy results for existing products in North America .....	40
Table 4 Proportional differencing summary values .....	47
Table 5 Freeze/thaw temperature distribution summary characteristics .....	59
Table 6 Modified Köppen-Geiger climate classifications .....	74
Table 7 Modified IGBP land cover classifications from MODIS MCD12Q1 .....	75
Table 8 SMAP-FT to SSM/I-FT agreement statistics .....	86
Table 9 Product agreement by freezing days .....	89
Table 10 SMAP-FT to SSM/I-FT agreement by landcover .....	92
Table 11 Product agreement proportion by elevation standard deviation .....	95
Table 12 Summary of ensemble validation results .....	107
Table 13 Observational and satellite derived datasets .....	110
Table 14 Characteristics within analysis targeted freeze/thaw pixels .....	113
Table 15 Test model point site descriptions .....	183
Table 16 Site specific performance results for freeze/thaw classification algorithms....	189
Table 17 Model training summary and performance scores by climate region .....	196
Table 18 GMU-FT summary and validation statistics for 2020 period .....	202

## LIST OF FIGURES

Figure	Page
Figure 1 The relevance of freeze/thaw cycles.....	3
Figure 2 Techniques for freeze/thaw observation.....	6
Figure 3 Real and imaginary permittivity of soil.....	9
Figure 4 SMAP-FT summer freeze classifications map.....	38
Figure 5 Classification accuracy of freeze/thaw products to air temperature.....	43
Figure 6 Classification accuracy of freeze/thaw products to soil temperature.....	44
Figure 7 Classification accuracy of freeze/thaw products to NLDAS skin temperature..	45
Figure 8 Classification accuracy of freeze/thaw products to MODIS skin temperature ..	46
Figure 9 Mean proportional difference of freeze/thaw products to air temperature.....	48
Figure 10 Mean proportional difference of freeze/thaw products to soil temperature .....	50
Figure 11 Mean proportional difference of freeze/thaw products to NLDAS skin temperature .....	51
Figure 12 Mean proportional difference of freeze/thaw products to MODIS skin temperature .....	52
Figure 13 Mean frozen proportion differences timeseries.....	54
Figure 14 Absolute mean proportional differences by overpass .....	55
Figure 15 Boxplot temperature distributions of frozen and thawed states .....	57
Figure 16 Probability distribution function histogram AMSR-FT to air temperature.....	60
Figure 17 Distribution of global Agreement Proportion SMAP to SSM/I-FT .....	71
Figure 18 Timeseries detrending and seasonality removal.....	72
Figure 19 Spatial comparison variable maps .....	77
Figure 20 Total frozen classifications by hemisphere .....	78
Figure 21 Agreement proportion between global P-MW freeze/thaw products.....	79
Figure 22 Comparing global records .....	81
Figure 23 Time series FT agreement proportion separated by climate class.....	84
Figure 24 Density plot of freezing days against agreement proportion.....	88
Figure 25 Time series of product FT agreement separated by land cover.....	90
Figure 26 Density plot of product agreement compared to elevation variability .....	94
Figure 27 Focus pixels, representative regions, and in-situ site locations.....	108
Figure 28 North American climate, land cover, and elevation profile maps.....	115
Figure 29 Spatial mean sub-grid frozen proportion when classified as frozen.....	117
Figure 30 Sub-grid frozen proportion means when classified frozen by class.....	118
Figure 31 Percentage of agreement on frozen classifications.....	122
Figure 32 Freeze/thaw state to temperature comparison for Tundra pixel .....	124
Figure 33 Freeze/thaw state to temperature comparison for Southern Plains pixel .....	125
Figure 34 Freeze/thaw state to temperature comparison for Mountainous pixel.....	126
Figure 35 Freeze/thaw state to temperature comparison for Northern Forest pixel .....	127
Figure 36 Freeze/thaw state to temperature comparison for Northern Plains pixel .....	128
Figure 37 Summary of results for each focus region.....	128

Figure 38 Grand Mesa, CO study area.....	137
Figure 39 UAV thermal and visible image pairs captured in varied conditions.....	144
Figure 40 Timeseries of in-situ observations, FT classifications, and UAV imagery ....	145
Figure 41 Surface temperature imagery from UAV collections.....	149
Figure 42 Timeseries of near site observations and energy balance.....	151
Figure 43 R <sup>2</sup> improvement by adding additional model predictors.....	152
Figure 44 Distribution of global network of in-situ soil temperature sites.....	167
Figure 45 Model accuracy and training time of machine learning methods.....	171
Figure 46 Soil temperature observational sites map.....	182
Figure 47 Decision tree modeling for Alaska, USA site .....	184
Figure 48 Decision tree modeling for High Mountain Asia (HMA) site.....	185
Figure 49 Decision tree modeling for Northern Italy site.....	186
Figure 50 Decision tree modeling for Illinois, USA site .....	187
Figure 51 Decision tree modeling for Texas, USA site .....	188
Figure 52 Sensitivity analysis for select random forest parameters .....	192
Figure 53 Normalized feature importance by climate class.....	194
Figure 54 Framework for global freeze/thaw classification using random forests.....	197
Figure 55 Example outputs from GMU-FT global model.....	198
Figure 56 Global validation of GMU-FT against operational FT products.....	200
Figure 57 Comparison of total number of frozen days between FT classifiers .....	201
Figure 58 Classification accuracy of soil freeze/thaw state by temperature category ....	203

## LIST OF EQUATIONS

Equation	Page
Equation 1 Planck's law of blackbody radiation .....	7
Equation 2 Simplified formulation for brightness temperature .....	8
Equation 3 Normalized polarization ratio formula for SMAP bands .....	16
Equation 4 Seasonal scale factor used in SMAP-FT algorithm.....	16
Equation 5 Thresholding for freeze/thaw classification NPR algorithm .....	16
Equation 6 Freeze thaw single channel V-pol (SCV) algorithm .....	17
Equation 7 Classification Accuracy.....	35
Equation 8 Classification Accuracy.....	36
Equation 9 Agreement proportion .....	70
Equation 10 Frozen proportion .....	111
Equation 11 Sub-pixel frozen proportion when classified as frozen (FZp).....	111
Equation 12 Sub-pixel frozen proportion when classified as thawed (THWp).....	111
Equation 13 Three-product agreement proportion.....	112
Equation 14 Precision .....	180
Equation 15 Recall.....	180
Equation 16 F1-Score .....	181

## LIST OF ABBREVIATIONS AND SYMBOLS

Freeze/thaw state.....	FT
Freeze/thaw State Product.....	FTSP
Earth Observing System .....	EOS
Passive Microwave .....	P-MW
Soil Moisture Active/Passive.....	SMAP
Special Sensor Microwave Imager .....	SSMI or SSM/I
Special Sensor Microwave Imager/Sounder.....	SSMIS
Advanced Scanning Microwave Radiometer (-EOS, -Version 2).....	AMSR (-E, -2)
Brightness Temperature .....	$T_b$
Unpiloted Aerial Vehicle .....	UAV
Scanning L-band Active/Passive Experiment.....	SLAPEX
Single Channel Algorithm .....	SCA
Normalized Polarization Ratio.....	NPR
Spectral Gradient .....	SG
Machine Learning .....	ML
Decision Tree .....	DT
Random Forest.....	RF
Making Earth System Data Records for Use in Research Environments.....	MEaSUREs
Geostationary Operational Environmental Satellite .....	GOES
The Continental United States .....	CONUS
World Meteorological Organization .....	WMO
Thermal Infrared .....	TIR
Agreement proportion.....	$A_p$
Classification Accuracy .....	CA
Celsius.....	C
Kelvin.....	K
Degree.....	$^{\circ}$
Correlation .....	R
Root Mean Square Error .....	RMSE
Mean Absolute Error.....	MAE
Gini's Diversity Index.....	GDI
radiometer version of Scatterometer Image Reconstruction.....	rSIR
National Snow and Ice Data Center.....	NSIDC
Volumetric Water Content.....	VWC

## **ABSTRACT**

### **ENHANCING GLOBAL OBSERVATION OF SOIL FREEZE/THAW STATE THROUGH MULTI-SCALE DATA FUSION**

Jeremy M. Johnston, Ph.D.

George Mason University, 2022

Dissertation Director: Dr. Viviana Maggioni

In cold regions, temperature acts as a hydrologic and ecological switch. As extended periods of sub-freezing temperatures can constrain various parts of the water cycle by preventing evaporation, infiltration, and runoff as well as defining the transition from periods of plant growth and ecologic activity to ecosystem dormancy. Hence, global knowledge of surface characteristics such as freeze/thaw state are exceedingly important for monitoring of biogeochemical cycles, such as the carbon cycle, and the implications of a changing climate. Current approaches leverage satellite observations in the microwave portion of the electro-magnetic spectrum to detect the transition between frozen and thawed states. However, investigations into the use of such observations have suggested freeze/thaw classifications remain uncertain and lack a clear physical meaning as to the type of freeze indicated (e.g., soil, surface snow and ice, frost). This project focuses on improving the characterization of freeze/thaw processes, in terms of both the observational techniques and the methods used to extract information from observations at multiple

scales. To achieve this, we employ extensive analyses of existing satellite derived freeze/thaw classifications, the introduction of new predictive variables (i.e., land surface temperature and land cover), targeted field campaigns, and the integration of thousands of global soil temperature observing sites. The result, a new and robust decision tree-based classification algorithm which introduces a freeze/thaw probability metric, soil specific classifications, and improved resolution. This information is expected to provide better inputs to hydrologic models used to predict runoff and flooding, allow for more accurate estimates of the growing season, and enable assessments of the agricultural implications of changing melt and refreeze cycles. At its core, the implementation of an improved freeze/thaw classification can enhance our understanding of fundamental physical processes which occur on our planet.

# **1 BACKGROUND AND PROJECT MOTIVATION**

This dissertation details the path to improving observations of water within the Earth system. With the growth of technological capabilities, especially in remote sensing, a breadth of opportunity has been unlocked in Earth observation. Specifically, capturing detailed changes in the cryosphere and in regions experiencing seasonal melt and refreeze cycles. This chapter provides a baseline of understanding for water freezing and thawing processes at the Earth's surface and the associated importance. This includes detailing the relevance of these cycles to our lives as well as the many exchanges occurring between the land, environment, and atmosphere. We introduce the physics of water state changes, different observational techniques used to track these essential processes, and briefly explore limitations and opportunities that exist in the context of improving these techniques. This highlights the primary objective underpinning this work, which is to develop the next generation of algorithms to enhance observation of freeze/thaw cycling at a global scale.

## **1.1 Significance of Freeze/Thaw Processes**

Seasonal freezing and thawing cycles at the Earth's surface act as a critical control of global climate and the hydrologic cycle. Water not only facilitates many important processes within the human body, such as nutrient transport to cells and the regulation of body temperature, but provides a similar role at a global scale, by moving

energy (e.g., thermal, latent fluxes), sediment, and regulating weather and climate. As a result, both ecological activity and the prevalence of life as we know it are constrained by the movement and state of water. Extensive research has been devoted to characterizing the effects of freeze/thaw (FT) transitions on carbon, water, and energy cycles, further revealing the link between water phase changes and these global systems. For instance, changes in the frozen season timing and duration, especially at high elevations and latitudes, have significant implications for the length of the growing season, plant productivity, and the associated carbon uptake (*Goulden et al. 1996; Kimball et al. 2004; Kreyling et al. 2008*).

Moreover, FT cycles play a substantial role in regulating energy exchanges between the Earth's surface and atmosphere through impacts to evapotranspiration, sensible and latent heat flux, runoff, and albedo (*Shanley and Chalmers 1999; Zhang et al. 2011; Zhang and Sun 2011; Betts et al. 2014*). As a result, surface FT state can directly influence temperature regimes both at micro- and macro-scales, which is relevant to hydrologic, weather, and climate modeling (*Stadler et al. 1997; Henry 2008; Koren et al. 2014; Xue et al. 2019*). Essentially, surface freezes can act as a cap on the exchange of water between the soil and atmosphere. Frozen ground can prevent absorption of rainwater or snowmelt leading to increased runoff while also inhibiting evaporation by locking water in ice within the snowpack or soil. Similarly, global FT processes can constrain groundwater recharge, soil moisture, and infiltration rates, which can directly impact flood intensity through increases in runoff from snowmelt and rain on frozen ground events (*Dunne and Black 1971; Koren et al. 1999; Bayard et al. 2005; Niu and*

Yang 2006; Iwata et al. 2011). Freeze/thaw state and cycles are also linked to socio-economic impacts through their connections to flooding, water availability, ground instability, and weathering of infrastructure such as roadways (Kalra et al. 2008; Martins et al. 2016; Daniel et al. 2017; Teufel et al. 2017; Hjort et al. 2018; Chen et al. 2021).

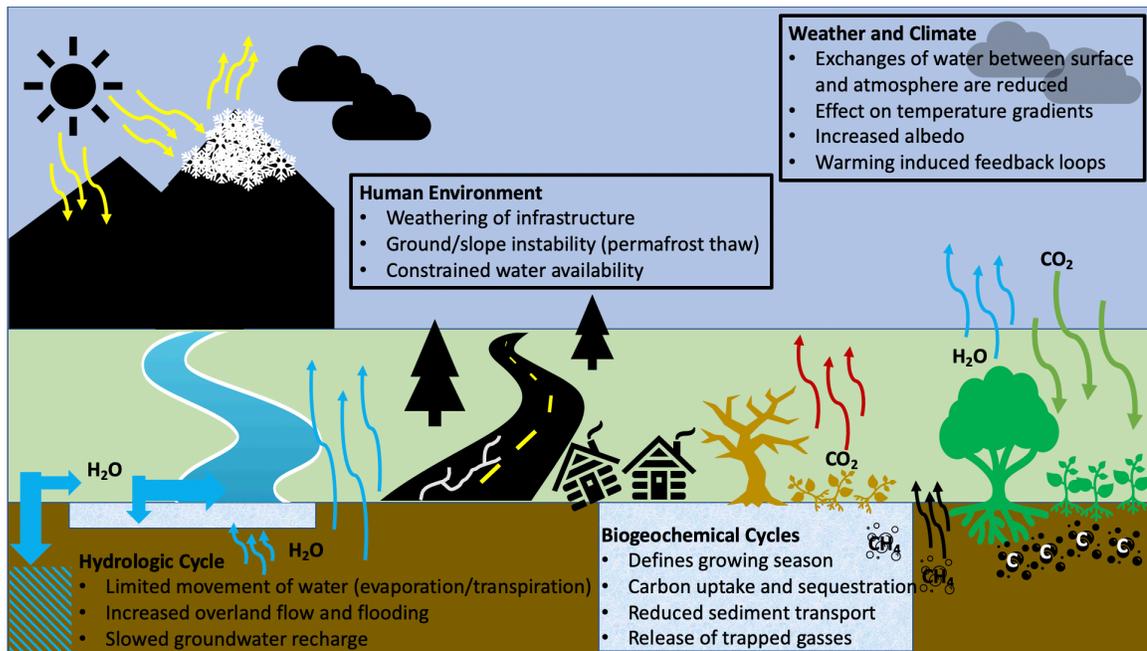


Figure 1 The effects of freeze/thaw cycles on various components of the Earth system

Detecting shifts in the duration of the frozen season using satellite observations has provided a new method to monitor a changing climate (Kimball et al. 2009; Kim et al. 2014; Du et al. 2019). In turn, due to the interconnected nature of FT and global climate, changes in FT seasonality can also have considerable impacts on climate, creating feedback loops. Existing records have been applied to identify a shortening of the frozen season (Kim et al. 2012), finding reductions in the frozen period by nearly 1

week and an increase in the number of diurnal FT cycles by 6 days across the Northern Hemisphere between 1979 and 2008. This carries direct implications for permafrost thaw, the release of trapped gases such as methane, and the duration and quantity of carbon uptake (*Arndt et al. 2019; Biskaborn et al. 2019*). The northern hemisphere growing season largely controls carbon sequestration as nearly 70% of the global land area exists in the northern hemisphere, including over 90% of the FT constrained regions. Regarding thawing permafrost, as the frozen season duration decreases and temperatures warm, the soil active layer (unfrozen layer) is expected to also expand (*Hugelius et al. 2011*). This may result in increased potential for plant productivity, an extended period of carbon uptake, and may also allow certain species to expand pole-ward while contributing to the extinction of others. Freeze/thaw cycles remain relevant in areas ranging from civil infrastructure and water resources to global circulation patterns and climate feedbacks. The extent to which these cycles effect the Earth system make their accurate observation chiefly important for scientific research and in other operational applications.

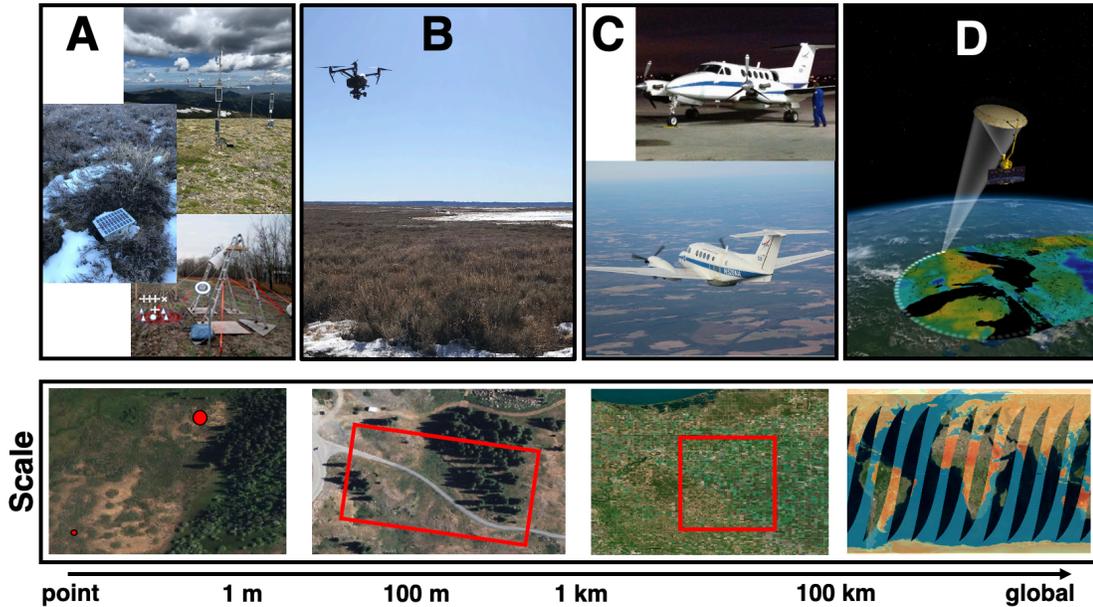
## **1.2 Current Observational Approaches**

Due to the fundamental role and impact of freeze/thaw on climate, hydrology, and ecology, widespread efforts have been invested in the observation of these cycles. Initially, observational approaches of surface FT states relied on sparse networks of ground (in-situ) observing sites. In most cases these sites record air temperature, with some collecting information on soil temperature and moisture. An even smaller subset collects information on snow depth and surface layer temperatures. As a result, traditional estimates of freeze/thaw transitions have been made largely using air temperatures as

they fall below the freezing point (0 °C). Insulation of the soil due to snow or soil can result in considerable lag between air temperatures and observed soil FT states. Warming due to exposure to direct sunlight and freezing point depression can also result in considerable FT uncertainty, especially when using air temperature observations alone. Therefore, the combination of air and soil temperatures have provided a new means of validating estimates of the surface freeze condition. Still, these types of observations are only representative of single points in space (**Figure 2A**) and remain sparsely distributed around the world. This means, while useful, these observations may not be representative of regional scales as surface heterogeneity introduces spatial variability in surface states. For example, the difference between a forested northern facing slope with abundant moisture and the reduced solar exposure to that of a drier south facing slope leads to very different microclimates over a relatively small distance.

In recent decades, the introduction of microwave remote sensing techniques, first using ground mounted sensors, became capable of observing FT transitions and soil moisture regionally (**Figure 2A**). The use of aircraft like in the Scanning L-Band Active Passive Experiment (SLAPEX; *Kim et al. 2016; Rowlandson et al. 2018*) has proven effective at collecting observational information over regions covering hundreds of square kilometers. Even more recently, unpiloted aerial vehicles (UAVs) have provided a method of quick deployment for field-scale (<1 km<sup>2</sup>) observations of soil temperatures and moisture using thermal imagery (*Berni et al. 2009*) and miniaturized microwave-band sensors (*Acevo-Herrera et al. 2010; Houtz et al. 2020*). These approaches are information dense and spatially distributed but remain limited for timeseries analyses

(barring regular operational flight schedules) or studies at the regional or continental scale (**Figure 2B-C**).



**Figure 2** Techniques for freeze/thaw observation and associated observational scale. (A) In-situ sites and ground-based radiometers (*Kim et al. 2016*), (B) UAVs and other small airborne vehicles, (C) Aircraft, from SLAPEX (*Kim et al. 2016*), and (D) Satellite; Soil Moisture Active Passive (SMAP, NASA JPL) is shown

Satellites enable both regular revisit times and global observational coverage (**Figure 2D**). The use of similar microwave sensors on board polar orbiting satellites have provided FT relevant observations at a global scale. These approaches capture continuous records that can extend back decades while also providing daily products in near real time. These can be useful for long term climate studies as well as operational work like flood prediction and agricultural decision making. With continued improvements in technology and thus satellite coverage and resolution, remote sensing provides the best opportunity for capturing freeze/thaw states within the Earth system. Still, with increases

in spatial coverage come reductions in resolution, meaning both field-scale and sub-daily (i.e., hourly) collections remain largely beyond satellite capabilities. Integration of these multi-scale datasets can thus provide the best means of accurately capturing these processes.

### 1.2.1 Physics of Microwave Remote Sensing and Freeze/Thaw Transitions

All materials emit energy within the electromagnetic spectrum. The magnitude of this energy within a given portion of the spectrum (i.e., spectral energy density) depends on both the surface properties and the kinetic (or physical) temperature of an object. This decreases significantly in wavelengths above the visible portion of the spectrum ( $>1 \mu\text{m}$ ). The theoretical magnitude of this emitted energy for a perfect emitter (i.e., black body) is determined by Planck's law for blackbody radiation (**Equation 1**) and is a function of wavelength ( $\lambda$ , meters) and temperature ( $T$ , Kelvin). Constants include the Planck constant ( $h$ , J-s), Boltzmann constant ( $k$ , J/K), and the speed of light ( $c$ , m/s):

**Equation 1 Planck's law of blackbody radiation**

$$E(\lambda, T) = \frac{2hc^2}{\lambda^5} \cdot \frac{1}{e^{\frac{hc}{\lambda kT}} - 1}$$

In nature, perfect black body emitters are unrealistic as material properties affect the emission of energy. Thus, observations acquired through remote sensing methods are considered as brightness temperatures ( $T_b$ , **Equation 2**), or the observed magnitude of energy of an equivalent black body with a known temperature ( $T_{\text{eff}}$ ).  $T_b$  values are calibrated using hot and cold targets with known emissive properties.  $T_b$  can then be considered as a function of both the surface emissivity ( $\epsilon$ , unitless), which changes based

on surface physical properties, and  $T_{eff}$  (Ulaby *et al.* 1986; Kerr and Njoku 1990). In other words, these observations provide a measure of emitted radiation at a specific wavelength ( $\lambda$ ), assuming the target acts as a black body ( $\epsilon = 1$ ).

**Equation 2 Simplified formulation for brightness temperature**

$$T_b = \epsilon T_{eff}$$

However, since true emissivity at the Earth's surface varies considerably due to material physical properties and sensing band, brightness temperature observations can be used to infer properties of the surface. One such property is the dielectric constant (or relative permittivity), which measures the ability of a material to store energy in an electric field. Importantly, water is known to have high permittivity relative to that of ice (**Figure 3**). In terms of **Equation 2**, these changes have relevance to the emissivity ( $\epsilon$ ) of an object particularly in the microwave portion of the spectrum, as  $\epsilon$  varies inversely with permittivity. Therefore, if the temperature and moisture content of a surface is held constant and permittivity is reduced (e.g., less water and more ice)  $T_b$  is expected to increase along with  $\epsilon$ .

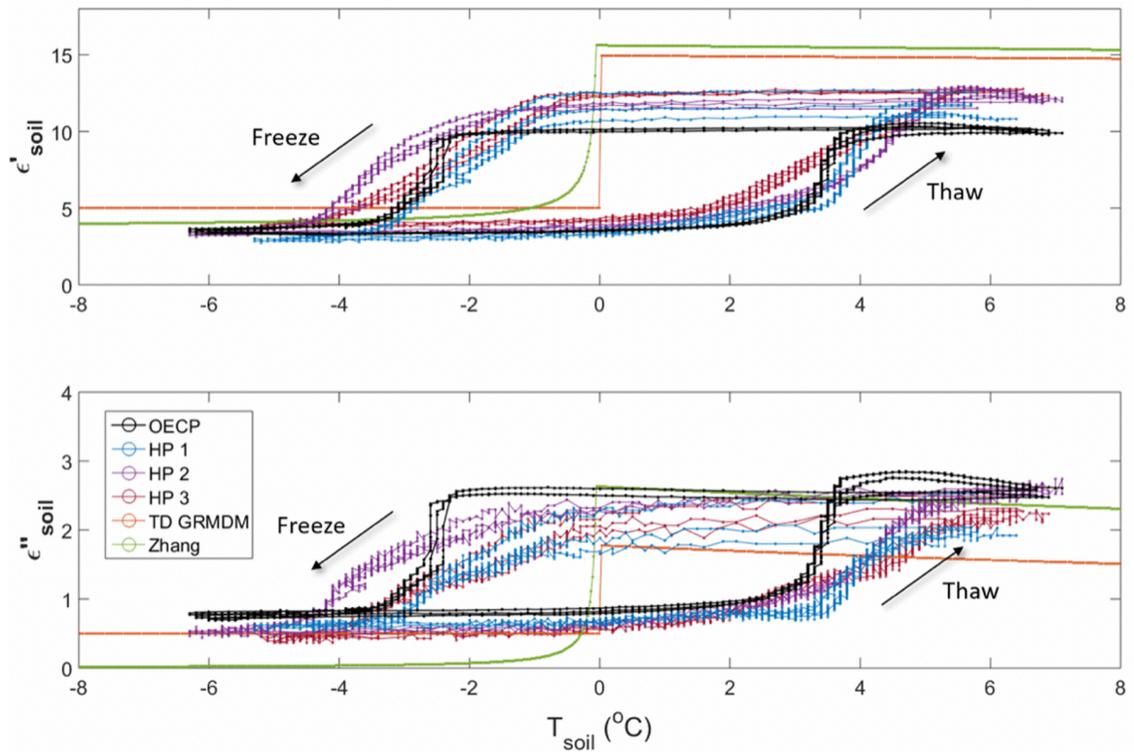


Figure 3 Real ( $\epsilon'$ ) and imaginary ( $\epsilon''$ ) permittivity of an organic soil sample using various instruments (OECP and HP) and model results (TD and Zhang), from Mavrovic et al. (2021)

### *Interpretation of Microwave Brightness Temperature Observations*

In theory,  $T_b$  changes are easy to interpret when the factors contributing to emissivity changes or physical temperatures are held constant. In reality, these factors can vary simultaneously and spatially, complicating the interpretation observed brightness temperatures. Such factors, beyond just the presence of water, include its quantity. As increased volumetric water content (VWC) results in higher permittivity and reduced  $T_b$ . This sensitivity has been widely leveraged for soil moisture estimation (Zuerndorfer et al. 1990; Kerr et al. 2012; Dunbar et al. 2018). In other words, as more liquid water exists within in a surface such as soil or snow, the more microwave energy is dissipated,

absorbed, or stored, resulting in lower magnitudes of emitted microwave radiation and lower  $T_b$ . As an extreme case, surface waters (e.g., lakes, rivers, and ponds) tend to have very low emissivity in the microwave band (**Figure 2D**). As a result, emissivity can vary considerably over regions with abundant surface water features and land (i.e., mixed pixels) which is known to complicate the observation of soil moisture or freeze/thaw transitions (*England 1990*).

Physical temperatures also remain an important consideration, as increases in  $T_{\text{eff}}$  are proportional to increases in  $T_b$ . However, both temperature and moisture states tend to vary considerably throughout a given year. Variations of which can occur coincidentally. For example, soil freezing requires colder physical temperatures (i.e., a reduction of  $T_{\text{eff}}$ ) and results in lower magnitudes of observed energy. But, as soil freezes, this also results in increased  $\epsilon$  due to decreased permittivity from phase changes (*Zhang and Armstrong 2001*). In moist soils, this becomes the dominant response since  $T_b$  increases more due to the transition between states than due to the minimal decrease in kinetic temperature required for freezing (*Dobson et al. 1985*). Nevertheless, these changes in properties may not occur right at 0 °C which can make the interpretation more challenging, but also may present opportunities to identify intermediate phases during such transitions (**Figure 3**). Over dry regions, moisture influence is minimal and because of the relatively small VWC, emissive properties remain relatively constant even as phase changes occur. In these regimes, the physical temperature of a surface is considered to provide the dominant contribution to microwave  $T_b$  observations.

### ***Emission Depth in Microwave Remote Sensing***

The frequency ( $f$ ) of the observing band provides critical controls on the relevance of any remote sensing observation. Band properties control which components of the surface dominate the observed emissions. As an example, sensors measuring shorter yet higher energy wavelengths, such as those in the thermal infrared band, only detect temperatures representative of a very small layer at the surface (skin). Due to short wavelengths ( $\sim 7 - 13 \mu\text{m}$ ), these observations are also impeded by water droplets ( $10 \mu\text{m}$  to  $>1000 \mu\text{m}$ ) within clouds. This allows for observations of cloud top temperatures for meteorology but prohibits all-weather surface temperature observation.

Similar considerations exist in microwave radiometry and are usually referenced as the emission (or penetration) depth, or depth to which observed energy is emitted from within a medium. This is known to be a function of frequency (or wavelength,  $f = \frac{c}{\lambda}$ ). Longer wavelengths  $>15\text{cm}$  (L-band microwave and above) thus are less affected by surface volume scattering (e.g., due to surface heterogeneity, vegetation, or snowpack water content) and have increased capabilities for sub-surface estimation compared to Ka-band ( $0.75 - 1.1 \text{ cm}$ ), which has higher correlations with air and soil surface temperatures (*Zuerndorfer et al. 1990*). This is because in higher frequencies (shorter wavelengths), the soil column ‘appears’ more heterogenous resulting in increased volume scattering at the surface thus relative reductions in the observed  $T_b$  during soil freezing. This is compared to longer wavelengths which respond more to soil properties, which can result in increased  $T_b$  due to soil freezing (*Zuerndorfer et al. 1990; England et al. 1991*).

In microwave bands, an effective emission depth of around 10% of the free space wavelength is expected in moist soils. In commonly used L- and Ka-bands this is equivalent to 1.5 – 3 cm and ~0.1 cm emission depths, respectively. Meaning these layers provide the dominant contribution to observed signals. For frozen soils, this can increase by around 3x as dielectric loss factors are reduced when freezing occurs, meaning less energy is attenuated or scattered within the medium (*Zuerndorfer et al., 1990; Jin et al. 2009*). Over dry regions with VWC around 4%, emission depth can extend to approximately one wavelength (L-band, 15 – 30cm; Ka-band ~1 cm). In especially dry, frozen, and/or sandy regions, microwave emission can even be representative of depths exceeding 50 cm below the surface (*Ulaby et al. 1986; Owe and Van de Griend 1998*).

#### ***Active vs. Passive Remote Sensing and Band Polarizations***

All of these inferences are made possible due to the relative transparency of water vapor and the minimal sensitivity of microwaves to atmospheric effects. This makes microwave remote sensing methods, both using active and passive approaches, effective for all weather monitoring of surface states (i.e., not effected by cloud cover). While this work focuses on the use of passive observations (i.e., radiometry), which relies on observation of naturally emitted energy, active approaches such as Radio Detection and Ranging (Radar) also exist using microwave bands. These techniques differ from passive ones in that they provide their own illumination source, similar to the flash on a digital camera. This can enhance resolution of observations and increase interpretability by using a known illumination source. Even so, these approaches come with drawbacks regarding observational coverage, sensor cost, and complexity.

In the case of linear polarization, the orientation of electromagnetic waves can be considered as either vertical or horizontal. This signifies waves either on the vertical or horizontal plane relative to the surface of the Earth. Many spaceborne remote sensing platforms collect observations across these polarizations, as the presence of highly polarized water molecules result in variable emission characteristics (*Ulaby et al. 1986*). Generally, regarding the phase change of water, horizontal polarizations have been observed to have greater magnitude changes relative to vertical polarizations (*Brucker et al. 2014*).

### ***Applied Microwave Remote Sensing***

In summary, interpreting changes in brightness temperatures ( $T_b$ ) depends on moisture content, surface temperature, water phase, and band specific sensitivities. Other differences in surface properties between various soils and vegetations can also induce changes in  $T_b$ . This can complicate determination of signal contributions in the microwave portion of the spectrum making both soil moisture and freeze/thaw detection non-trivial. Still, due to the particular sensitivities of microwave bands to water relevant variables, passive microwave (P-MW) remote sensing has proven effective at capturing state changes of water at the Earth's surface. Preceding studies have confirmed distinct changes in time series of  $T_b$  from both ground and satellite-based radiometers in L-band (1-2 GHz) and Ka-band (26.5 – 40 GHz) in order to characterize FT transitions (*Zhang and Armstrong 2001; Zhang et al. 2010; Kim et al. 2011; Kerr et al. 2012; Rautiainen et al. 2012, 2014, 2016; Derksen et al. 2017*). These efforts, among others, provide the

foundations for applying P-MW sensing approaches for the detection of frozen surfaces at global scales.

### **1.2.2 An Introduction to Existing Sensors and Available Freeze/Thaw Products**

As has been established, in the right bands, satellite remote sensing approaches are directly sensitive to changes in surface properties due to the presence of water or ice. In some cases, freezing point depression can result in soil freezing at temperatures below 0 °C, which can complicate FT observation even when using contact surface temperature observations. Impressively, microwave observations may present the capability for detection of ice even in these cases, providing a better true measure of surface states. Other observations such as those in the thermal infrared (TIR) band, may detect FT relevant information about surface temperatures, but do not have direct sensitivity to surface FT states as do microwaves.

Current global FT products are limited to those developed as part of the FT-ESDR (1979-2020, Version 5; *Kim et al. 2021a*) and the recent SMAP FT product (April 2015 – present, Version 3; *Xu et al. 2020*). Both use similar seasonal thresholding algorithms (STAs) to distinguish between frozen and thawed states. To reach current versions, these products have been continuously refined since their inception; FT-ESDR (2009; *Kimball et al. 2009*) and SMAP-FT (2015; *Derksen et al. 2017*). General enhancements include:

- 1) Improved resolution through use of higher resolution data inputs
- 2) Improved masking procedures and reduction of false freezes or reduced accuracy over snow/ice covered regions
- 3) Updates of freeze/thaw reference states

- 4) Implementation of regionally variable algorithms
- 5) Extension of record length and more rigorous validation

The products detailed herein provide a reference baseline of current global FT detection techniques and operational records.

### ***Soil Moisture Active Passive Freeze/Thaw (SMAP-FT)***

Following the recent launch of the Soil Moisture Active Passive (SMAP) mission in 2015, FT classifications have been derived from 1.41 GHz radiometer (L-band) observations and gridded to 36 km (or ~9 km) cell size (SMAP-FT) (*Derksen et al. 2017; Dunbar et al. 2014-2020*). This instrument follows a near-polar orbit passing the equator twice-daily at approximately 6 a.m. and 6 p.m. local solar time. The sensor follows an 8-day orbit repeat cycle, though the majority of the global surface, especially in polar regions, is imaged twice-daily. Thus, daily coverage gaps tend to exist in the mid-latitudes. Valuable precursor investigations demonstrated L-band sensitivity to FT state including the use of Soil Moisture Ocean Salinity (SMOS) mission observations (*Rautiainen et al. 2012, 2014, 2016*), ground-based radiometer and in-situ observation studies (*Roy et al. 2015, 2017a, 2020*), and focused field experiments such as SMAPVEX (2012) and Scanning L-Band Active Passive experiment (SLAPEX 2015) (*McNairn et al. 2015; Kim et al. 2016; Rowlandson et al. 2018*). These studies have demonstrated the ability of L-band observations to detect soil specific freezing processes more effectively than higher frequency observations (Ka-band), whereas also revealing FT classification challenges in dry regions, near water, and over complex terrain.

SMAP-FT is based on the combination of the normalized polarization ratio (NPR) seasonal thresholding algorithm (STA) and the single channel vertically polarized (SCV) algorithm. In the colder portion of the domain, which experiences clear and cyclical seasonality of freeze/thaw cycles, using information from each polarization (V- and H-, NPR) provides an improved classifier. **Equations 3, 4, and 5** detail the NPR-STA which is applied to all grids meeting algorithm requirements (*Dunbar et al. 2014-2020*).

**Equation 3 Normalized polarization ratio formula for SMAP bands**

$$NPR(t) = \frac{Tb_{1.41V} - Tb_{1.41H}}{Tb_{1.41V} + Tb_{1.41H}} \times 100$$

First the NPR value is calculated as a unitless value between 0 and 100 (**Equation 3**). For each location, reference values for frozen and thawed conditions are then defined. These are considered to be the standard frozen and thawed signatures for a given location, as estimated from air temperatures (**Equation 4**).

**Equation 4 Seasonal scale factor used in SMAP-FT algorithm**

$$\Delta t = \frac{NPR(t) - NPR_{fz}}{NPR_{thw} - NPR_{fz}}$$

**Equation 5 Thresholding for freeze/thaw classification NPR algorithm**

$$\begin{aligned} \Delta t &\geq T, \text{thaw} \\ \Delta t &< T, \text{freeze} \end{aligned}$$

Finally, using the seasonal scale factor to indicate similarity to thawed conditions, a seasonal threshold (T) is estimated for each pixel using air temperature records to capture FT states (**Equation 5**).

A modified baseline STA approach is applied for all other locations, which primarily include the mid- and lower latitudes. In these regions, only vertically polarized

$T_b$  measurements in the L-band (1.41 GHz) are used to determine surface states. These efforts first determine the correlation (R) of SMAP  $T_{b_{1.41V}}$  observations to air temperatures, which then are used to determine regional FT thresholds (**Equation 6**).

**Equation 6 Freeze thaw single channel V-pol (SCV) algorithm**

$$\begin{aligned}
 R > 0.5, \quad FT &= \begin{cases} \text{thaw, if } T_{b_{1.41V}} > \text{threshold} \\ \text{freeze, if } T_{b_{1.41V}} \leq \text{threshold} \end{cases} \\
 R < -0.5, \quad FT &= \begin{cases} \text{thaw, if } T_{b_{1.41V}} < \text{threshold} \\ \text{freeze, if } T_{b_{1.41V}} \geq \text{threshold} \end{cases}
 \end{aligned}$$

Recently, SMAP-FT has also implemented masking procedures to prevent false freeze flags through a combination of thresholding ( $T_b > 273$  °K indicates thaw) and by applying climatological rules (i.e., never frozen) learned from existing long-term records of FT states (i.e., the Freeze/Thaw Earth System Data Record).

Numerous efforts into SMAP FT validation have been performed (*Lyu et al. 2018; Rowlandson et al. 2018; Xu et al. 2018; Kim et al., 2019*). Thorough analysis at SMAP core validation sites was performed over tundra and boreal environments showing the impacts of vegetation and deep snowpack on FT classification, with strongest performance over tundra sites occurring in areas of shallow snow and minimal vegetative cover (*Xu et al. 2018*). Additionally, these studies have shown similar characteristics of improved agreement with in-situ soil and air measurements in the afternoon (PM) as compared to morning (AM) observations as the poorest agreement occurs in spring when melting is widely prevalent. Since algorithm thresholds are optimized based on air temperatures, this may have an effect on increases in morning (77.1% soil to 81.0% air) and afternoon (78.6% to 86.6%) classification agreement at SNOTEL sites in Alaska relative to soil temperatures (*Xu et al. 2018*). Utilizing air temperature observations at

nearly 5,000 World Meteorological Organization (WMO) stations SMAP was found to agree 78% and 90% in the AM and PM respectively, mirroring results at SNOTEL and core validation sites (*Kim et al. 2019*). Further details regarding recent product versions can be found in recent SMAP algorithm theoretical basis documents (ATBD; *Dunbar et al. 2014-2020*).

### ***Freeze/Thaw Earth System Data Record (FT-ESDR)***

Estimates of freeze/thaw states from Measure's Earth System Data Record (FT-ESDR) are acquired through a methodology similar to SMAP-FT but using different observing bands and with varied observational times. FT-ESDR, consists of multiple twice-daily Ka-band based FT records:

- 1) SSM/I-FT, derived from Special Sensor Microwave Imager (Sounder) series (SMMR, SSM/I, SSMIS) 37.0 GHz  $T_b$  and,
- 2) AMSR-FT, Advanced Scanning Microwave Imager series (AMSR, AMSR-E, AMSR-2) 36.0 GHz  $T_b$  observations (*Kim et al. 2011*)

This remains the longest existing global record of FT classifications and extends from 1979-2020 using 25 km grid spacing. A new spatially enhanced ESDR product (6 km) has also been recently made available as of April 2021. Observations are representative of overpass times for SSMIS which are similar to SMAP (6 a.m./p.m.), while AMSR-2 has equatorial crossings at approximately 1:30 a.m. and 1:30 p.m. local solar time.

Vertically polarized microwave observations from these platforms have been shown to have the highest classification accuracies for FT detection, relative to horizontal

(Kim et al. 2011). Thus, FT records are derived from twice-daily overpasses using  $T_{bv}$  observations. In lieu of the NPR approach, the single channel seasonal threshold approach is used within ESDR (Kimball et al. 2009; Kim et al. 2017). Expressly, the vertically polarized  $T_b$  observations from both satellite platforms are used to determine surface states through locally optimized thresholds (SCA-SCV). In other words, an identical algorithm is used as described in **Equations 4 and 5**, although NPR is replaced with band specific  $T_{bv}$  observations. Earlier releases of FT-ESDR data records included optional edge detection and moving window approaches in addition to the aforementioned STA baseline algorithm. These have since been removed.

Recent validation efforts for new product versions rely on a combination of lake and river ice phenology records and global air temperature observations (Kim et al. 2021b). One such validation study of FT-ESDR, performed against WMO station air temperature measurements, showed annual agreement of 90.3% and 84.3% when using afternoon and morning retrievals, respectively (Kim et al. 2017). This indicates the strong capability of Ka-band P-MW sensors to detect changes at the soil surface. Observations from ground-based radiometers in the Ka-band have also proved highly correlated to soil FT states in the 0 to 5 cm soil layer (Zhang et al. 2010). Further information on these products and related research activities are available via the University of Montana (<https://www.ntsg.umt.edu/freeze-thaw/>).

### ***Other Efforts Towards Freeze/Thaw Estimation***

While no global product exists outside of SMAP and FT-ESDR records, additional freeze detection techniques have been employed on regional scales to enhance

FT classification, including over China and Alaska. These algorithms, extending beyond STAs and change detection, have been applied to capture the complexity of FT classification. Such approaches include decision trees (*Jin et al. 2009*), multi-band discriminant and probabilistic algorithms (*Zhao et al. 2011; Zwieback et al. 2012*), moving window or change detection approaches (*Mortin et al. 2012; Chen et al. 2019*), estimation of frozen soil fraction (*Prince et al. 2019*), the use of multiple MW band reflectance values (*Muzalevskiy and Ruzicka 2020*), as well as the combination of surface skin temperature with satellite brightness temperatures (*Zhao et al. 2017; Hu et al. 2017*). *Chai et al. (2014)* demonstrated the application of several algorithms to SSMIS and AMSR-E including decision trees, the use of FT indices, and discriminant functions and found that no one algorithm had dominant classification accuracy in all areas (i.e., region, time of day).

Binary landscape scale observations used in both global (SMAP-FT and FT-ESDR) records simplify complex freeze/thaw processes down to a coarse binary determination. Aforementioned novel classification approaches incorporate additional information combined with increasingly advanced computing techniques to achieve comparable or improved accuracy relative to global records. As a result, major steps have been taken towards improving the representation of FT states beyond coarse resolution binary products and in efficiently leveraging information from different sources. Similar approaches are explored within this work in the development of global data fusion approaches for FT classification.

### 1.2.3 Known Limitations

As with any observational record, there remain limitations and uncertainties resulting from spatially varying and complex band sensitivities covered in prior sections. These uncertainties have been explored to a degree prior to the release of global scale FT records (Kimball et al. 2009; Derksen et al. 2017). Major factors known to affect the interpretability of P-MW observations for freeze/thaw detection include:

- 1) **Moisture.** *Current approaches are challenged over dry areas due to the lack of a clear freeze/thaw signal. Also, in moist regions variability in moisture content can change  $T_b$  reference states corresponding to the thawed or frozen conditions*
- 2) **Surface Water and Mixed Pixels.** *Surface water within mixed land-water pixels can result in significant reductions in the observed P-MW signal, thus complicating freeze/thaw detection*
- 3) **Snow.** *Shallow and dry snowpacks are largely transparent to microwaves especially in longer wavelengths. Even still, deeper, or wet snow at the surface can challenge retrieval algorithms*
- 4) **Vegetation and Biomass.** *While vegetation can be transparent to longer wavelength microwaves, the thickness of the vegetative layer and its water content can complicate extraction of soil information from below this layer (Ulaby et al. 1986). The canopy can also intercept snow and cause more rapid melt above the surface, leading to more challenging signal interpretation*

5) **Resolution and Topography.** *There remain limits on the scales that can be observed using P-MW techniques alone (~10 km). In areas with considerable variability in climates (i.e., complex terrain, sharp elevation gradients) can induce in a complex mixture of climates and surface states*

6) **Validation Approach Variations.** *Differences in product calibration/validation, such as the use of soil or air temperature observations, can induce uncertainties*

While P-MW remote sensing approaches have known sensitivities to various land surface properties as detailed here, the degree to which these factors impact freeze detection is not well understood.

### **1.3 Project Overview**

This section provides a concise overview of each section contained herein and defines research objectives underpinning this project. Extensive efforts went into data collection and analysis techniques included within this venture. The breadth of and unique strengths of various data inputs (i.e., in-situ/fieldwork, land cover, satellite observation) can help to enhance physical meanings and the observational approaches used for FT detection. *Note: FT detection approaches have continuously been improved throughout the duration of this study (2017-2022). Through releases of new product versions some of the clear deficiencies detailed both in this document and by algorithm development teams are addressed. Due to continued change within the field, products used in earlier efforts (Chapter 2) may be outdated even as many of the conclusions remain accurate.*

### 1.3.1 Research Objectives

- 1) *Improve characterization of existing freeze/thaw products by identifying their limitations and exploring the relationships between existing records*
- 2) *Determine the utility of additional observations and new variables for improving freeze/thaw classification using focused regional-scale studies and field work*
- 3) *Explore the value of and potential for implementation of new freeze/thaw metrics (i.e., fractional, probabilistic, enhanced resolution, and/or new freeze type identifiers)*
- 4) *Develop a freeze/thaw classification methodology with improved accuracy at a global scale by leveraging additional data and global networks of ground-truth observations*

### 1.3.2 Project Tasks

To accomplish the objectives highlighted above, this work begins with foundational explorations of existing classification approaches and moves towards the development of novel methods for freeze/thaw classification. The primary tasks are split into three stages (with the relevant chapters indicated):

**Task 1. Exploring existing freeze/thaw records; their limitations, sensitivities, and relationships to land surface variables (Chapters 2 - 3)**

**Task 2. Determine the driving factors of freeze/thaw transitions at higher resolutions, the potential for resolution enhancement, and utility of new classifiers through focused regional studies and in-situ/ground observations (Chapter 3)**

### **Task 3. Algorithm and predictor selection, sensitivity analyses, and validation of a new data driven global freeze/thaw classifier (Chapter 4)**

While included herein in a modified state for continuity, the chapters as presented in this work largely combine efforts from several research articles both previously published in leading peer-reviewed journals or that are currently undergoing revision. Each of the following chapters are founded on these works, detailed below.

**Chapter 2.1:** Johnston, J., Maggioni, V., & Houser, P. (2019). *Investigating the Relationship Between Satellite-Based Freeze/Thaw Products and Land Surface Temperature. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1–25.

<https://doi.org/10.1109/JSTARS.2019.2926942>

**Chapter 2.2:** Johnston, J., Maggioni, V., & Houser, P. (2020). *Comparing global passive microwave freeze/thaw records: Investigating differences between Ka- and L-band products. Remote Sensing of Environment*, 247, 111936. <https://doi.org/10.1016/j.rse.2020.111936>

**Chapter 3.1:** Johnston, J. M., Houser, P. R., Maggioni, V., Kim, R. S., & Vuyovich, C. (2021). *Informing Improvements in Freeze/Thaw State Classification Using Subpixel Temperature. IEEE Transactions on Geoscience and Remote Sensing*, 1–19.

<https://doi.org/10.1109/TGRS.2021.3099292>

**Chapter 3.2:** [Under revision] Johnston, J. M., Houser, P. R., Maggioni, V., & Pestana, S. (2022). *Exploring Fine-Scale Temperature Processes with Unpiloted Aerial Vehicles (UAVs)*.

**Chapter 4:** [Under revision] Johnston, J. M., Houser, P. R., Maggioni, V., & Xue, Y. (2022). *A New Machine Learning Framework for Enhanced Global Freeze/Thaw Estimation*.

## 2 IMPROVING CHARACTERIZATION OF EXISTING FREEZE/THAW OBSERVATION PRODUCTS

This chapter focuses on analyzing existing freeze/thaw (FT) products and their respective differences, capabilities, and limitations. While FT estimates using remote sensing data at continental scales have been around for decades (*Rignot and Way 1994; Judge et al. 1997*), no long-term continental scale records of FT states had been produced prior to 2009 (*Kimball et al. 2009*). In recent years, technological advancements have led to rapid growth within the field of cryosphere sciences as well as in remote sensing in general. While meteorology, visible imagery and land use, and communications applications have historically received the majority of funding for spaceborne endeavors, an array of new Earth Observing Systems (EOS) have provided a collection of new observational bands for emerging scientific applications such as FT state observation.

As a result, methods to extract FT states from microwave observations have remained a growing topic of scientific interest. Still, determinations of the physical surface characteristics being observed, such as snow melt, soil freeze/thaw, or changes in moisture status remain challenging to separate, most notably across heterogeneous regions. Variability across such land surface characteristics is known to effect microwave (MW) emissions, thus supporting the importance of improved regional product characterization. However, the true observational sensitivities are not well understood at regional scales. One challenge is the lack of appropriate validation datasets. In-situ sensing sites, which can be sparsely dispersed (especially in remote regions) and are unrepresentative of larger areas, have traditionally provided the validation baseline for

assessing FT classification accuracy using soil and air temperature (*Zhang et al. 2003; Kimball et al. 2009; Dunbar et al. 2014-2020*). Recent advances in modeling and data assimilation capabilities have opened up new opportunities for validation using reanalysis products which rely both on physically based models and ground observations. These approaches can provide data that is accurate, continuous, and can be derived at regularly spaced grids.

To explore questions into the characterization of existing FT records and advance the literature on existing remote sensing FT classification approaches, this chapter examines how near-surface temperature observations such as those of soil, air, and land surface (skin) temperature relate to various operational passive microwave (MW) derived FT classification products. Also, through inter-comparisons between existing global FT records, this chapter explores where existing products diverge to assess where FT classifications remain uncertain and may require additional refinement.

## **2.1 Examining Near Surface Temperature Variables and their Relationship to the Microwave Derived Freeze/Thaw Observations**

To address gaps in knowledge regarding existing remote sensing observations of surface freeze/thaw, the relationship between surface temperature variables and passive microwave-derived surface freeze/thaw states is explored. Utilizing retrievals from the Soil Moisture Active/Passive (SMAP), Advanced Microwave Scanning Radiometer (AMSR-2, AMSR-E), and Special Sensor Microwave Imager/Sounder instruments (SSM/I, SSMIS), surface FT records have previously been derived globally. Herein, these

records are compared against spaceborne surface temperature observations from the Moderate Resolution Imaging Spectroradiometer (MODIS), and North American Land Data Assimilation System (NLDAS) skin, 0–10-cm soil layer, and 2-meter air temperatures. These datasets provide dispersed temperature data at regularly spaced grids and enable a regional scale assessment in a FT constrained region stretching across southern Canada and the northern portion of the Continental United States (CONUS). The relationship of binary FT classifications to temperatures are explored through thresholding (at 0 °C), comparing sub-grid scale variability in temperatures, and by analyzing the associated probability distribution functions of regional temperatures when classified as frozen or thawed. These efforts contribute to an improved characterization of which surface components provide the dominant contribution to the FT signal in existing products and has relevance to the enhancement of FT products moving forward. Improved knowledge as to which components of the landscape undergo freeze/thaw cycles has considerable relevance to hydrology, carbon and nutrient cycling, and climate studies.

### **2.1.1 Background and Objectives**

As satellite retrievals are acquired at various passive microwave (P-MW) frequencies, times, and spatial resolutions, significant variations in FT classifications can occur. These P-MW observations are generally sensitive to the holistic landscape FT state, which can include a heterogeneous combination of different vegetation covers, soils, snow, and terrain. This landscape complexity is not well understood, as defined FT state can represent canopy, snow surface, or soil FT state depending on observation band,

location, and time (Lyu et al. 2018). Similarly, these classifications have been shown to diverge from air temperature measurements by exhibiting sensitivity to land surface FT properties which may remain either colder or warmer than the surrounding air. This occurs especially in times of ephemeral FT events in which snow cover and soil thermal inertia can contribute to prolonged frozen periods even as air temperatures have climbed above freezing (Kim et al. 2015; Wang et al. 2016). Lower frequency L-band (1.41 GHz) brightness temperature ( $T_b$ ) measurements, as are utilized in the Soil Moisture Active Passive (SMAP) (Dunbar et al. 2014-2020) FT products, have been shown to have an increased emission depth sensitivity as compared to Ka-band retrievals (36.5, 37 GHz) (Rautiainen et al. 2014). These Ka-band retrievals, as are observed by the Advanced Microwave Scanning Radiometer series (AMSR, E, 2) (Kawanishi et al. 2003; Zhao et al. 2011) and the Special Sensor Microwave Imager/Sounder (SSM/I, SSMIS) (Grody 1991; Kim et al. 2017), are characterized by wavelengths ( $\lambda$ ) of around 1 cm. As a result, Ka-band retrievals exhibit increased sensitivity to surface features such as vegetation compared to SMAP L-band ( $\lambda = 15\text{--}30$  cm) retrievals (Zhang et al. 2010; Rautiainen et al. 2012, 2014). In both bands, both cloud cover and thin vegetation layers are largely transparent. However, dense canopies, snow interception and melt, and vegetation moisture conditions can affect MW emissions through changes in surface permittivity, especially in the Ka-band.

These differences in P-MW frequencies have resulted in two unique types of FT state products (FTSPs). Including; i) L-band derived, represented by the SMAP passive FT products (Version 1; Dunbar et al. 2014-2016), produced at both  $36\text{ km} \times 36\text{ km}$

(SMAP-FT) (Xu et al. 2016) and at an enhanced  $9 \text{ km} \times 9 \text{ km}$  (SMAP<sub>E</sub>-FT) (Xu et al. 2017; Long et al. 2019) resolution (**Table 1**), and ii) Ka-band based FTSPs, produced as part of the Freeze/Thaw Earth System Data Record (FT-ESDR) (Kimball et al. 2009), which are derived from AMSR and SSMI(S) instruments on a  $25 \text{ km} \times 25 \text{ km}$  grid (Version 4; Kim et al. 2018).

**Table 1 Selected satellite derived freeze/thaw products. Note: SMAP products been expanded to the global domain**

Freeze/Thaw (FT) Product (Sensors)	Microwave (MW) band	Spatial Resolution	Period of Record	Domain	Overpass Times (local solar)	
					Descending (AM)	Ascending (PM)
SMAP <sub>E</sub> - FT (SMAP)	L-band (1.41 GHz)	9 km	April 2015 - present	> 45° N*	6:00	18:00
SMAP - FT (SMAP)	L-band (1.41 GHz)	36 km	April 2015 - present	> 45° N*	6:00	18:00
AMSR - FT (AMSR-E, AMSR2)	Ka-band (36.5 GHz)	25 km	2002 - 2016*	Global	1:30	13:30
SSMI - FT (SSMR, SSM/I, SSMIS)	Ka-band (37 GHz)	25 km	1987 - 2016*	Global	6:30	18:30

\*at time of study completion

Both SMAP and FT-ESDR derived products utilize a seasonal thresholding approach, in which retrievals are compared to  $T_b$ -derived FT reference states in order to classify areas as frozen or thawed. FT product validation remains an ongoing process, though in-situ air and soil temperature measurements have been the predominant evaluation dataset in meeting classification accuracy goals of both SMAP and ESDR FT records. In validating FTSPs, the use of in-situ validation sites has in many cases been limited to areas of primarily homogenous terrain and vegetation. This is done in order to best represent landscape FT through point measurements but tend to be unrepresentative of large heterogenous areas. Moreover, the sparse availability of in-situ observations,

especially in high latitude regions, has made validation incomprehensive. The use of optical and thermal infrared RS techniques has been proposed to bolster validation efforts, but inconsistent coverage due to clouds, effects of snow cover, limited emission depth, and low solar illumination during FT transitions limit their use as effective methods (Kimball et al., 2009). *Of note, these products were the most up to date versions at the time this study was undertaken in 2018. As of 2022, SMAP-FT is currently on Version 3 and FT-ESDR Version 5 and include some algorithm updates such as a single channel approach in the mid-latitudes, extended record length, and resolution enhancements through down-sampling of input  $T_b$  observations.*

Previous studies have found high correlation of FT-ESDR products to air and soil temperature, as thresholding algorithms in many cases have been optimized based on these observations (Kimball et al. 2009; Kim et al. 2011). SMAP-FT validation efforts have followed a similar path by utilizing in-situ air and soil temperatures to evaluate and improve FT classification algorithms. This is done through the use of flags marking disagreement between FT state and in-situ observations (Dunbar et al. 2014-2020; Derksen et al. 2017). Evaluation and improvement of SMAP FTSPs is ongoing which considers the effects of terrain, surface water, snow cover, and vegetation through sensitivity analysis and masking procedures. Validation efforts, in a limited capacity compared to the use of air temperature observations, have been performed using modeled skin and soil temperatures as well as multi-scale measurements intended to reduce representation errors incurred through the use of point observations (Lyu et al. 2018). Previous studies have also applied SMAP based FT algorithms to latitudes south of 45 °N and validated these classifications utilizing air, 5 cm soil temperatures, and modeling

approaches (*Kraatz et al. 2018*). Historically, FT validation efforts have relied on imperfect measures, due to the difficulty in representing aggregate landscape FT state as is measured by P-MW remote sensing techniques. The SLAPEX field experiments demonstrated the effects of spatial heterogeneity on how P-MW observations relate to large scale FT classifications (*Rowlandson et al. 2018*). However, the majority of validation efforts continue to rely on point observations. As ground-based radiometer studies have shown varied retrievals depending on surface vegetation, soil, and moisture characteristics (*Zhang et al. 2010; Rautiainen et al. 2012*), satellite-based FT observations tend to only capture the aggregate emissions of these components making it difficult to accurately characterize entire areas at the kilometer scale.

As such, the primary goals of this study are:

- 1) *To determine the land surface temperature variables which act as the best surrogates to satellite FT product defined surface states; and*
- 2) *to investigate the differences and uncertainties among various FTSPs*

To do so, we assess the accuracy and limitations of four FTSPs over a study area encompassing varied terrain, climate, and land cover. Through a spatial analysis, we examine the factors contributing to FT classification differences between FTSPs and temperature-threshold based FT classifications. This study aims to improve our understanding of what current satellite FT products represent and ways in which they can be enhanced through a regional assessment of landscape FT complexity. Comparisons across existing global P-MW FT products have not been previously performed at this scale utilizing both modeled and satellite observed temperature variables. Hence, this

study fills a gap in the literature, contributing to the development of improved classification products.

## 2.1.2 Comparison Methods

### *Data Products*

This study compares the following four temperature datasets (**Table 2**) to SMAP, AMSR, and SSM/I FT products (detailed in **Table 1**):

- 1) National Land Data Assimilation System Version 2 (NLDAS) Noah land surface modeled skin temperature ( $T_{N\text{-skin}}$ ) (*Xia et al. 2009*)
- 2) NLDAS Noah 0-10 cm average soil temperature ( $T_{\text{soil0-10cm}}$ )
- 3) NLDAS forcing 2-meter air temperature ( $T_{2\text{m-air}}$ ) (*Xia et al. 2009*)
- 4) MODIS thermal infrared-based skin temperature ( $T_{M\text{-skin}}$ ) on board the Aqua satellite platform (MYD11C1 V006, *Wan et al. 2015*).

**Table 2 Selected temperature products**

Temperature Variable	Product Name	Spatial Resolution	Temporal Resolution	Period of Record	Domain
NLDAS 2-meter Air Temperature	NLDAS FORA0125 V2	0.125° (~11 km)	Hourly	1979 - Present	25 - 53° N, 67 - 125° W
NLDAS 0-10cm Averaged Soil Temperature	NLDAS NOAH0125 V2				
NLDAS Surface Skin Temperature	NLDAS NOAH0125 V2				
MODIS-Aqua Land Surface Temperature	MYD11C1 V006	0.05° (~4 km)	1:30 AM/PM local solar time	2002 - Present	Global

These comparisons are completed based upon product overlapping temporal (April 2015 – 2016) and spatial domains (45-53 °N, 67-125 °W), which includes the northern conterminous United States and southern Canada. This time series was sufficient to capture multiple seasonal and ephemeral FT transitions occurring within a 21-month period of FT product overlap.

The Phase 2 of NLDAS project (NLDAS-2) dataset has been improved over the previous version of NLDAS (*Cosgrove et al. 2003; Rodell et al. 2004*, <http://ldas.gsfc.nasa.gov>) based on output from the atmospheric data assimilation system at the NASA Global Modeling and Assimilation Office (*Suarez et al. 2005*). NLDAS forcing variables such as 2-meter air temperature are derived from North American Regional Reanalysis (NARR) datasets, which are based on in-situ station, radiosonde, aircraft, and satellite observations (*Mesinger et al. 2006*). Significant positive biases in shortwave radiation have been observed in NARR datasets but have been improved upon in NLDAS-2 forcing by performing bias correction with Geostationary Operational Environmental Satellite (GOES) observations. Several studies have validated NLDAS  $T_{2m-air}$  against observations from the Oklahoma Mesonet, the Atmospheric Radiation Measurement (ARM) program, and Cloud and Radiation Test Bed (CART) sites. High correlation of NLDAS to air temperature observations was shown across Mesonet, ARM, and CART sites having a small negative bias (*Cosgrove et al. 2003; Luo et al. 2018*). Other modeled surface temperature variables including  $T_{N-skin}$  and  $T_{soil0-10cm}$ , have also performed well compared to in-situ observations. Still,  $T_{N-skin}$  has been shown to tend towards a warm bias during the mid-day period as well as a 3-5 °C cold bias during the

nighttime hours. These skin temperature biases are also expected to contribute to error in soil temperatures, though near surface layer soil temperatures have been shown to compare best with observations relative to lower layers. This is reflected by a 2.5 °C annual negative bias when compared to soil temperature observations at 10 cm, best among modeled soil layers (Xia *et al.* 2012b). This cold bias in  $T_{\text{soil0-10cm}}$  is most pronounced in the late fall and early winter, while  $T_{\text{soil0-10cm}}$  tends to be lower than observations most notably during the nighttime hours of the winter months. Inaccuracies in modeling snow cover insulation remains a contributor. Multiple assimilation inputs along with imperfect model forcing have been shown to introduce small errors into NLDAS temperature data (Xia *et al.* 2012c). In contrast, MODIS  $T_{\text{M-skin}}$  observations are derived from surface emissivity estimates and at sensor observations of emitted thermal energy, with errors on the order of 2 °K (Wan *et al.* 1999; Duan *et al.* 2019). This must be considered when assessing study results, especially as it pertains to uncertainty around the freezing point.

Both standard (36 km) and enhanced resolution (9 km) SMAP FTSPs are used to investigate the degree to which resolution effects FT classification. This is particularly useful as temperature data are available at finer spatial resolutions, 0.05° (~5 km) and 0.125° (~12 km), as compared to the moderate resolution of FT-ESDR (25 km) and SMAP FT products (36 km). Additionally, both Ka-band based FTSPs included in the FT-ESDR are included to evaluate the effect of different retrieval times when using the similar sensing frequencies of AMSR (36.5 GHz) and SSM/I (37 GHz). Intercomparison between SMAP L-band and FT-ESDR K-band FTSPs is also of particular interest.

## ***Comparing Temperatures to Binary Freeze/Thaw Classifications***

### *Classification Accuracy*

Direct comparison of binary FT state and temperature is performed by thresholding the temperature values at 0 °C (~273 °K) and matching products to a common 0.5° grid (~50 km). The continuous temperature datasets are converted into a binary one with values either greater than 0 °C (thawed state) or lower than/equal to 0 °C (frozen state). Realistically, the ground surface may remain thawed at temperatures at or below the freezing point or frozen at temperatures above, due to freezing point depression and thermal inertia (*Campbell 1951; Suzuki 2004*). Nevertheless, this threshold provides a reasonable proxy for examining FT state under the assumption that phase changes at the surface will begin to occur at this temperature. After some investigation into threshold optimization, this value was set to 0 °C as to avoid the introduction of any additional biases. Overall classification accuracies (CA) are calculated for all overpasses within the study period by comparing the two binary datasets. CA quantifies the agreement between the temperature products (NLDAS and MODIS) and the FTSPs across the study region on a 0.5° grid using a cell-by-cell comparison:

**Equation 7 Classification Accuracy**

$$CA = \frac{FTy}{FTy + FTn}$$

where *FTy* indicates agreement, and *FTn* disagreement between data pairs. Large water bodies and grid cells missing frequent temperature or FT state classifications are excluded. Additionally, MODIS Aqua retrievals are limited in coverage due to cloud cover, therefore all regions are not equally represented. Cells containing no data in either

comparison member are excluded from overall CA calculations. Previously, similar validation approaches have been performed over China utilizing both AMSR and SSM/I-based FT classifications (Chai et al. 2014).

*Proportional Comparison Technique*

In fitting data to a coarser resolution grid, simple averaging can result in mischaracterizing the true ground state, especially in regions with complex topography where the extent of frozen ground can vary considerably. Alternately, a grid-by-grid proportional analysis is adopted. Frozen proportions are calculated as the number of defined frozen values divided by the total number of data points within each 0.5° grid cell. After these values are calculated for both temperature-based FT classes and FTSPs, a difference in frozen proportion ( $D_{FTp}$ ) is defined by subtracting temperature product frozen proportion ( $F_{Tp}$ ) from the corresponding FTSP frozen proportion ( $F_{FTp}$ ):

**Equation 8 Classification Accuracy**

$$D_{FTp} = F_{FTp} - F_{Tp}$$

For example, if all NLDAS air temperatures (1/8°) within a 0.5° grid cell (16 values) are below the freezing point and only half of the FTSP pixels are classified as frozen,  $D_{FTp}$  would be -0.5, signifying 50% less frozen area from the FTSP compared to the temperature-based frozen area. Anywhere from 2 to 25 FT classifications are contained within each 0.5° cell, dependent on location and FTSP resolution. This method avoids effects of significant temperature variations and rounding errors incurred when upscaling FT state variables to a relatively coarse 0.5° grid. The proportional comparison methodology also enables a spatial representation in which positive values represent

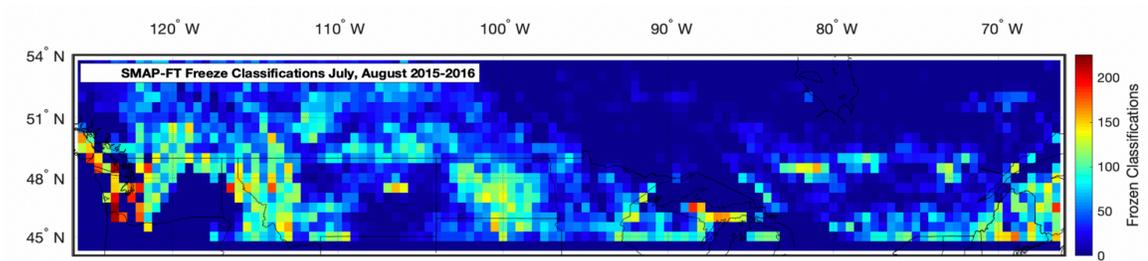
greater FTSP defined frozen area as compared to temperature products, whereas negative values define a lesser FTSP defined frozen area in relation to temperature threshold-based products. This methodology moves beyond the binary agreement approach and contributes to the identification of contributors to FT classification disagreement.

#### *Freeze/Thaw Temperature Distributions*

To determine the temperature variables with the strongest relationship to FTSP states, all corresponding temperature values deemed frozen and thawed by each of the FTSPs are extracted and used to define two unique probability distributions. Overlap area between the normalized distributions is also computed. This methodology is used to enumerate the degree to which land surface temperature products distinguish FT states similar to those derived from P-MW emissions, where perfectly unique FT state distributions are equal to 0 (no overlap), and identical distributions equal to 1 (completely overlapping). Additionally, the value ranges at which land surface temperatures diverge in FT classification are also identified. These temperature values are determined as those in which 1% of the data has been classified below a given temperature as thawed, and above as frozen. This is done to summarize critical distribution characteristics, as the 1% threshold is employed to remove outliers in which values at very low or high temperatures are classified as both frozen and thawed. Similarly, the Two-sample Kolmogorov-Smirnov test is also employed to verify the relationship between the FT distributions. The test statistic is calculated using the greatest difference between the frozen and thawed cumulative distribution functions, where larger values indicate increasingly different distributions (*Massey 1951; Miller 1956*).

### *SMAP-FT False Freezes*

Errors are expected to occur when high temperatures are associated with frozen classifications due to topographic effects, temporal or spatial resolution mismatch, temperature uncertainties, and FT algorithm limitations. However, known challenges in establishing robust FT reference metrics in SMAP-FT Version 1 products can cause a substantial reduction in product performance (Lyu *et al.* 2018; Dunbar *et al.* 2014-2020). Here, to investigate this, the location and time of occurrence of high temperature freeze classifications is examined. In July and August, characterized by minimum temperatures well above the freezing point across the domain, we identify regions in which freeze classifications were likely false (**Figure 4**).



**Figure 4** SMAP-FT summer freeze classifications

These regions correspond to areas with poor agreement in comparison metrics for SMAP-FT explored later in this work. Outside of high elevation regions with persistent snow cover (many classifications west of 110 °W), areas indicating frequent frozen classifications, including areas in the Upper Midwest, Maine, Central and Eastern Canada, can be classified as false freezes by SMAP-FT. It is likely that dry conditions and the relatively

dense presence of surface water in some of these regions result in similar FT reference states. FTSPs in such regions will need to be improved moving forward to correct for false freeze classifications, especially in the mid-latitudes. Algorithm updates have been made in the recent SMAP Version 2 & 3 FT products in an attempt to improve upon this and extend to global product coverage (*Xu et al. 2018*).

### 2.1.3 Results and Discussion

#### *Classification Accuracy*

First, classification accuracy (CA) of FT products with respect to various temperature datasets is explored during the period April 2015 – December 2016 (**Table 3**). Overall, FTSPs agreed the best with  $T_{2m-air}$  (CA = 81–91%) and show the least agreement with  $T_{N-skin}$  (76–85%). Conversely,  $T_{soil0-10cm}$  shows a comparatively stronger relationship with SMAP FTSPs than with FT-ESDR records, likely due to the deeper emission depth of SMAP L-band retrievals. MODIS  $T_{M-skin}$ -based FT classifications show intermediate agreement with all products and are slightly improved relative to NLDAS skin temperatures ( $T_{N-skin}$ ). Also, increases in CA from SMAP-FT to SMAP<sub>E</sub>-FT ranged from 1.7-2.4% across all comparisons, suggesting higher resolution FTSPs match more closely with temperature implied states. Also of note, all SMAP FTSPs showed a clear reduction (~8%) in CA as compared to FT-ESDR products. The causes of this are investigated further by examining seasonal performance, proportional frozen percentage, and corresponding frozen and thawed temperature distributions in the following sections. This reduction may also be explained through the known prevalence of summer false freezes.

**Table 3 Classification accuracy (CA) results for period of April 2015 – December 2016. Results are presented as totals throughout the study period and divided seasonally**

Freeze/Thaw Product	Comparison Variable	Classification Accuracy (April 2015 - 2016)	2015				2016		
			AMJ	JAS	OND	JFM	AMJ	JAS	OND
SMAP <sub>E</sub> -FT	NLDAS 2m Air Temperature	0.835	0.84	0.88	0.77	0.80	0.85	0.90	0.81
	NLDAS 0-10cm Avg Soil Temperature	0.822	0.81	0.88	0.78	0.75	0.83	0.90	0.82
	MODIS Skin Temperature	0.786	0.76	0.86	0.70	0.80	0.78	0.87	0.73
	NLDAS Skin Temperature	0.777	0.80	0.86	0.67	0.72	0.82	0.88	0.69
SMAP-FT	NLDAS 2m Air Temperature	0.811	0.81	0.83	0.75	0.80	0.83	0.85	0.81
	NLDAS 0-10cm Avg Soil Temperature	0.800	0.79	0.83	0.76	0.75	0.81	0.85	0.82
	MODIS Skin Temperature	0.766	0.74	0.81	0.69	0.80	0.77	0.83	0.73
	NLDAS Skin Temperature	0.760	0.78	0.82	0.67	0.72	0.80	0.84	0.70
AMSR-FT	NLDAS 2m Air Temperature	0.913	0.94	1.00	0.80	0.86	0.94	1.00	0.86
	NLDAS 0-10cm Avg Soil Temperature	0.874	0.89	1.00	0.72	0.79	0.91	1.00	0.82
	MODIS Skin Temperature	0.860	0.84	0.98	0.69	0.86	0.86	0.97	0.75
	NLDAS Skin Temperature	0.849	0.84	0.97	0.71	0.82	0.86	0.97	0.78
SSM <sub>I</sub> -FT	NLDAS 2m Air Temperature	0.898	0.94	0.99	0.75	0.84	0.94	1.00	0.82
	NLDAS 0-10cm Avg Soil Temperature	0.874	0.90	1.00	0.71	0.80	0.91	1.00	0.80
	MODIS Skin Temperature	0.864	0.83	0.98	0.71	0.87	0.85	0.97	0.76
	NLDAS Skin Temperature	0.841	0.90	0.99	0.63	0.78	0.90	0.99	0.70

Seasonal analyses provide information on the dynamics of FTSPs during FT onset as well as during the winter frozen and summer thaw periods. During summer (JAS), when temperatures are largely well above freezing, all temperature products perform best at classifying the FT state (>83% SMAP-FT, and ~100% in FT-ESDR). As the fall transitions (OND) to winter freeze (JFM), NLDAS air (75-81%) and top layer soil temperatures (76-82%) become the best surrogates to SMAP FT states, as characterized by a CA increase relative to skin temperatures of 6-13% during the fall transition period. While this relationship also holds for Ka-band FT-ESDR products, it is less pronounced (0-12%). Additionally, during the spring thaw period (AMJ), slight increases on the order of 3-12% in CA are observed across most members compared to winter period, excluding  $T_{M-skin}$ . All comparison members show a relative increase in CA when comparing the 2015 freeze onset to the 2016 spring period (5-17%). This indicates improved agreement between temperature products and FTSPs when identifying FT conditions during the

thaw period (74-94%) as compared to the freeze onset (63-86%). Poor relative agreement with skin temperatures is likely a result of sensor sensitivity deeper than the surface skin temperatures, in addition to potential NLDAS model deficiencies in representing temperatures of complex snow-covered surfaces. In support of this, FTSPs are shown to have better agreement with MODIS  $T_{M\text{-skin}}$  than modeled NLDAS  $T_{N\text{-skin}}$ . Small variations in CA between AMSR and SSM/I-FT can likely be attributed to the varied overpass times and is explored further in the coming sections.

Large increases in classification accuracy are also observed during the primary freeze onset period (OND) from 2015 (63-80%) to 2016 (69-86%). The presence of near record high temperatures in December 2015 across the domain, likely resulting in more ephemeral FT-events, may be the cause (*NCEI 2019*). Depending on the characteristics of freeze and thaw onset as well as snow cover variability, CA is expected to vary annually. An uptick in accuracy over nearly all CA proportions occurs during the mid-winter period (JFM), as classified FT state and surface temperatures are continuously frozen across much of the domain. This is true for all but SMAP FTSP comparisons with soil temperature, in which the thermal insulation properties of snow cover should be considered as SMAP exhibits a reduced frozen area relative to FT-ESDR. During spring thaw (AMJ),  $T_{2m\text{-air}}$  appears to be the best surrogate to P-MW measured FT state. As P-MW based RS techniques begin to observe water in the landscape, FT products appear to closely resemble  $T_{2m\text{-air}}$  as air temperature tends to define surface melt onset. This is because soil and skin temperatures can reflect the temperatures of snow insulated soil and surface snow temperatures prior to the disappearance of the snowpack, making these

products less likely to indicate the presence of ephemeral thawed conditions as air temperatures creep above freezing (*Lyu et al. 2018*).

Maps of classification accuracy illustrate spatial variability in the relationship between FTSPs and surface temperatures (**Figures 5 - 8**). FT-ESDR products show comparable accuracy across the domain with higher accuracies in homogenous regions without prevalent surface water features and complex topography. FT-ESDR products show highest CA in regions downwind of the Rocky Mountains and slightly reduced CA in eastern portions of the domain with more vegetation. However, comparisons of FT-ESDR products to MODIS skin temperature show increased CA over mountainous regions as compared to NLDAS parameters ( $T_{N\text{-skin}}$  and  $T_{\text{soil}0\text{-}10\text{cm}}$ , **Figures 6 - 8**). SMAP FTSPs illustrate similar dynamics to an exaggerated degree, as decreased CA values are prevalent in the Rocky and Cascade Mountain regions. In contrast to FT-ESDR products, SMAP FTSPs exhibit the strongest relationship to temperature variables in central Canada, which is a cold region with increased moisture and abundant vegetation relative to west-central Canada (west of 97 °N). The reductions in CA across portions of western and eastern Canada, the Pacific Northwest, central North Dakota, and Maine are primarily attributed to false freeze flags.

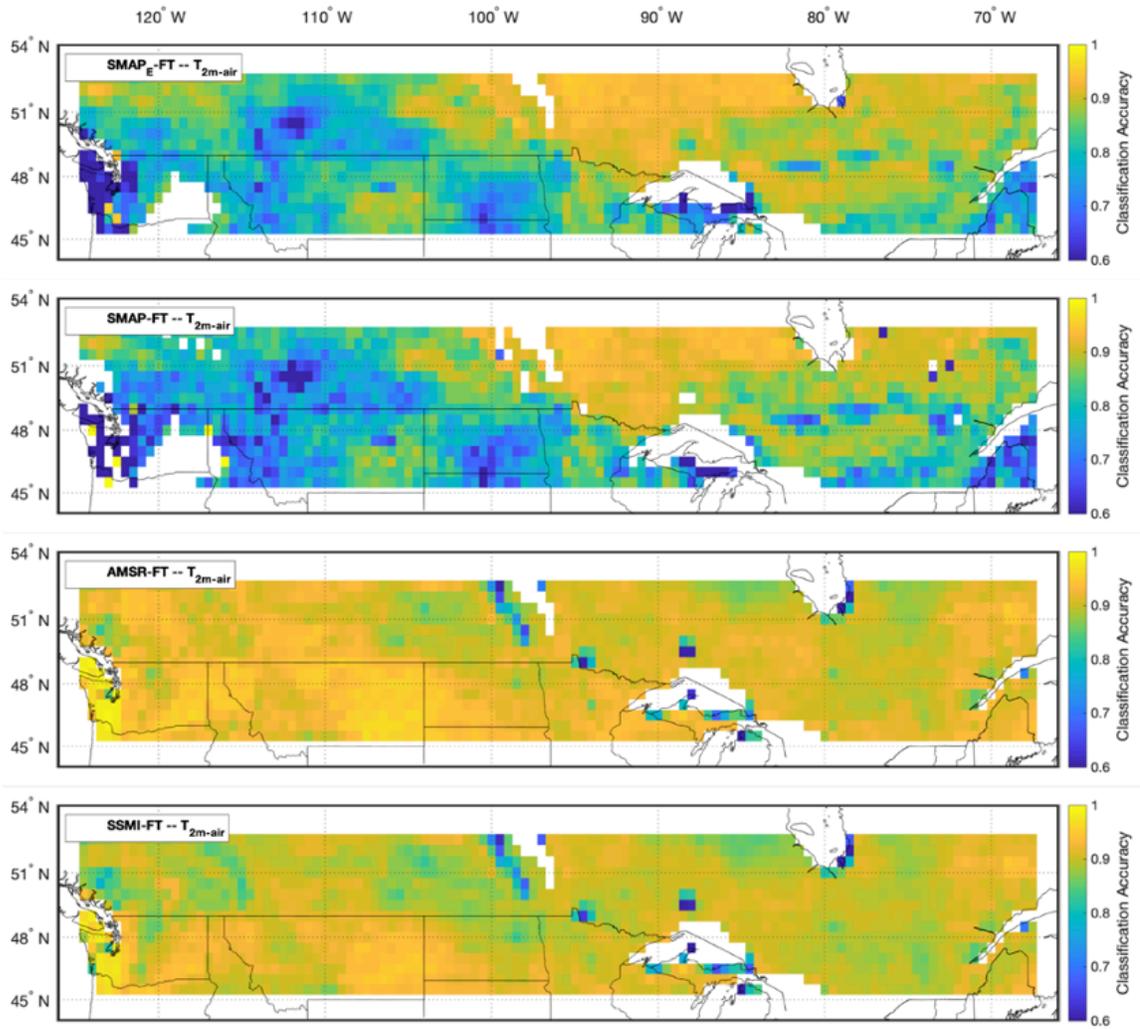


Figure 5 Gridded (0.5°) classification accuracy of freeze/thaw products compared to NLDAS-2 2-meter air temperature (April 2015 – December 2016). White areas signify no data or surface water

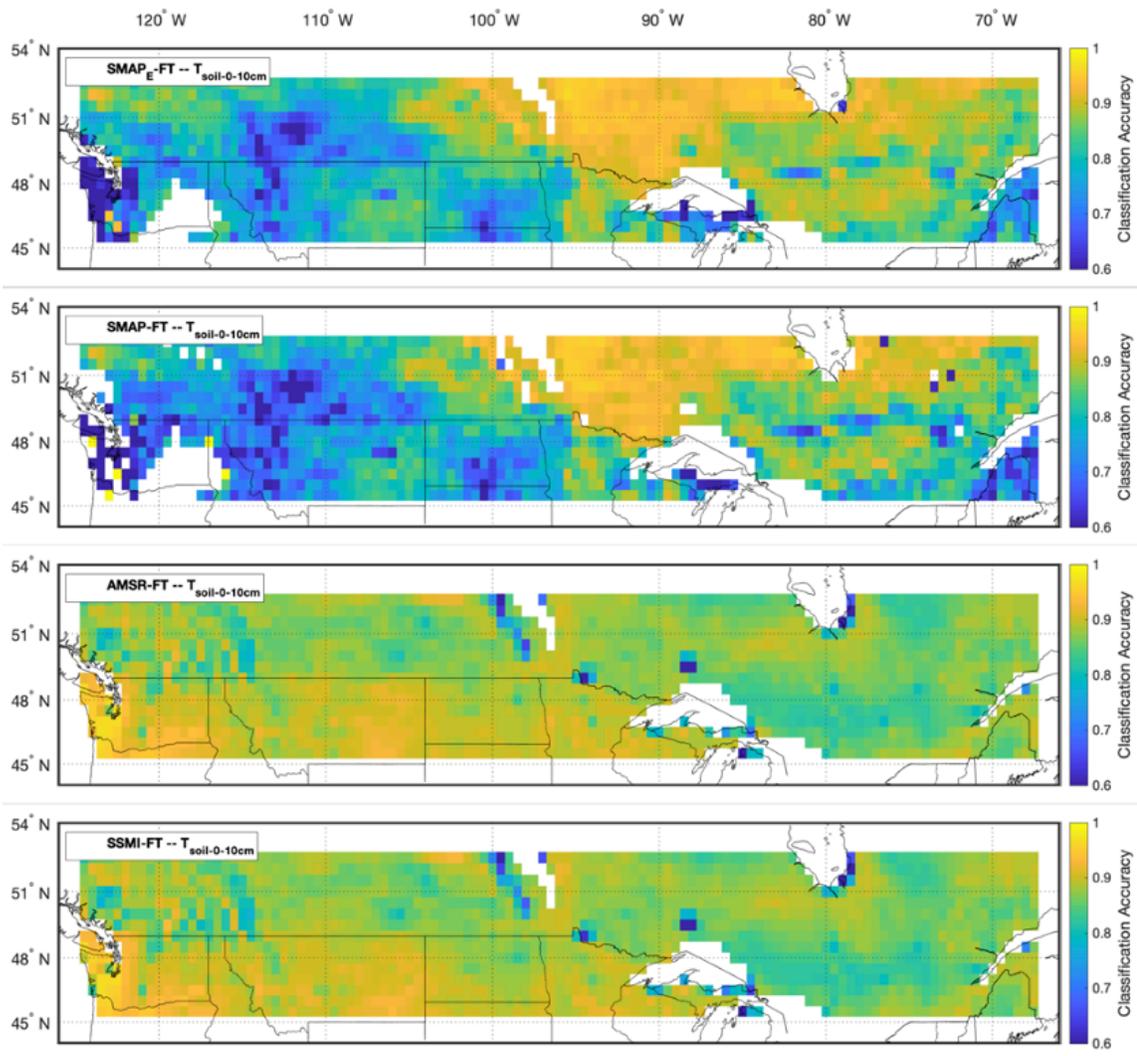


Figure 6 Same as Figure 5, but for NLDAS-2 0-10 cm soil layer temperature

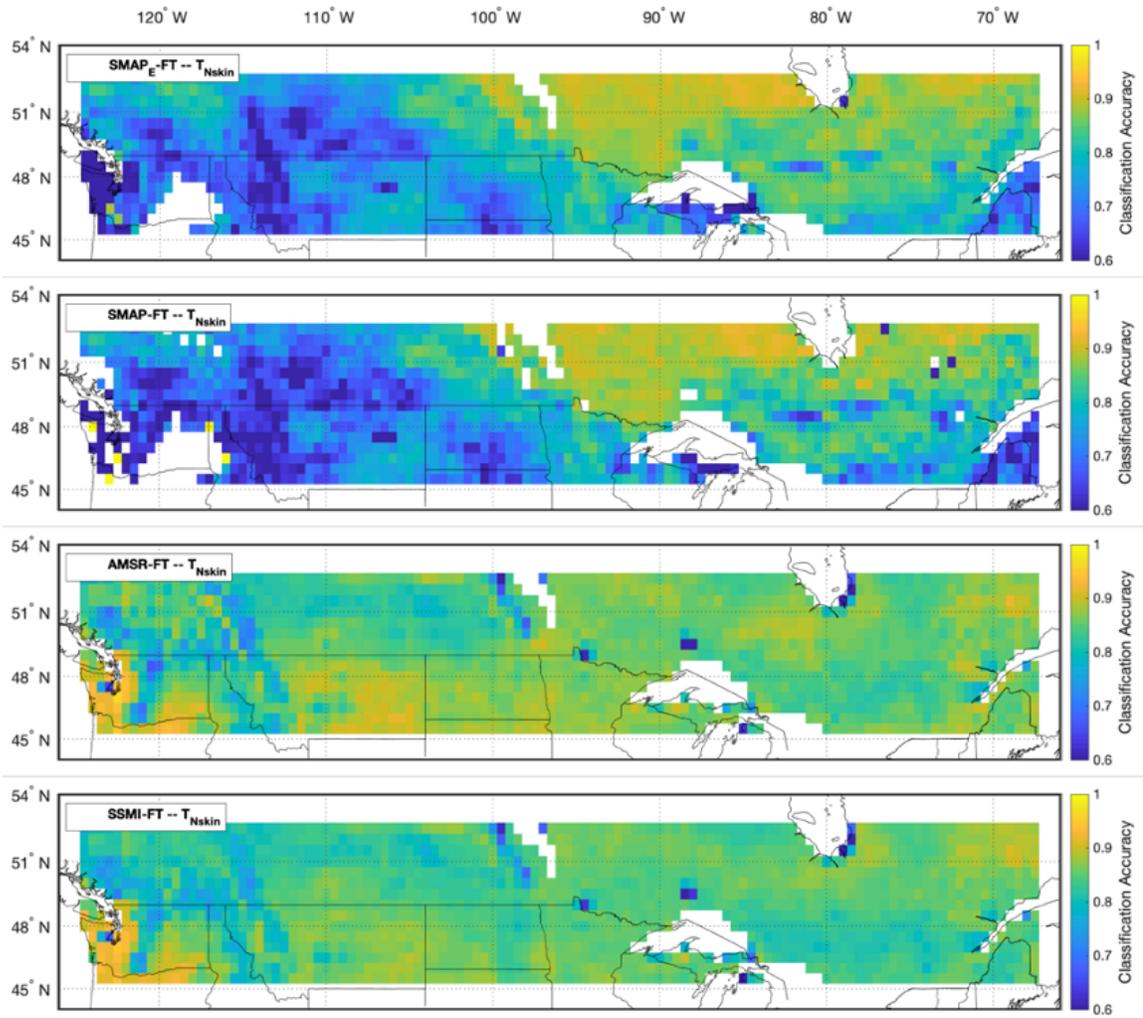


Figure 7 Same as Figure 5, but for NLDAS-2 skin temperature

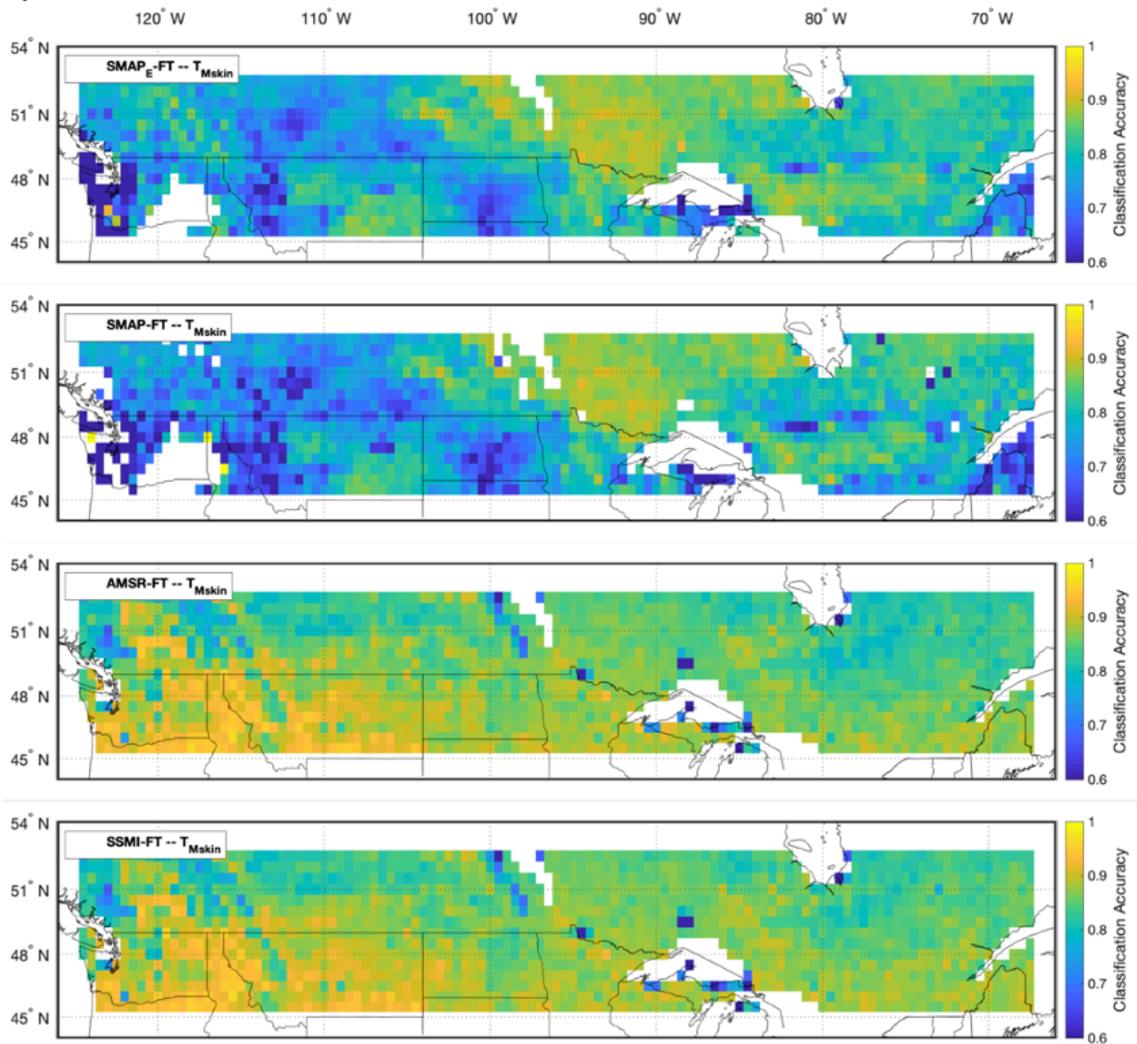


Figure 8 Same as Figure 5, but for MODIS skin temperature

### *Proportional Differencing*

#### *Regional Proportional Summary Results*

Two statistical measures are computed for each FT/temperature comparison:

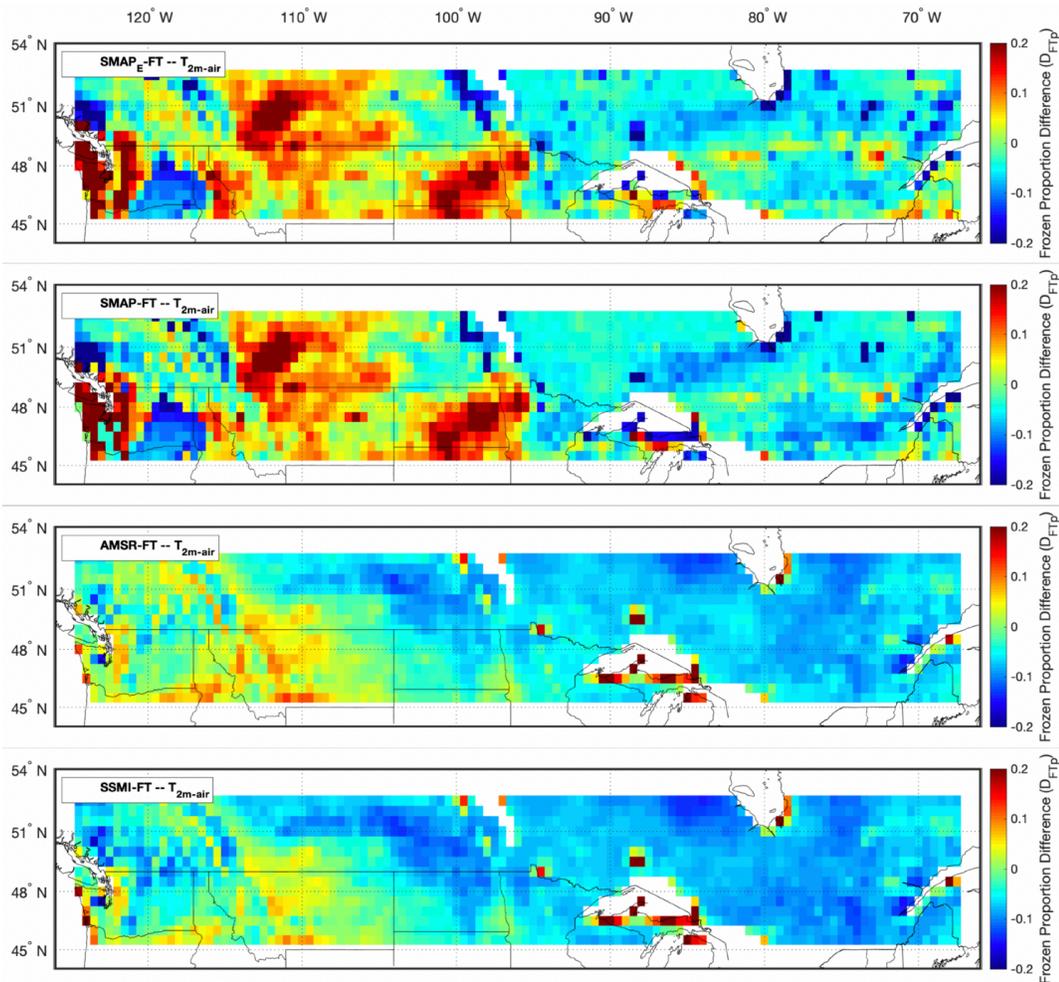
- 1) proportional mean ( $D_{FTp}$ ): representing FTSPs tending to overestimate (+) or underestimate (-) the temperature threshold-based frozen proportion; and
- 2) absolute  $D_{FTp}$ : used to represent the magnitude of overall average difference

**Table 4 Proportional differencing summary values. Computed as average across all pixels and complete time period (April 2015 – 2016); proportional mean ( $D_{FTp}$ ) and absolute proportional mean ( $D_{FTp-abs}$ ), (Equation 8)**

Proportional Differencing Results ( $D_{FTp}$ )			
Freeze/Thaw Product	(-) Temperature Product	April 2015 - 2016	
		$D_{FTp}$	$D_{FTp-abs}$
SMAP <sub>E</sub> -FT	NLDAS 2m Air Temperature	0.007	0.189
	NLDAS 0-10cm Avg Soil Temperature	-0.026	0.204
	MODIS Skin Temperature	-0.087	0.247
	NLDAS Skin Temperature	-0.088	0.247
SMAP-FT	NLDAS 2m Air Temperature	-0.004	0.179
	NLDAS 0-10cm Avg Soil Temperature	-0.036	0.195
	NLDAS Skin Temperature	-0.099	0.240
	MODIS Skin Temperature	-0.098	0.243
AMSR-FT	NLDAS 2m Air Temperature	-0.035	0.089
	NLDAS 0-10cm Avg Soil Temperature	-0.083	0.131
	NLDAS Skin Temperature	-0.129	0.155
	MODIS Skin Temperature	-0.141	0.163
SSMI-FT	NLDAS 2m Air Temperature	-0.050	0.104
	NLDAS 0-10cm Avg Soil Temperature	-0.082	0.130
	MODIS Skin Temperature	-0.132	0.159
	NLDAS Skin Temperature	-0.130	0.159

The absolute average results mirror what was found in the CA analysis, with all FT products most closely comparable to  $T_{2m-air}$  (**Table 4**). Overall, the results indicate an average frozen proportion mismatch across all grid cells ranging from 8.9% (AMSR-FT,  $T_{2m-air}$ ) to 24.7% (SMAP<sub>E</sub>-FT,  $T_{N/M-skin}$ ). In contrast to CA results, the 9 km SMAP<sub>E</sub>-FT has an increased average magnitude of disagreement as compared to the 36 km SMAP-FT. Presumably, the use of significantly more data points introduces increased variability in  $D_{FTp}$  between overpasses. Also notable, is nearly all  $D_{FTp}$  comparison values shown in **Table 4** are negative, indicating that FT products defined less frozen area as compared to temperature products. The use of a lower temperature threshold ( $< 0$  °C) would result in an increased  $D_{FTp}$  by reducing the frozen extent accordingly. Still, when using a threshold at 0 °C, the only case in which  $D_{FTp}$  was positive occurred when comparing SMAP<sub>E</sub>-FT to  $T_{2m-air}$  (**Table 4**). These results support the ability of P-MW techniques to capture phase changes in the surface that do not generally occur until temperatures drop below freezing

for an extended period. Documented NLDAS cold biases may also contribute (*Xia et al. 2012b, c*). The prevalence of summer false freeze classifications in SMAP FTSPs can also affect this tendency, as spatial plots (**Figures 9 – 12**) show overestimated regions (in red) skew the mean results. Even with SMAP-FT mid-summer false freezes, FTSPs used in this study are much more likely to define a lesser frozen extent than temperature threshold-based FT classifications, especially in the eastern portion of the domain.



**Figure 9** Mean proportional differences ( $D_{FTp}$ ) maps of FT products to NLDAS 2m Air Temperature for  $SMAP_E - FT$ ,  $SMAP - FT$ ,  $AMSR - FT$ , and  $SSMI - FT$

FT-ESDR products tend towards underestimation of temperature-based frozen proportions across a wider area of the domain (compared to SMAP FTSPs), with the closest FT to temperature agreement occurring in the western plains. FT products follow similar patterns of variation in  $D_{FTp}$  regionally (**Figures 9 – 12**), as compared to CA results (**Figures 5 – 8**) when averaged over the study period. However, regional causes of reduced CA are highlighted by FTSPs exhibiting less (blue) or more (red) frozen classification than temperature products. In central Canada, FTSPs show increased variability as compared to surface temperatures with  $D_{FTp}$  ranging from around -0.1 (FT-ESDR) or -0.05 to 0.05 (SMAP-FT). Eastern regions tend towards underestimation of temperature defined frozen area, whereas more mountainous regions display increased heterogeneity in classification  $D_{FTp}$  (**Figures 9 - 12**). This result can be expected as a byproduct of imperfect modeled temperature parameters, limited spatial resolution affecting the ability to resolve smaller scale topographic features, FT sensitivities to other surface components (i.e., moisture), and known deficiencies in FTSPs at observing FT dynamics across complex topography (*Kim et al. 2017; Derksen et al. 2017*). A reduction in  $D_{FTp}$  from SMAP-FT to SMAP<sub>E</sub>-FT in portions of the domain indicate improvements in the enhanced SMAP product's ability to resolve complex features (except in coastal areas). Interestingly, there appears to be extensive regions where  $F_{FTp}$  exceeds  $F_{Tp}$  focused on the leeward side of the mountains, which tend to be inherently drier. While they do share similarities, these results verify clear spatial inconsistency across FTSPs as compared to various land surface temperature variables. Even so, the small magnitude of proportional difference suggests that the satellite

observed FT state is closely linked to these variables, especially air temperature. While also suggesting that as skin temperatures may provide an indication of the underlying FT state, microwave-based approaches remain more sensitive to temperature changes beyond the surface skin layer.

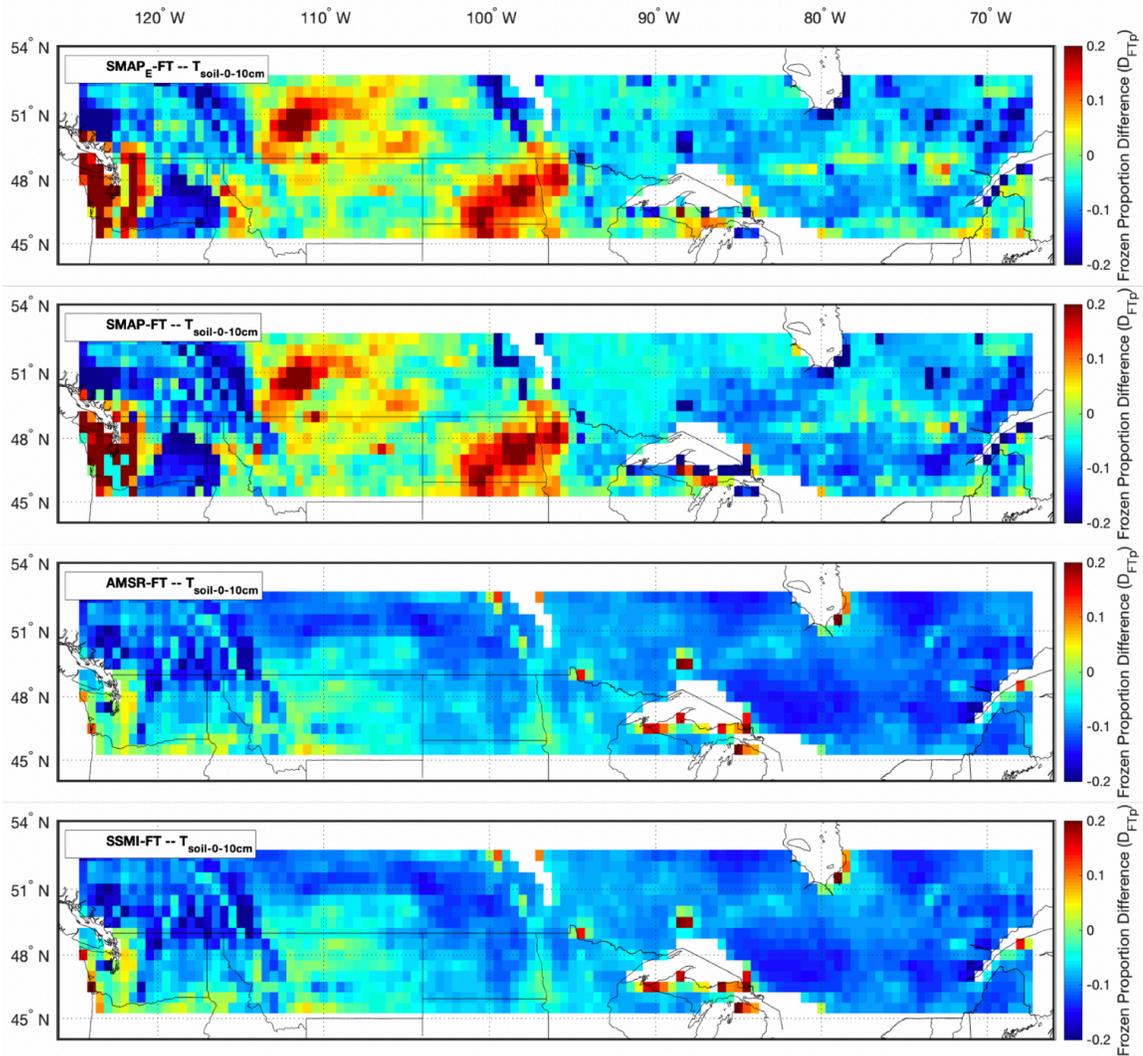


Figure 10 Same as Figure 9, but for NLDAS 0 – 10 cm depth soil layer temperatures

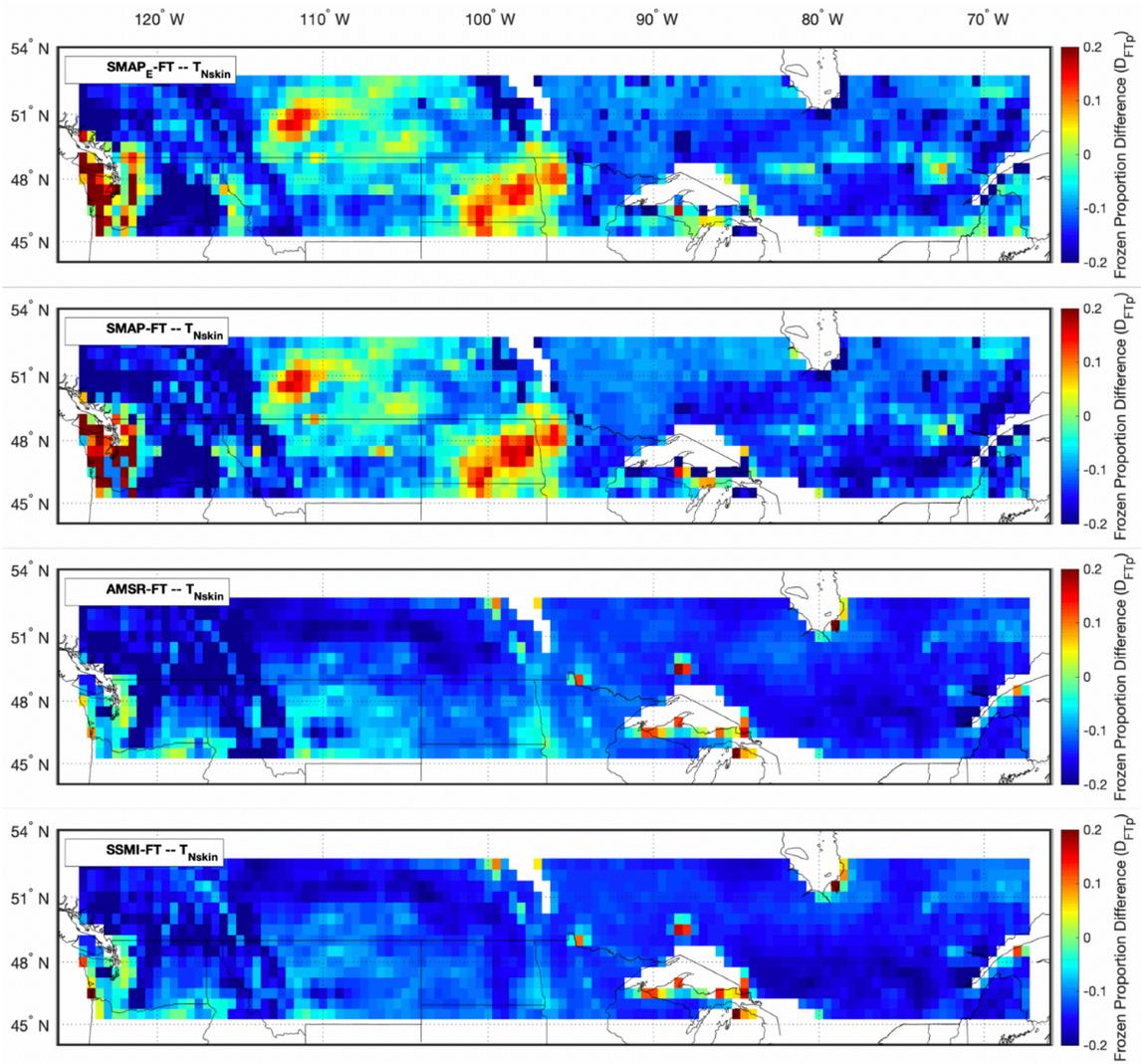


Figure 11 Same as Figure 9, but for NLDAS skin temperatures

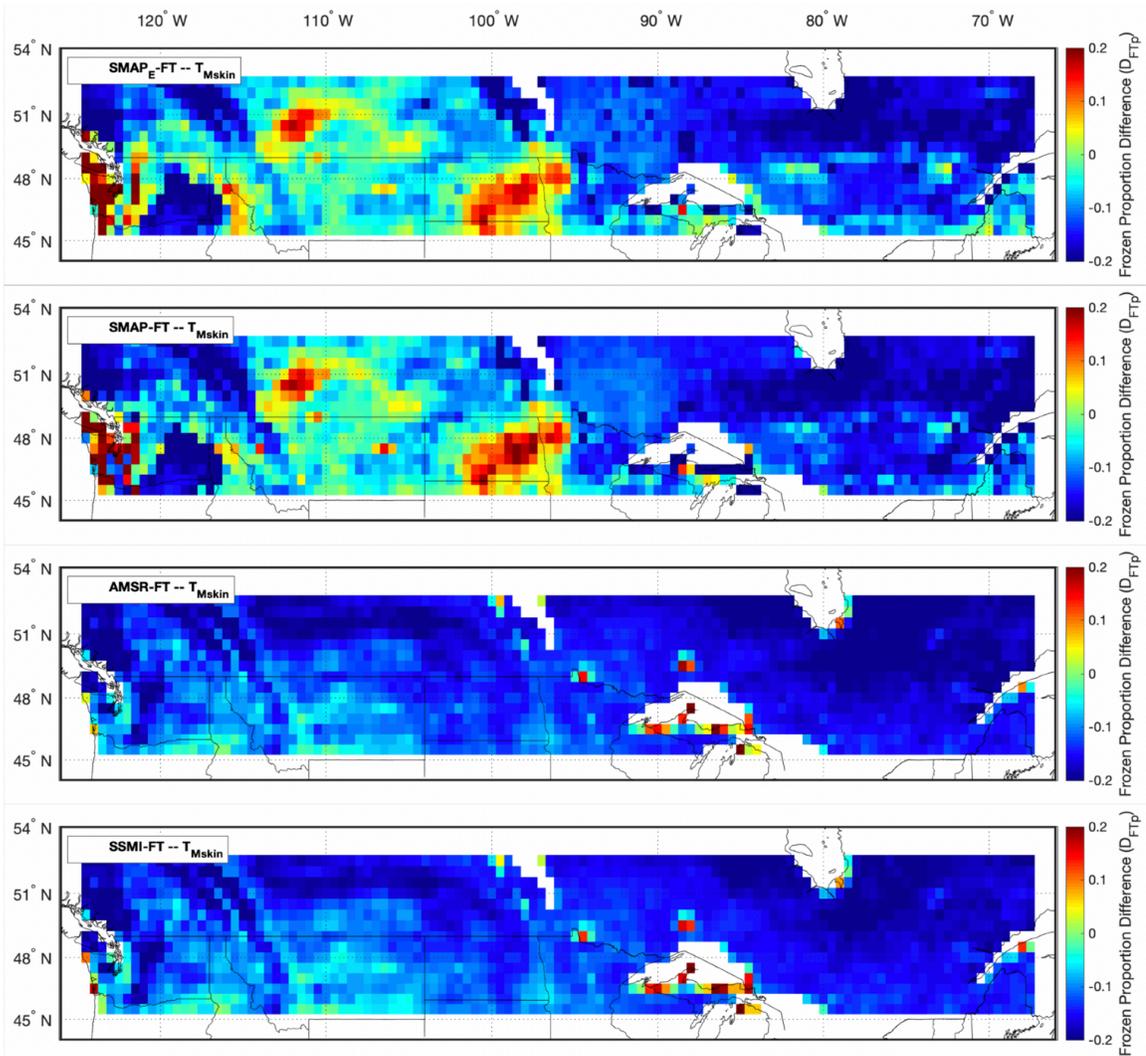
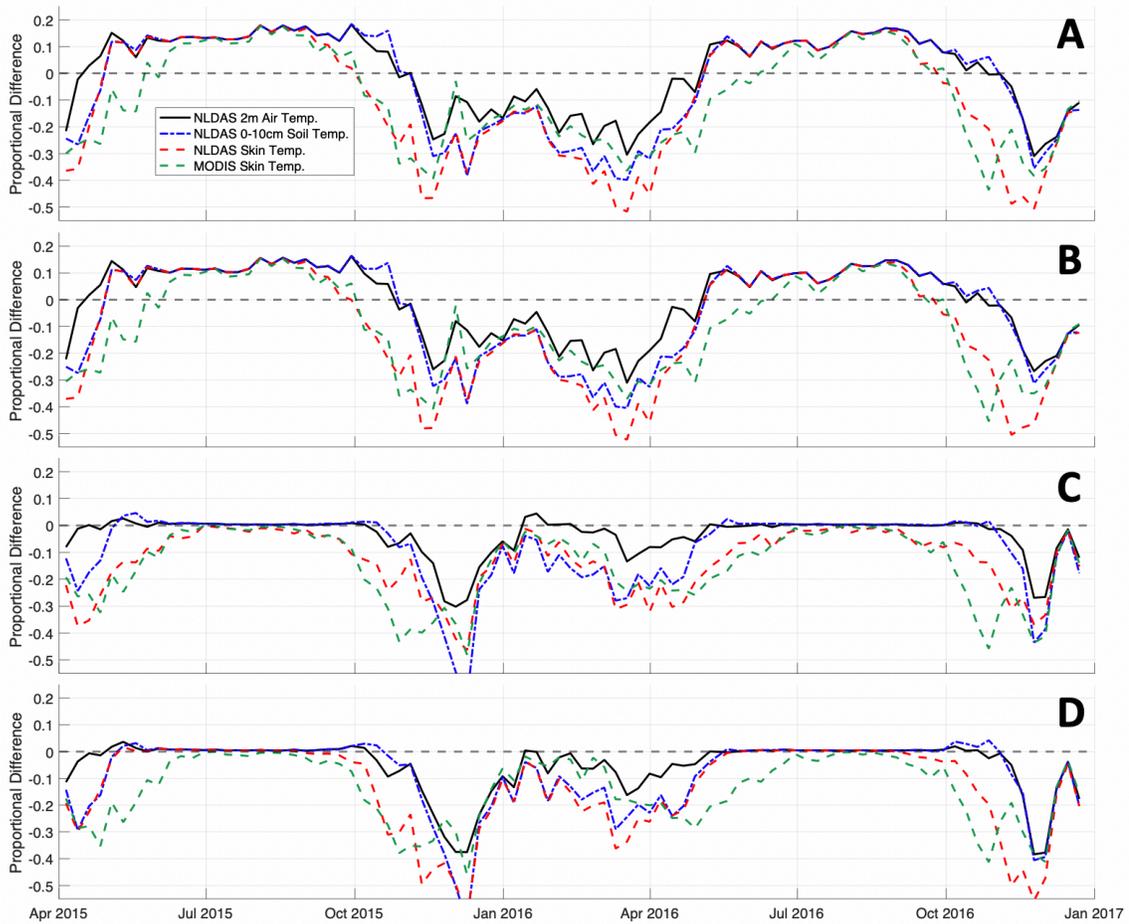


Figure 12 Same as Figure 9, but for MODIS skin temperatures

Time-series based analysis utilizing monthly  $T_{2m-air}$ ,  $T_{soil0-10cm}$ ,  $T_{N-skin}$ , and  $T_{M-skin}$  mean  $D_{FTP}$  values is shown in **Figure 13** to assess changes in agreement through time. Temperature-based frozen proportion values show the strongest relationship with FTSPs May through October. As FT-ESDR products closely resemble NLDAS air and soil temperatures during this period and  $D_{FTP}$  values fall well within the -0.1 to 0.1 range. Skin temperatures tend to lead air and soil temperature in estimating increased frozen extent

(negative values) relative to FTSPs by September due to the onset of sub-zero temperatures occurring before soil freeze up. Additionally, FTSP defined frozen extent is exceeded by NLDAS  $T_{N-skin}$  throughout much of the study period as MODIS skin temperature shows comparatively increased agreement with FTSPs through transitional periods and the January-February extended frozen period (**Figure 13**). For SSM/I-FT, MODIS skin temperature has a similar relationship to FTSPs as air temperature during the mid-winter period. As discussed, FTSPs exhibit reduced agreement to temperature products during freeze and thaw onset tending to underestimate temperature derived frozen proportions. Potential false freezes in SMAP FTSPs exhibit the opposite, as mean  $D_{FTp}$  values show SMAP overestimating temperature-based frozen proportions even in mid-summer. Summer frozen areas in SMAP products are contributors to the increased absolute mean  $D_{FTp}$  values and reduced CA over the summer period relative to ESDR FT products. However, agreement between temperature and FT products continues to be at its lowest during the shoulder seasons. This is likely due to increased ephemeral and diurnal FT events which tend to occur across the domain in spring and fall, making it more difficult to capture clear FT signals.

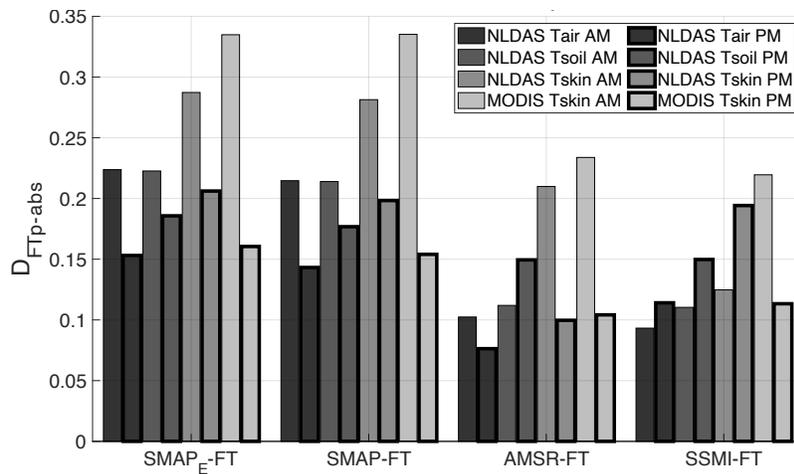


**Figure 13** Mean frozen proportion differences ( $D_{FTp}$ ) of MODIS skin, NLDAS 2m air, NLDAS 0-10 cm soil, and NLDAS skin temperatures as compared to FT products over the northern CONUS, southern Canada domain. (A) SMAP<sub>E</sub>-FT, (B) SMAP-FT, (C) AMSR-FT, and (D) SSM/I – FT

### *Comparing Morning and Evening Overpasses*

All FTSPs evaluated as part of this study are derived from both descending (AM) and ascending (PM) data from twice-daily satellite overpasses (**Table 1**). Previous studies have identified significant variations in the relationship between FT classifications and temperature variables dependent on overpass time (*Kim et al. 2011; Lyu et al. 2018*). Characteristically, morning overpasses (completed at either 1:30 a.m. or 6:00 a.m. local

solar time) have cooler temperatures and little solar exposure as compared to afternoon overpasses at 1:30 p.m. or 6:00 p.m. As a result, AM FT classifications have proven more susceptible to mismatches with surface temperature due to diurnal freeze events in which surface temperatures may creep below freezing while satellite FT retrievals continue to measure surface emissions of a thawed landscape, or vice versa.



**Figure 14 Absolute mean proportional differences ( $D_{FT^p-abs}$ ) separated into morning (AM) and afternoon (PM) periods for all comparisons. Unbounded bars indicate AM overpasses and unbounded PM**

As presented in **Figure 14**, the general rule holds true as absolute mean proportional differences are significantly reduced from morning (AM) to afternoon (PM) retrievals across nearly all comparison members. This is most notable in SMAP FTSPs and comparisons involving  $T_{M-skin}$ . The exceptions include SSM/I to  $T_{2m-air}$ ,  $T_{soil0-10cm}$ ,  $T_{N-skin}$  and AMSR to  $T_{soil0-10cm}$ . As shown across all FTSPs to  $T_{soil0-10cm}$  comparisons, the relatively small reductions or increases in absolute  $D_{FT^p}$  can be attributed to increased soil insulation capacity relative to air. As a result, soil temperature tends to be less susceptible to slight

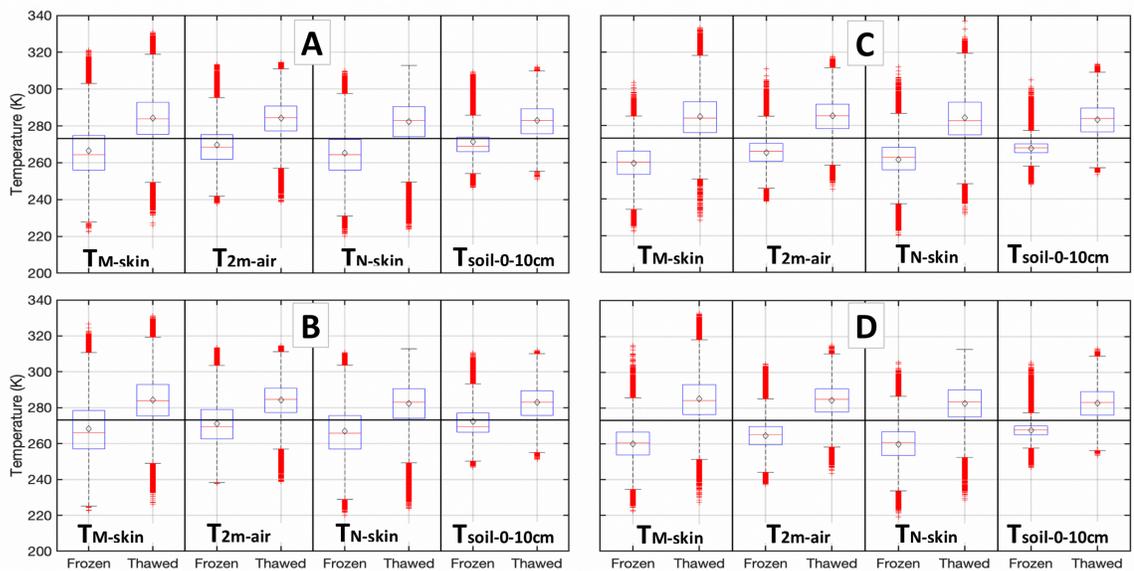
temperature variations and avoids falling below freezing during ephemeral freeze events in which FT classifications remain thawed. Conversely, soil can also retain colder temperatures as surface thaws, as the presence of wet surfaces may result in thawed P-MW based FT classifications while soil remains frozen at depth. The most significant reduction in absolute  $D_{FTp}$  occurs in all FTSP comparisons with MODIS skin temperature. This occurs as a result of MODIS defining far more frozen extent in the morning as compared to afternoon observations relative to FTSPs.

There are several complex dynamics driving the FT classifications of FTSPs, including surface characteristics, sensor configurations, and varied overpass times. During transitional periods where ephemeral FT events are common, slight temperature differences can lead to large differences in temperature threshold-based FT classifications. This is especially pronounced when dealing with varied retrieval times of different FTSPs. As such, precise time-targeted approaches would be valuable to draw improved information on FTSP-temperature relationships across different satellite-based FT products.

### *Exploring Freeze/Thaw Temperature Distributions*

Following the association of NLDAS and MODIS observed temperatures with FT states, distributions are created for each temperature variable (**Figure 15**). Boxplots show clear variations between frozen and thawed classes across all temperature variables. In SMAP derived products, a clear overlap is observed between FT distributions well above and below the freezing point of 273 °K (0 °C). However, temperature distributions derived from FT-ESDR do not show overlap between the central 50% of the data in any case. Much

of this can be attributed to frequent SMAP FT classifications of regions as frozen while surface temperature variables indicate values  $>300\text{ °K}$  ( $>27\text{ °C}$ ). Slight reductions in these high temperature frozen classifications were observed when using SMAP<sub>E</sub>-FT (9 km) as compared to SMAP-FT (36 km). As a result, slight increases were observed in separation between FT distribution values as compared to SMAP-FT reflected by a shift of frozen distribution median values by 3 to 5 °K. This increased separation between distributions supports the idea that SMAP<sub>E</sub>-FT shows improvement at accurately distinguishing surface FT states relative to SMAP-FT.



**Figure 15** Boxplot distributions of frozen and thawed states as associated with various temperature products. (A) SMAP<sub>E</sub>-FT, (B) SMAP-FT, (C) AMSR-FT, and (D) SSM/I-FT

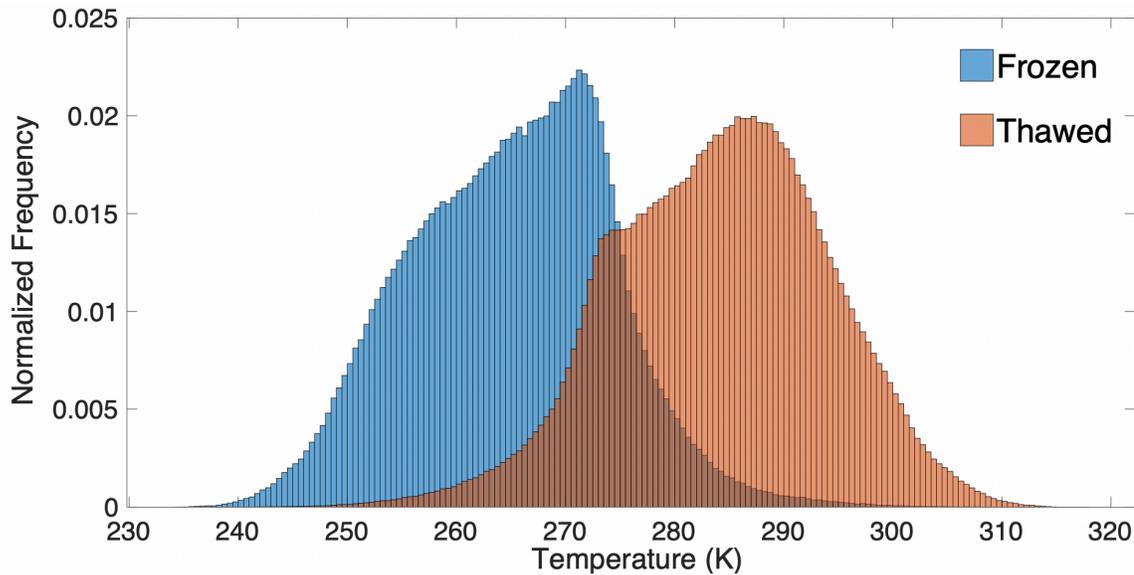
By calculating the proportion of overlap between normalized temperature distributions (or probability distribution functions, PDFs), we are able to quantify which associated temperature values defined the most unique frozen and thawed distributions

relative to various FTSPs. A minimal overlap signifies a comparison in which the frozen and thawed states were most clearly discerned by temperature variables (**Table 5**). As may be expected, all PDFs show highest values of overlap and associated classification uncertainty at the freezing point generally decreasing logarithmically away from this point in both directions (**Figure 16**). The resulting overlap proportions help to verify that none of the temperature variables defined particularly unique FT distributions when calculated with SMAP products (>20% overlap). This is compared to overlap values as low as 9% when comparing NLDAS  $T_{2m-air}$  to associated AMSR FT classifications. Additionally, in order to summarize the characteristics of each PDF, the temperature at which 1% of data is classified below as thawed ('Low') and above as frozen ('High') are presented in **Table 5**. The Two-sample Kolmogorov-Smirnov Test is performed as an additional method to determine the relationship between the two temperature distributions. A larger test statistic indicates that the two distributions are increasingly different, or more well separated. The resulting outputs are shown to mirror the PDF overlap proportions, providing verification in identifying which FTSP-temperature combinations provide the most unique frozen and thawed temperature classifications.

**Table 5 Mean temperature of frozen and thawed distributions, distribution overlap proportions, Kolmogorov-Smirnov test statistics, and other distribution summary characteristics**

Comparison Members			Distribution Mean Temperature (°K)		Distribution Overlap Proportion	Temperature (°K) of Uncertainty: ~1% of data classified below Low (as thawed) and above High (as frozen)		Two sample Kolmogorov Smirnov Test Statistic
			Frozen	Thawed		Low	High	
Freeze/Thaw Product	SMAP-FT	NLDAS-ST 0-10cm	271	283	0.20	266	296	0.596
		NLDAS-2m	270	284	0.20	264	297	0.593
		NLDAS-SkinT	265	282	0.21	256	297	0.574
		MODIS-SkinT	266	284	0.23	257	307	0.539
	SMAP-FT	NLDAS-ST 0-10cm	272	283	0.23	266	297	0.541
		NLDAS-2m	271	284	0.23	263	299	0.541
		NLDAS-SkinT	267	282	0.24	255	299	0.523
		MODIS-SkinT	268	284	0.26	256	308	0.485
	AMSR-FT	NLDAS-2m	265	285	0.09	268	281	0.820
		MODIS-SkinT	260	285	0.10	260	279	0.803
		NLDAS-SkinT	261	284	0.10	263	278	0.794
		NLDAS-ST 0-10cm	268	283	0.11	266	279	0.779
	SMMI-FT	NLDAS-2m	265	284	0.10	266	281	0.802
		MODIS-SkinT	260	285	0.10	260	281	0.795
		NLDAS-ST 0-10cm	268	283	0.11	266	279	0.785
		NLDAS-SkinT	260	283	0.11	258	279	0.779

Notably, the comparisons showed very distinctive temperature distributions between both  $T_{2m-air}$  and  $T_{M-skin}$  when compared to FT-ESDR products, with overlap proportions ranged from 9.0-9.9% ( $T_{2m-air}$ ) to 9.9-10.3% ( $T_{M-skin}$ ). MODIS distributions were derived utilizing around 300,000 frozen and 900,000 thawed classifications. This amounted to about 1/3<sup>rd</sup> the total data-points as compared to NLDAS parameters, due to MODIS temperature data availability. Even so, the ability of  $T_{M-skin}$  to define clearly separate temperature distributions when associated with ESDR FT classifications can help justify the use of satellite-based thermal products in estimating and evaluating the accuracy of FT states, especially in remote regions. Using MODIS skin temperature in efforts to validate SMAP-FT products may not be as suitable. As temperature distributions were most poorly defined for SMAP products, 21.3-25.7% overlap, when associated with skin temperatures.



**Figure 16** Probability distribution function histogram for AMSR-FT defined frozen and thawed states, matched to NLDAS air temperatures. Overlap indicated by dark shaded region

Characteristics of each FTSP-temperature distribution are uniquely representative of variations in each FT product over the domain. These characteristics are especially useful when evaluating whether skin temperature, soil temperature, or air temperature is most representative of current FTSPs along with associated measurement dynamics. The results show a tendency of P-MW FT classifications to observe the top few centimeters of the surface, as soil temperatures can remain below freezing, while the P-MW observation indicates surface thaw. This increase in thawed classifications below the freezing point for  $T_{\text{soil}0-10\text{cm}}$  can also be related to periods in which wet snow is prevalent at the surface, resulting in a thawed P-MW  $T_b$  observation, even as soil temperatures are modeled to be frozen. In reality this can also occur in the opposite direction as soils remain thawed even as surface temperatures indicate frozen conditions due to thermal insulation of the snow and underlying soils. This illustrates a weakness of using air temperature as a direct

indicator of FT as it can both ignore the deeper surface state in addition to P-MW methods not being directly sensitive to air temperature. Relative to soil, the comparatively wide ranges of skin and air temperatures in which frozen classifications occur (**Table 5**) support this assertion.

### ***Study Limitations and Relevance***

Satellite FT products have shown an improved agreement as compared to in-situ temperatures over more non-arid northern regions with homogenous land cover. This is in contrast to performance in mountainous regions, areas with frequent freezing and thawing transitions, those with heterogeneous vegetation cover, frequent large precipitation events, and in dry regions in which FT product reliability can be reduced (*Kimball et al. 2009; Kim et al. 2011*). Various landscape characteristics can contribute to erroneous FT classifications especially in regions with small differences between frozen and thawed reference states. This can be due to high densities of water bodies or over dry regions as substantiated by this study. It is also important to note that radiometer-based FT products tend to retrieve less frozen area as compared to active products such as SMAP and Aquarius radars (*Derksen et al. 2017*). While active FT RS methods are not assessed in this study, comparisons have been performed over Alaska utilizing L-, C-, and Ku-band radars showing a similar ability to detect to changes in the surface FT state (*Podest et al. 2014*).

Regarding study limitations, sparse coverage due to cloud cover of MODIS Aqua skin temperature retrievals increases comparison uncertainty by reducing the amount of evaluation data. Furthermore, as FT and temperature products are matched to a lower resolution grid, distortions of FT classifications due to classification rounding and

temperature averaging can induce further error. This use of different resolution products can increase uncertainty especially in mountainous regions where temperatures can vary significantly within a single grid pixel.

NLDAS skin temperature ( $T_{N\text{-skin}}$ ) was found to over represent the frequency of temperatures around 273 °K (0 °C) compared to other temperature variables. NLDAS modeled  $T_{N\text{-skin}}$  has minimal thermal inertia and is very sensitive to the longwave radiation balance. Therefore, it may cool (or warm) quickly and not be represented in the broader landscape FT state as observed using P-MW emissions. As  $T_{\text{soil0-10cm}}$  is related to surface skin temperature in NLDAS assimilation systems, the freezing point bias can also be identified in the resulting  $T_{\text{soil0-10cm}}$  values. An increased frequency of values around the freezing point are also associated with the increased energy required in phase changes, in which a surface can remain at freezing for an extended period of time during freeze up, or thaw. Due to NLDAS model physics (*Cosgrove et al. 2003; Xia et al. 2012b*) it is possible for the modeled temperatures to hover slightly above freezing as freeze up occurs before falling below 0 °C. When in reality, the majority of a surface may be frozen and deemed as such by FTSPs.

Methodology presented herein has the potential to be improved through the removal of clearly false classifications, examination of an extended spatial and temporal record, the utilization of additional temperature datasets (geostationary satellites and in-situ networks), the assessment of microwave backscatter observations, and precise time matching techniques over an extended spatial and temporal domain. Still, by utilizing regional NLDAS and MODIS temperature data, we are able to identify spatial dynamics across FT

products that are difficult to examine through the use of limited point measurements and core validation sites. Moving forward, it will be critical to further assess SMAP L-band FT products as aspects of this analysis showed variations in the relationship to surface temperatures compared to Ka-band retrievals. Namely, increased similarity to soil temperatures relative to NLDAS in areas void of frequent summer false freezes. As likely false freeze classifications are removed in the future, relationships of SMAP FTSPs to land surface variables should become even clearer. The need to further investigate thermal profiles as they relate to regions with snow cover, precipitation, and varied landcover types will help to improve understanding of FT dynamics under various conditions as they relate to microwave RS techniques. Additionally, inter-grid cell heterogeneity in FT state is often not well captured at these product resolutions (~10 km +). The implementation of fractional or higher-resolution FT products will likely improve the representation of FT state across the global domain, especially during freeze and thaw onset. While there are limitations to the methodology utilized in this study, the resulting classification accuracy, proportional differencing, and FT-temperature associated distribution analyses have proven insightful in assessing the relationship between FTSPs and surface temperature. These methods provide a comprehensive assessment of these relationships through space and time, resulting in the ability to compare AMSR and SSM/I FTSPs, examine differences between SMAP 36 km and enhanced 9 km products, as well as those between K- and L-band based freeze/thaw classifications across an extended domain.

## **2.2 A Global Comparison of Satellite Freeze/Thaw Records**

Using lessons from regional-scale studies of satellite-based freeze/thaw estimates and surface temperatures, we employ a large-scale comparison to examine the limitations of current freeze/thaw observational approaches on a global scale. Specifically, this includes exploring where FT uncertainty is the highest by examining where products diverge. Preceding efforts to compare existing records are largely absent, in part because of known differences between microwave observations of different bands, but more so due to the relatively short duration of the SMAP-based FT record at the time of this project's inception (April 2015 – 2017). In order to fill this gap, this study investigates regional agreement between ESDR and SMAP FTSPs and identifies potential sources of classification variability. The SMAP and SSM/I-FT records are compared over an extended period covering multiple seasonal cycles from April 2015 through December 2017. The spatially and temporally varying relationship between these products is examined in relation to climate, land cover, and topography.

While reiterating some challenges of classifying frozen ground identified by prior studies, this work also contributes new insights by providing detailed geospatial and seasonal analyses into the factors contributing to classification uncertainty over the global domain. This information can in turn be applied to address current product limitations and advance spaceborne observation of land surface freezing and thawing processes.

### **2.2.1 The Need for Global Scale Comparisons of Freeze/Thaw Records**

The use of satellite P-MW remote sensing techniques has enabled the development of global twice-daily observations of surface properties over an extended

period. As mentioned, long term freeze/thaw records (1979 – present) have been derived using  $T_b$  retrievals from Ka-band sensors like the Advanced Scanning Microwave Radiometer (AMSR, AMSR-E, AMSR-2) and the Special Sensor Microwave/Imager Sounder (SSM/I or SSMI, SSMIS) series of instruments (*Kim et al. 2011, 2017; Zhao et al. 2011*). L-band P-MW retrievals from the Soil Moisture Ocean Salinity (SMOS) mission (*Rautiainen et al. 2016*) and the Soil Moisture Active Passive (SMAP) mission have also been useful for classifying surface FT state (*Derksen et al. 2017*). These records of surface FT are detailed in **Table 1**, with the exception of SMAP-FT products having been recently extended to global coverage (Version 2, *Dunbar et al. 2018*).

However, the use of varied microwave frequencies, algorithms, and resolutions (spatial and temporal) can result in very different classifications due to physical differences in surface properties and sensing band capabilities (*Colliander et al. 2010; Podest et al. 2014; Chai et al. 2014; Lyu et al. 2018; Johnston et al. 2019*). Different retrieval timings of around 1:30 a.m./p.m. local solar time (AMSR-2) and approximately 6 a.m./p.m. (SSM/I and SMAP) have resulted in significantly different observed responses, especially during transitional periods when surface states can change frequently (i.e., spring and fall).

Still, the identification of areas in which P-MW derived FT classifications are most variable and uncertain has not been comprehensively assessed at a global scale. Regions with frequent FT transitions (*Roy et al. 2015*), snow cover/density, abundant surface water (*Roy et al. 2017*), topography and/or complex vegetation and soil layers (*Ulaby et al. 1986; Kimball et al. 2009; Podest et al. 2014; Kraatz et al. 2018*) have been

identified as challenging regions to accurately assess surface FT state. Many of these challenges also exist in the collection of any satellite-derived data records. Thus, a global comparison allows the many dynamics driving P-MW FT product uncertainty to be assessed across bands. Lower frequency L-band microwave retrievals have been shown more effective at estimating the underlying state of the soil, being less impacted by vegetation and snow cover (*Kimball et al. 2009; Rautiainen et al. 2012; Bateni et al. 2013*). Still, SMAP field validation efforts have shown emissions and FT classification response are most dominated by the top few cm of the surface as frequent frozen classifications were made even as soil was not frozen at 5 cm depths (*Rowlandson et al. 2018*). An assessment of SMAP FTSPs by *Kim et al. (2019)* confirmed reduced agreement of FT state classifications to ground temperature stations due to terrain complexity, open water, and vegetative cover in L-band. The FT product pixels adjacent to water bodies are shown to have diminished FT agreement with ground stations, most notably when compared against the enhanced 9 km SMAP product.

Similarly, measured emissions in Ka-band as part of FT-ESDR have been shown least accurate in areas with open water, drier climates, and high grid cell heterogeneity (*Kim et al. 2017*). The introduction of such sub-grid-scale FT heterogeneity over mountainous and transitional regions introduces greater uncertainty in FT classification and reference states (*Du et al. 2015*). Other studies have indicated that P-MW FT products align more closely with air temperature than soil temperature when estimating FT state, as algorithms are frequently calibrated using synoptic air temperature measurements (*Roy et al. 2015; Dunbar et al. 2018; Rowlandson et al. 2018; Johnston et*

*al. 2019*). This indicates that current FT products may only capture the very top surface layer state, even in L-band. Seasonal dynamics can also lead to different landscape properties dominating the microwave FT response in spring and fall. In fall, during freeze onset with limited snow cover, soil temperature more closely mirrors MW-based FT products, whereas in spring the presence of wet snow tends to dominate the observed MW response and is better captured by air temperature (*Johnston et al. 2019*).

A direct time-series comparison examining the seasonal contribution to variability in frozen extent is performed to identify the general trends in agreement or disagreement between these products. Spatial agreement trends are also examined along with differences in FT retrieval algorithms. Moreover, through the evaluation of FT classifications by topography, land cover, and climate classifiers, we assess the degree to which these variables impact FT classification globally. These results are processed during the time of record overlap of the FT-ESDR and SMAP Version 2 (R16) FT records from April 2015 – December 2017 and provide an extension to existing assessments of these products. The ability to accurately represent FT state has implications beyond hydrology and ecology as FT state and land surface temperature are also valuable inputs in global climate modeling (*Farhadi et al. 2014; Reichle et al. 2010; Xue et al. 2019*). As the first extended global comparison of P-MW FT products, this study provides valuable insight into the identification of challenging classification regions and climates as well as probable causes of FT state disagreement. The outcomes of this work bolster efforts towards improving algorithms and increasing the accuracy

and representativeness of freeze/thaw classifications. Outcomes from this work are expected to be used in climate studies and future modeling efforts.

## 2.2.2 Methods for Global Comparison

### *Freeze/thaw Detection Considerations*

As a result of largely similar classification methodologies (**Chapter 1.2**) of the SMAP FT (SMAP\_L3\_FT\_P\_E R16010, *Xu et al. 2018*) and FT-ESDR (AMSR\_36V\_FT & SSMI\_37V\_FT, *Kim et al. 2017*) products used herein, reasonable comparisons can be made in regard to differences between L- and Ka-band based FT classifications. Even still, algorithm variability over northern regions and masking procedures must be considered when examining results, as mask variability can artificially produce differences in FT classification agreement. Both products are masked over regions with permanent ice and snow (i.e., Greenland, Antarctica), open water, and non-FT constrained areas (*Kimball et al. 2009; Dunbar et al. 2018*). In order to mitigate false classifications, two additional masking procedures are applied to SMAP FTSPs. This consists of marking all cells in which vertically or horizontally polarized  $T_b$  is larger than 273 °K as thawed and applying daily ‘never frozen’ and ‘never thawed’ masks over regions that AMSR-E and AMSR2 derived FT maps have historically classified as only frozen or thawed (*Kim et al. 2012; Dunbar et al. 2018; Kim et al. 2019*). For example, this includes the high northern latitudes in January (‘never thawed’) and much of the northern hemisphere domain in late July (‘never frozen’).

A limited number of studies have been performed analyzing the differences between P-MW derived FT state in various bands. Synthetic experiments have been

performed using passive (1.4, 18.7, 36.5 GHz) and active microwave (1.4, 5.4, 12 GHz) concluding that lower L-band frequencies are more effective at soil FT state estimation while higher frequencies were affected more by snow properties (*Bateni et al. 2013*). In a study by *Podest et al. (2014)*, backscatter in L-, C-, and Ku-bands was assessed against vegetation, elevation, and aspect over Alaskan sites showing general agreement in FT classification. The largest differences were attributed to terrain and varied product resolutions. Global studies have not been performed comparing P-MW records, though Prince et al. (2018) showed that L- (Aquarius scatterometer) and Ka-band (FT-ESDR radiometers) FT classifications had agreement over 80% in the majority of grid cells  $>50^{\circ}\text{N}$ . They also identified product differences relative to freeze/thaw onset dates and across tundra, forest, and open lands. However, these efforts provided a limited assessment over only a few broad land cover classes.

#### ***Agreement Proportion Metric***

Prior to classification comparisons, all products are resampled to a  $0.5^{\circ}$  global grid ( $\sim 49$  km or  $\sim 2,400$  km<sup>2</sup> at  $40^{\circ}$  latitude). Both the 9 km ( $\sim 81$  km<sup>2</sup>) enhanced SMAP product and 25 km ( $\sim 625$  km<sup>2</sup>) FT-ESDR boast considerably higher resolution. This illustrates the relatively large regions covered by each comparison pixel, frequently capturing different terrain, water, soils, and vegetation types. For the SMAP enhanced product (9km) this results in having more than 5 datapoints in over 96% of half-degree cells, while the 25km resolution FT-ESDR products have a diminished data density in which 94% of cells had at-least 2 datapoints. These resolution differences can induce errors, especially in less data dense polar regions. Conversely, these regions are well

covered by the near-polar orbits of SMAP, AMSR-E, and SSM/I, which provide frequent coverage even if it is not well captured by the grid spacing used in this study. Grid matching was performed in order to match product resolutions while ensuring multiple classifications per cell and to reduce computational time. Overall, this methodology captures >40,000 comparable data cells globally within FT constrained areas matched to the same spatial domain. Variable datasets such as elevation, land cover, and climate classifications were also scaled to 0.5° resolution. While this does induce error by creating a lower resolution dataset, it is sufficient at capturing global patterns in spatial agreement between FT products and climate indicators, land cover, and topographic variables.

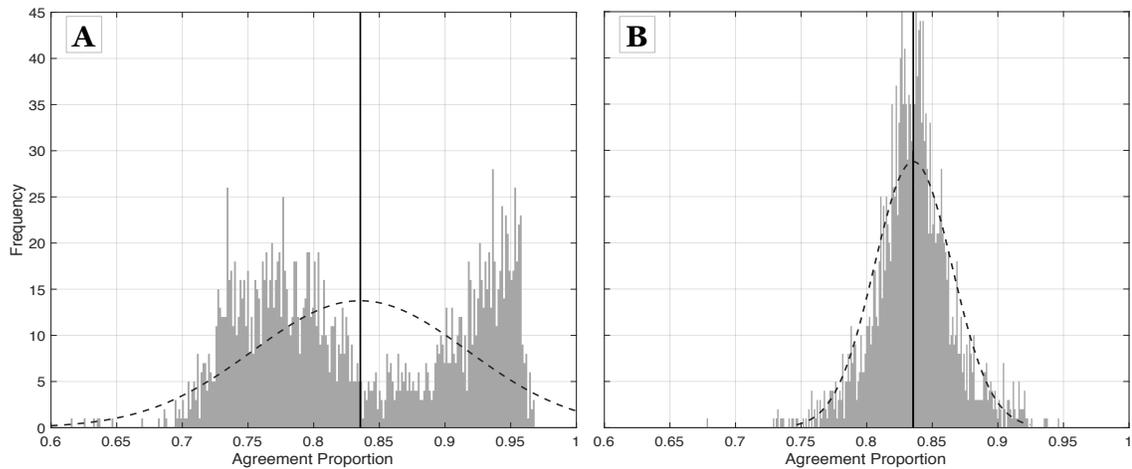
The primary comparison metric between products, agreement proportion ( $A_p$ ), is defined as the number of cells within a given class/location that agree ( $FT_y$ ) compared to the total number of comparison points.

**Equation 9 Agreement proportion**

$$A_p = \frac{FT_y}{FT_y + FT_n}$$

where  $FT_n$  is the total number of disagreeing FT classifications. This calculation is performed after matching SMAP and SSM/I ascending and descending classifications both to the FT constrained domain and same temporal range.  $A_p$  is then computed two ways, both geospatially and through time. The geospatial calculation computes agreement within a given cell over the entire study period, enabling a spatially distributed assessment of agreement. Alternatively, time series calculations are performed by computing daily  $A_p$  over an entire land surface category (e.g., deciduous forests).

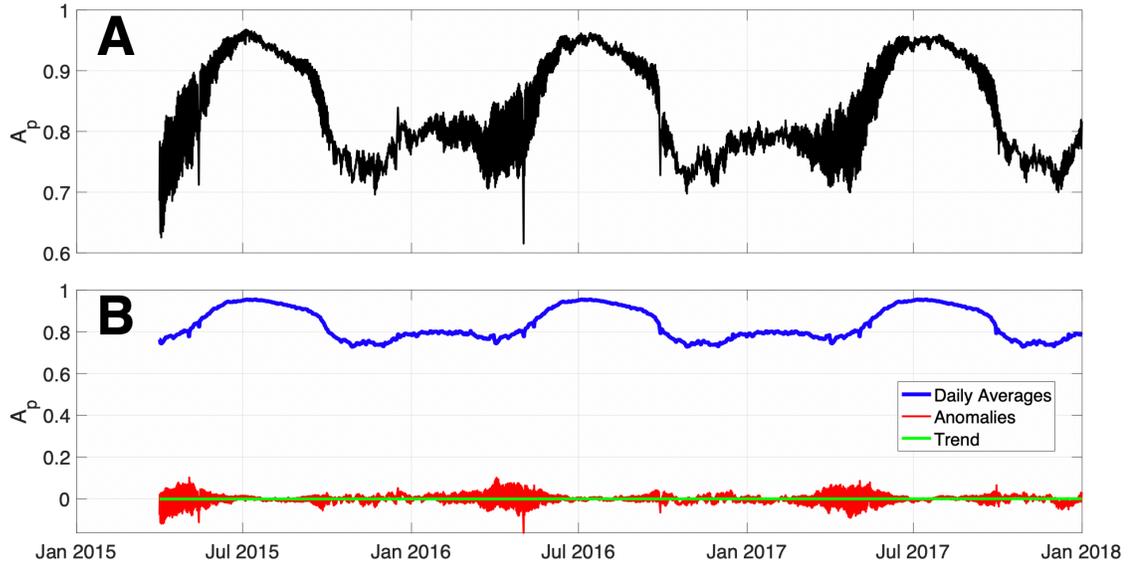
## Seasonal Adjustment for Statistical Analyses



**Figure 17** Distribution of global Agreement Proportion ( $A_p$ ) SMAP to SSM/I-FT, for entire study period (April 2015 – December 2017, includes 1,992 overpasses). (A) Raw distribution and (B) with seasonal adjustment to approximate normality. Distribution mean of 0.835 (~84% global agreement) indicated by black line. Normal distribution fits are shown by the dashed black curve

In order to apply standard statistical methods, a seasonal adjustment is performed across all comparison classes. This is critical in meeting t-test assumption of approximate distribution normality. Clear seasonal variability in global  $A_p$  is observed across the study period forming a bimodal distribution (**Figure 17A**), having high agreement during the summer and lower relative agreement during the freezing period (**Figure 18A**). The seasonal adjustment is performed using time series deconstruction techniques in which data are split into trend, seasonal, and random anomaly components (**Figure 18B**). The trend component is determined through the fitting of a regression line to approximate the dataset mean at a given point in time. Next,  $A_p$  averages are computed for each individual day of the year forming the seasonal component. Lastly, this seasonal component is

removed by correcting for the difference between trend and seasonal components, leaving only the regression and anomaly components.



**Figure 18** Timeseries detrending and seasonality removal. (A) Global product agreement showing seasonal pattern and (B) which shows each deconstructed component (average, trend, anomaly) from (A)

This process results in a new distribution of product agreement with seasonality removed, while preserving the overall distribution mean (**Figure 17B**). This method is applied to deconstruct time series of  $A_p$  from each comparison class, enabling the computation of 99% confidence intervals each classes' respective mean. As a result, classes statistically different from each other and/or the overall global mean are recognized. Similar techniques are used by U.S. Census Bureau in which seasonal cycles are removed from a variety of datasets to reveal underlying trends (*Monsell 2007*). This approach is applied only in data summary tables and does not impact the spatial calculation or time-series figures. Additionally, we compare the seasonally adjusted data

to uncorrected raw distributions and verified that the resulting conclusions from this study were unaffected.

### *Categories for Assessing Freeze/Thaw Agreement*

The agreement between the P-MW derived FT products is assessed in relation to four spatially distributed variables. These include temperature and precipitation relevant Köppen-Geiger climate classes and freezing days, as well as land cover classes, and topographic variability derived from a global elevation dataset. The original climate and land cover classifications have been slightly modified to reduce the number of comparison classes (**Tables 6 - 7**), whereas the number of freezing days is computed through a standard technique utilizing daily mean air temperatures from the Global Land Data Assimilation System (GLDAS).

The climate classification dataset was updated in 2006 to accurately represent the global climate over the second half of the 20th century (*Kottek et al. 2006*). Classifications are dependent on regional precipitation and temperature datasets and split into equatorial (A), arid (B), warm temperate (C), snow (D), and polar (E) climates. Originally comprised by 31 classes within these climate types, this number is reduced to 15 to include only classes included in FT constrained regions as defined by FT data products and to simplify comparison (**Figure 19A**). This is achieved through the removal of all equatorial classification types (4 classes), water, and the combination of several similar classifications into new classes (**Table 6**). Combined classes are determined by removing the distinction between hot and warm summers, steppe, and desert climates, as well as combining similar cold continental regions with few data cells.

**Table 6 Modified Köppen-Geiger Climate Classifications with original and modified abbreviations. Similar climate classes combined, and tropical classes removed due to having too few comparable cells within the freeze/thaw constrained domain. Number of cells indicate total 0.5° grid cells within each class**

Original Abbreviations	Modified Abbreviation	Label	Cells
BSh, BWh	Bh	Arid Hot	1,977
BSk, BWk	Bk	Arid Cold	4,778
Cfa	Cfa	Temperate Without dry season Hot summer	2,426
Cfb	Cfb	Temperate Without dry season Warm summer	1,954
Csa, Csb	Csh	Temperate Dry summer Hot/Warm summer	1,460
Cfc, Csc, Cwc	Cc	Temperate Cold summer	192
Cwa, Cwb	Cwh	Temperate Dry winter Hot/Warm summer	726
Dfa, Dfb	Dfh	Cold continental Without dry season Hot/Warm summer	5,212
Dfc, Dfd	Dfk	Cold continental Without dry season Cold summer/Very Cold winter	12,030
Dsa, Dsb	Dsh	Cold continental Dry summer Hot/Warm summer	286
Dsc	Dsc	Cold continental Dry summer Cold summer	262
Dwa, Dwb	Dwh	Cold continental Dry winter Hot/Warm summer	972
Dwc, Dwd	Dwk	Cold continental Dry winter Cold summer/Very Cold winter	1,415
EF	EF	Polar Eternal Winter/ Ice Cap	383
ET	ET	Polar Tundra	6,111

The International Geosphere-Biosphere Programme land cover classification (IGBP) utilized in this study consists of 17 land cover classes determined from MODIS spectral data (*Friedl et al. 2010*) accessed through the University of Maryland’s Global Land Cover Facility (GLCF; *Channan et al. 2014*). These classes have been reduced to 12 by combining evergreen and broadleaf deciduous forests, open and closed shrublands, savannas, and cropland dominated classes (**Figure 19B, Table 7**). Original classification legends with detailed class descriptions can be found in the Collection 6 MODIS Land Cover user guide (*Sulla-Menashe and Friedl 2018*).

**Table 7 Modified IGBP land cover classifications from MODIS MCD12Q1**

<b>IGBP Values</b>	<b>Label</b>	<b>Cells</b>
1	Evergreen needleleaf forest	2,214
2	Evergreen broadleaf forest	277
3 & 4	Deciduous forests	1,274
5	Mixed forest	5,083
6 & 7	Shrublands	9,559
8 & 9	Savannas	2,381
10	Grasslands	7,310
11	Permanent wetlands	343
12 & 14	Croplands	5,232
15	Snow and ice	1,385
16	Barren or sparsely vegetated	2,165
17	Water	3,642

Next, global topographic complexity is quantified through the use of 0.125° resolution elevation data. This global elevation dataset was developed by the United States Geological Survey (USGS) in 2010 as part of the Global Multi-resolution Terrain Elevation Datasets (GMTED) through a combination of existing 30-arc second elevation information (*Danielson and Gesch 2011*). Standard deviation is computed for each 0.5° cell using the higher resolution (0.125°) dataset. This results in elevation standard deviations ranging from 0 to 1300 meters over the study domain (**Figure 19C**).

Finally, not to be confused with the number of freezing days defined by FT products, freezing days in this context refers to the number of days in which mean daily temperatures are below 0 °C. Several studies have used similar approaches by using air temperature (freezing degree-days) to characterize and estimate permafrost depth (*Nelson and Outcalt 1987; Karunaratne and Burn 2003*). Here, average daily temperature is computed utilizing 3-hourly air temperatures from the GLDAS forced Noah Land

Surface Model L4 0.25° data product (GLDAS\_NOAH025\_3H; Rodell et al. 2004). This dataset was then resampled to 0.5° resolution.

*Ji et al. (2015)* performed a comprehensive evaluation of GLDAS air temperature products, showing good skill at reproducing air temperature observations. Using this data set, the sum of the mean daily temperatures below the 0 °C threshold is computed as the total freezing days during the study period. This comparison of agreement by freezing days allows an assessment of the degree to which FT product variability is related to freezing potential. Regions with high numbers of freezing days are expected to experience extended frozen periods and are generally located at high latitudes and elevations, whereas lower numbers of freezing days indicate regions characterized by ephemeral freezing events generally in the lower and mid-latitudes (**Figure 19D**).

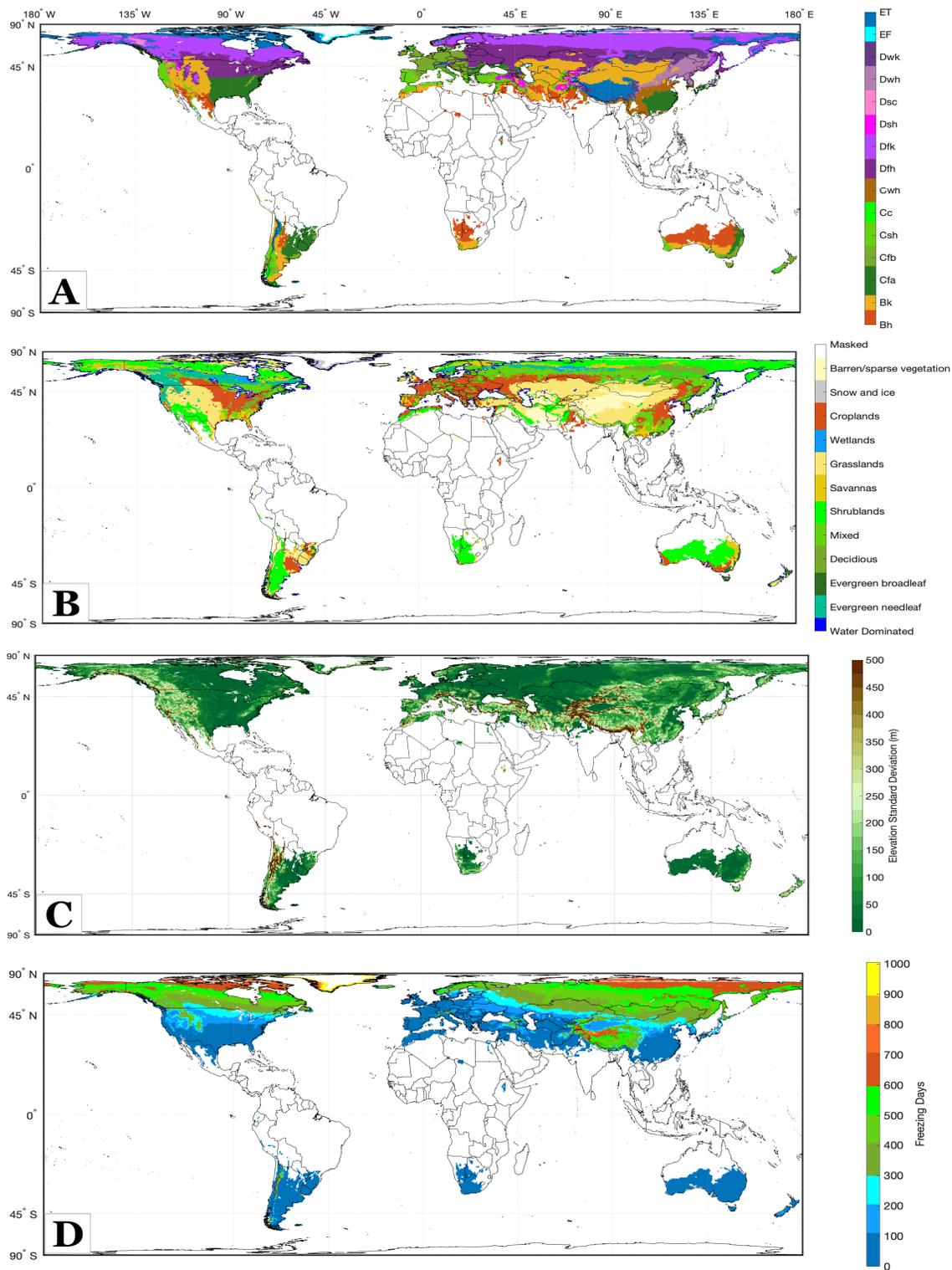
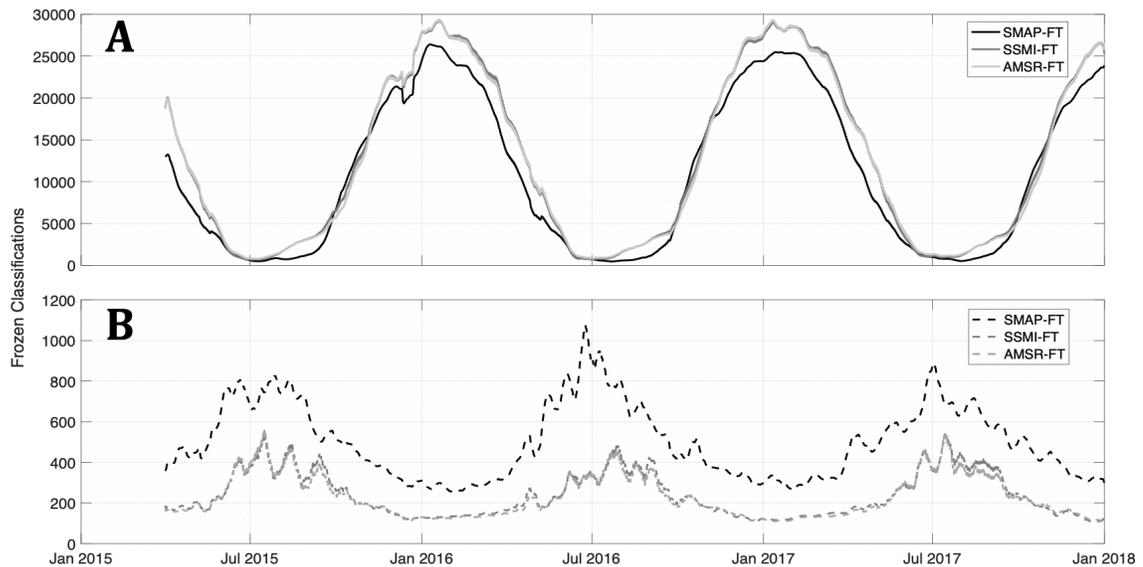


Figure 19 (A) Modified climate classes (see Table 6), (B) modified land cover classes (Table 7), (C) GMTED elevation standard deviation, and (D) freezing days. All non-FT constrained regions are masked

## 2.2.3 Global Comparison Results

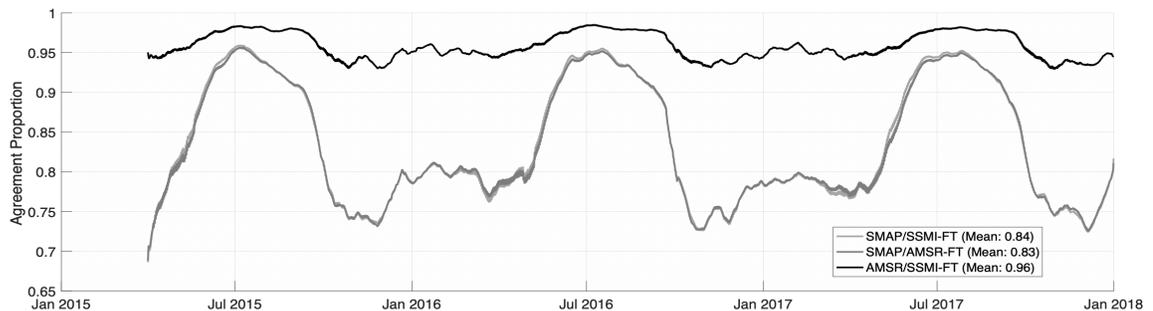
### *Exploring SMAP and ESDR Freeze/Thaw Products*

Time series of the total number of FT classifications for each product is shown in **Figure 20**, divided by hemispheres. Over the northern hemisphere (NH), products are characterized by high agreement in frozen extent during mid-summer (July) and during freeze onset. Relative to FT-ESDR, SMAP produces an increase in frozen classifications in the southern hemisphere (SH) year-round. This is in contrast to a reduction in SMAP frozen classifications relative to FT-ESDR records in NH. This also illustrates the larger area of FT constrained land in NH relative to SH (~25x more frozen classifications) (**Figure 21**). Similarities between FT-ESDR records are clear, as they are derived from identical algorithms using similar P-MW bands.



**Figure 20** Total frozen classifications by hemisphere. (A) Northern, (B) southern for select freeze/thaw products

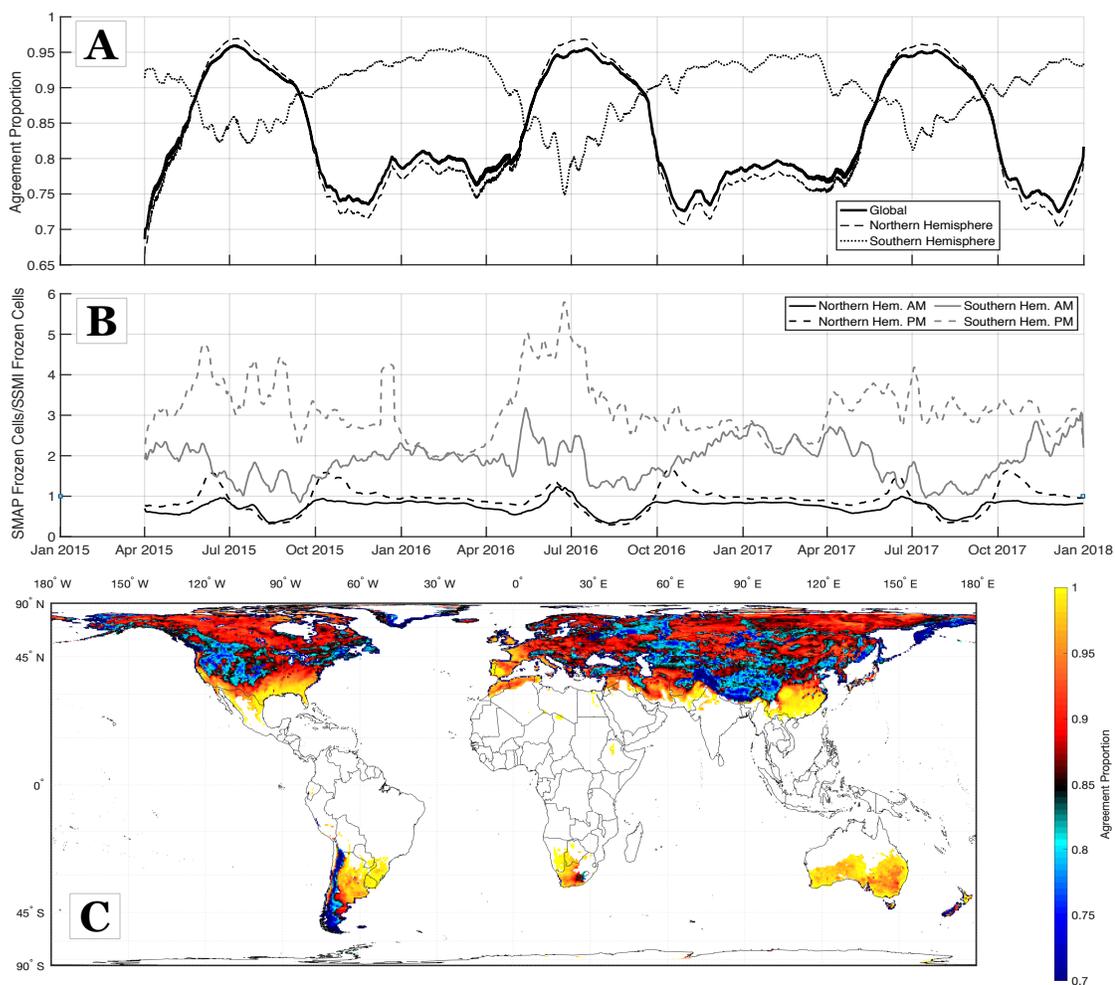
The resulting  $A_p$  time series for the entire FT domain are presented in **Figure 21**. The agreement of the FT-ESDR AMSR and SSM/I is shown to frequently exceed 95% throughout the study period (April 2015 – December 2017). Overall, Ka-band FT-ESDR products derived from SSM/I and AMSR observations provide very similar classifications during the study period, agreeing 96% of the time. The largest classification differences occur during the NH freeze (September – December) and thaw (April – May) periods in which agreement falls marginally below 95%. In contrast, SMAP and FT-ESDR products agree at 83-84% over the study period with well-defined reductions in  $A_p$  from late September (<75%), remaining relatively low through June. Due to comparison similarity between ESDR records (**Figure 21**) and overpass time similarity between SMAP and SSM/I, only  $A_p$  derived from SMAP and SSM/I FTSPs is assessed against global variables.



**Figure 21 Agreement proportion ( $A_p$ ) between global P-MW freeze/thaw products. Indicated values are mean  $A_p$  over the full study period**

### ***Global Agreement Trends***

A timeseries of global  $A_p$ , a comparison of SMAP frozen classifications relative to SSM/I, and the global distribution of  $A_p$  through time are detailed in **Figure 22**. As the northern hemisphere contains around 90% of the global FT constrained land mass, global agreement largely mirrors that of the region (**Figure 22A**). Slight  $A_p$  reductions are prevalent in June-August due to reduced winter agreement over the southern hemisphere, as well as small increases in October-May due to high summertime agreement. The seasonal relationship is apparent in both hemispheres, as the lowest  $A_p$  occurs during freeze onset during September-November in the NH (72-73%) and May-July in the SH (75-85%). Also, both hemispheres are characterized by the highest agreement in mid-summer when the domain is primarily thawed. Over the NH, slight increases in  $A_p$  during the thaw onset (75-78%) are observed relative to freeze onset (72-73%), supporting the tendency L- and Ka-band P-MW FT products to observe melt conditions more similarly than during the seasonal freeze onset period.



**Figure 22 (A) Mean agreement proportion between SMAP and SSM/I derived FT classifications, (B) normalized comparison of frozen extent by hemisphere and overpass, and (C) global map of agreement for entire period**

In the NH, mid-winter agreement increases to around 80% as SMAP and SSM/I retrievals agree over expansive frozen regions. While this leads to a slight increase in agreement, generally, as more area is defined frozen by these products, the more variation there is among them. However, it is important to note that much of the FT domain in South Africa and Australia is effectively thawed throughout the study period by both

products. This increased the agreement overall, while agreement in South America and New Zealand was considerably lower than the hemispheric average (**Figure 22C**).

**Figure 22B** shows the relationship in relation to frozen extent, where a value of 1 indicates an equivalent number of frozen classifications and  $>1$  indicates SMAP as having more frozen classifications. In the SH, SMAP consistently classifies more cells as frozen compared to SSM/I. This is most pronounced during PM overpasses where the SMAP product ranges from having two (2x) to as many as six (6x) times the frozen classifications as SSM/I during the winter period. The large discrepancy in frozen extent in the SH is likely linked to the relatively small number of FT constrained cells, increased Ka-band sensitivity to surface thaw as soil remains frozen, as well as product differences in coastal and mountainous portions of South America (**Figure 22C**). Conversely, in the NH, SSM/I FT classifications tend to define a slightly larger frozen extent compared to SMAP, except during the PM overpasses during late melt (June) and freeze onset (October) periods. This result supports the increased sensitivity of L-band SMAP observations beyond the very surface skin layer, as Ka-band is more sensitive to surface melt. This likely results in the increase in thawed classifications by the SSM/I FT product as daytime heating and sun-exposure can lead to the presence of liquid water at the surface. FT-ESDR products are also characterized by an increased variability from the AM to the PM when compared to SMAP. This translates into classifying more frozen area than L-band during diurnal freeze events in the morning as well as classifying larger regions as thawed during the afternoon hours. These comparisons underscore the

increased sensitivity of Ka-band to detect ephemeral changes (e.g., frost or moisture changes) in the top surface while underlying soil may remain thawed or frozen.

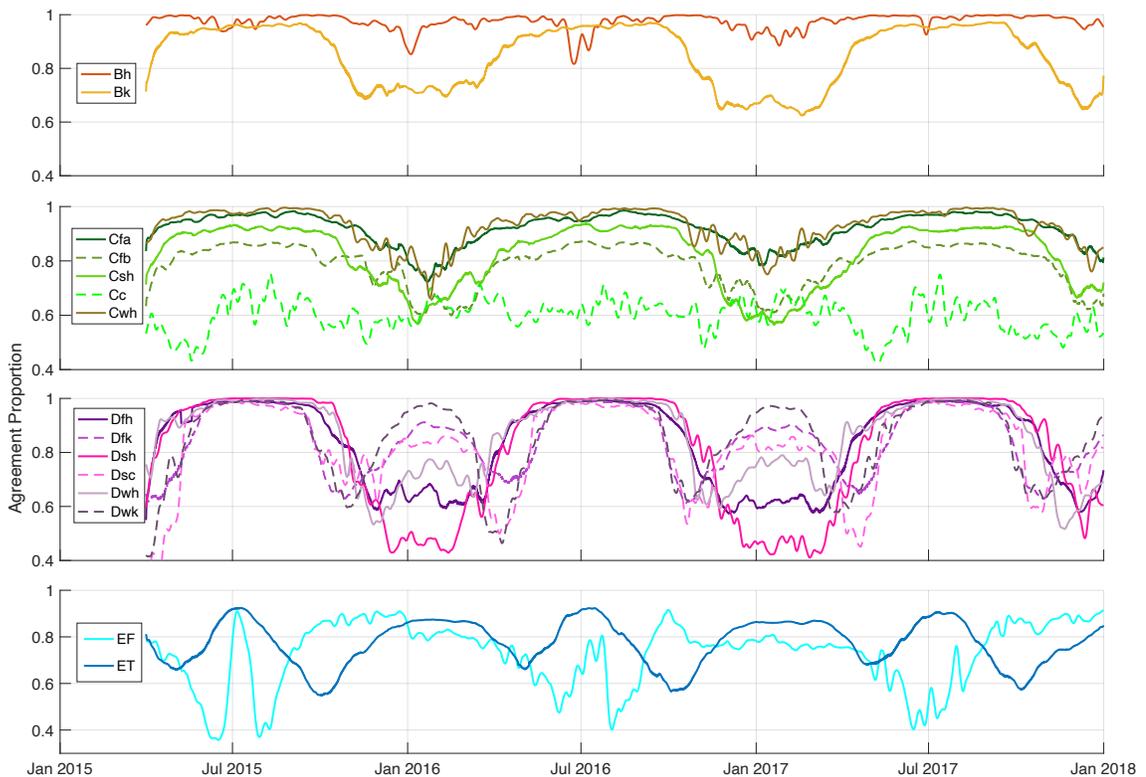
The spatial distribution of product agreement highlights regions in which P-MW-based FT classification is particularly challenging. Some more temperate areas have very limited periods of sub-freezing temperatures (freezing days < 100, **Figure 19D**) and include the southern U.S., Australia, southern China, and portions of South America in which products agree in excess of 95% (**Figure 22C**). This is primarily due to consistent thawed classifications throughout the year in both products. Coastal areas and portions of Greenland as well as the higher latitudes frequently have  $A_p$  below the global mean (83.5%). While this is due in part to the varied product resolutions and algorithms, the land water interface in coastal regions has proven difficult for FT classification in prior studies (*Kim et al. 2019*). In addition to coastal regimes, mountainous regions and those with higher terrain exhibit reduced agreement between products, most notably regions such as South America, the western U.S., and High Mountain Asia (HMA). These areas prove challenging for remote sensing of FT due to the potential for increased inter-pixel heterogeneity and an increased number of FT cycles due to rapidly changing climates frequently found in high elevations (*Podest et al. 2014*). Dry conditions in HMA may also challenge freeze detection.

### ***Effect of Spatial Variables on Freeze/Thaw Agreement***

#### *Climatology*

Uncertainty due to variations between SMAP-FT and FT-ESDR has been observed to follow clear seasonal cycles (**Figures 21, 22**). This variability occurs most

notably during periods in which large portions of the global domain are undergoing freezing or melt processes. The link between FT products and their performance over variable climate regions can provide insight as to what regions FT classifications remain uncertain and the contributing factors.



**Figure 23** Time series FT agreement proportion ( $A_p$ ) separated by climate classes (Table 6). Colors correspond to those in Figure 19A

**Figure 23** presents the time series variability in  $A_p$  across 15 global climate classes. Arid regions (Bh/k) have notable variability in product agreement (**Table 8**) as hot arid regions are minimally FT constrained (0.2-2.5% frozen during the study period) compared to cold arid regions (15.0-17.1%). The low frequency of FT transitions results

in increased  $A_p$  over hot arid regions (97.4%) compared to cold arid regions (85.0%) which has  $A_p$  near the global mean. Similar relationships are shown to exist across all temperate regions (C) excluding the temperate cold summer (Cc) class and the largely non-FT constrained Cfa (91.9%) and Cwh (92.8%) classes. In these temperate zones the SMAP-based FT product has a clear tendency to define a larger frozen extent compared to SSM/I. Large decreases in agreement over Cc regions is likely related to coastal and resolution impacts as these cells are focused over the Aleutian Islands, Pacific Northwest, New Zealand, Iceland, and Northern Ireland. These relatively cool coastal regions are shown to have  $A_p$  of 59.9% falling significantly below the global mean as SMAP defines nearly 42% of all possible cells as frozen compared to just 14.5% for SSM/I (**Table 8**). FT classification challenges in these regions have been documented as SMAP is observed to classify large parts of the cool temperate regions (Cc) as frozen leading to reduced agreement in mid-summer (**Figure 23**). Seasonal variability in  $A_p$  is recognizable across all but the Cc class with the overall agreement declining over temperate classes such as Csh (82.3%) and Cfb (78.7%). In these areas, SMAP derived FT products define a notably higher number of frozen classifications. This indicates that over regions with less well-defined FT seasonality, FT classifications remain relatively uncertain. Variability in the fundamental emitting layers as observed by L- and Ka-band radiometers combined with an increased frequency of borderline freezing conditions are likely contributors to this uncertainty.

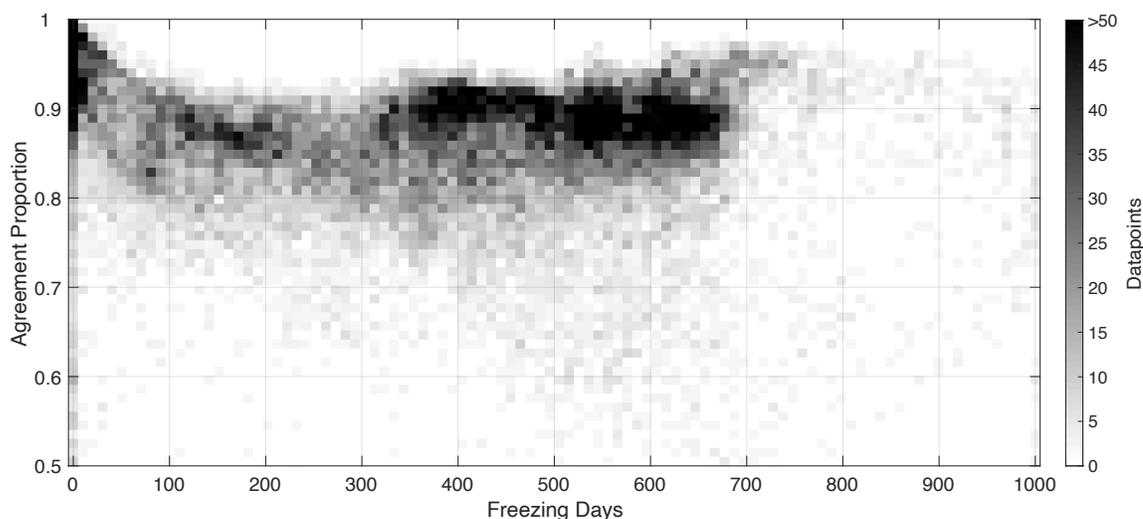
**Table 8 SMAP-FT to SSM/I-FT agreement statistics for  $A_p$  (mean, standard deviation, and confidence interval) separated by climate classifications. Percentage of class defined as frozen is computed over the entire study period. Global mean = 0.835**

Abbreviation	Label	Cells	Mean	Standard deviation	99% confidence interval	Percentage of Class Defined as Frozen	
						SMAP	SSM/I
Bh	Arid Hot	1,977	0.974	0.034	.972 - .975	3%	0%
Bk	Arid Cold	4,778	0.850	0.125	.847 - .854	15%	17%
Cfa	Temperate Without dry season Hot summer	2,426	0.919	0.066	.917 - .921	7%	3%
Cfb	Temperate Without dry season Warm summer	1,954	0.787	0.088	.785 - .790	20%	7%
Csh	Temperate Dry summer Hot/Warm summer	1,460	0.823	0.117	.821 - .826	16%	5%
Cc	Temperate Cold summer	192	0.599	0.084	.595 - .602	42%	15%
Cwh	Temperate Dry winter Hot/Warm summer	726	0.928	0.087	.925 - .931	5%	4%
Dfh	Cold continental Without dry season Hot/Warm summer	5,212	0.836	0.164	.832 - .839	16%	23%
Dfk	Cold continental Without dry season Cold summer/Very Cold winter	12,030	0.844	0.131	.841 - .847	38%	43%
Dsh	Cold continental Dry summer Hot/Warm summer	286	0.813	0.228	.808 - .818	11%	19%
Dsc	Cold continental Dry summer Cold summer	262	0.822	0.178	.816 - .827	34%	43%
Dwh	Cold continental Dry winter Hot/Warm summer	972	0.853	0.157	.849 - .857	20%	26%
Dwk	Cold continental Dry winter Cold summer/Very Cold winter	1,415	0.850	0.168	.846 - .855	35%	44%
EF	Polar Eternal Winter/ Ice Cap	383	0.727	0.156	.722 - .733	81%	72%
ET	Polar Tundra	6,111	0.776	0.104	.773 - .778	48%	56%

Continental zones can be described as having well-defined seasonality and generally frigid winters, including much of North America and Russia. Different from temperate zones, these regions are characterized by high agreement during the summer period (>95%) before a steep decline in the shoulder seasons (**Figure 23**). During the NH freezing period (September-June) agreement is shown to diverge greatly across continental sub-classes, ranging from 40% to 95%. In the northern most parts of the continental zone (Dfk, Dsc, and Dwk classes), following freeze onset and prior to thaw, products show high agreement >80%. Conversely, lower latitude regions are characterized by relatively low wintertime agreement, as low as 40%. This suggests extreme variability and uncertainty FT classifications, even in lower latitude continental zones even with clear FT seasonality (**Table 8**). In contrast to temperate regions, SSM/I regularly classifies more of the continental zones as frozen compared to SMAP, especially in drier regions (Dsh, Dsc, and Dwk 8.1-9.5% more frozen with SSM/I)

indicating challenges in capturing a clear FT transition in measured P-MW emissions over land surfaces with little moisture. This also indicates that Ka-band derived FT records have an increased tendency to classify more of the domain as frozen during the September-June period. In Dwk zones, further investigation indicates a tendency of SSM/I to define more frozen area by a large degree (+8.9%), especially during the thaw period. This would support potential of L-band emissions to observe FT deeper into the surface, potentially measuring liquid water near the base of the snowpack. The utilization of FT products which better capture the ephemeral nature, such as fractional classification, will likely improve the accuracy of FT state representations.

Regions covered by ice caps (EF) and polar tundra (ET) experience  $A_p$  significantly below the global mean of 83.5% (72.7%-77.6%) (**Figure 23**). Regions classified as ice caps are limited to outer boundaries of the Greenland ice sheet deeming any analysis over this region limited (**Figure 19A**). Nonetheless, these regions experience reduced agreement during the shoulder seasons (<50%) likely due to meltwater beginning to appear at the surface of the ice sheet. Interestingly, agreement in these regions is not near 100% even during the mid-winter period when regional temperatures are well below freezing, indicating difficulties with FT references over ice covered regions. SMAP (81.3%) also defines this region as frozen more frequently than SSM/I (71.6%) potentially showing the tendency of Ka-band to respond more to surface melt than SMAP (**Table 8**). When assessing contributors to differences in the number of frozen classifications, variability in algorithms over the high-latitudes and Greenland which use both the NPR and single channel SCA algorithms must also be considered.



**Figure 24** Density plot of freezing days against agreement proportion ( $A_p$ )

Further investigation into the relationship between climate, more specifically cold temperature periods, and product agreement is explored using freezing days. This relationship is presented in **Figure 24** and summarized by **Table 9**. As anticipated, P-MW FT products tend to agree (93.2%) over regions that are largely not FT constrained (0 freezing days). These regions include Australia, southern North America, as well as parts of South America and southern Africa in which zero days over the study period were modeled to have mean temperatures below freezing. Still, SSM/I and SMAP correspond poorly in the rare cases when either product classifies a cell as frozen. This is illustrated by SMAP classifying cells as frozen more than 7 times as frequently compared to SSM/I (6.6% to 0.9%). This discrepancy implies a cross platform challenge of using P-MW remote sensing to classify FT state in regions that only experience ephemeral freeze events. Regions experiencing 1-300 freezing days are largely located in temperate zones (**Figure 19D**) characterized by less well-defined FT seasonality classified 5.7-21.4% of

the study period as frozen. Regions with 300-700 freezing days exhibit an increase in mean agreement between products as these zones are largely within continental climate classes which have high agreement during the summer and long winter period. In these areas both products classify a similar number of frozen cells as frozen, 22.7-52.0% (SMAP) and 28.6-55.9% (SSM/I).

**Table 9 Agreement by number of days in which GLDAS mean daily air temperatures are < 0 °C (Freezing Days). Includes April 1, 2015 – December 31, 2017 (1005 days)**

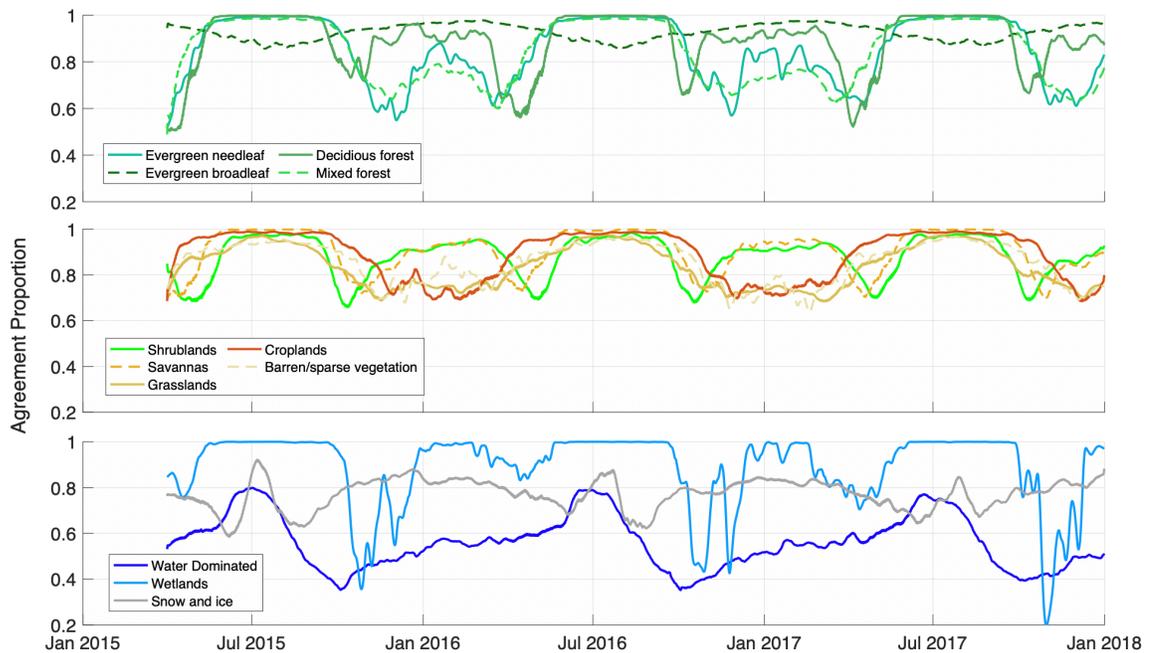
Freezing Days	0	0 - 100	100 - 200	200 - 300	300 - 400	400 - 500	500 - 600	600 - 700	700 - 800	800 - 900	900 - 1000	> 1000	
Cells	4,589	5,364	3,258	2,973	4,539	5,372	5,901	5,238	1,109	445	489	90	
Mean	0.932	0.847	0.834	0.824	0.842	0.840	0.837	0.843	0.786	0.748	0.760	0.626	
Standard deviation	0.025	0.104	0.171	0.179	0.160	0.155	0.157	0.154	0.126	0.117	0.134	0.218	
99% Confidence interval	.931 - .933	.845 - .849	.830 - .838	.820 - .828	.838 - .845	.836 - .844	.833 - .840	.840 - .847	.784 - .789	.745 - .751	.755 - .764	.617 - .635	
Percent of Period Classified Frozen	SMAP-FT	6.6%	13.6%	11.8%	15.2%	22.7%	30.7%	41.0%	52.0%	55.2%	58.9%	78.6%	84.0%
	SSM/FT	0.9%	5.7%	15.5%	21.4%	28.6%	37.1%	46.9%	55.9%	66.2%	71.2%	73.8%	63.4%

The increased number of freezing days generally results in improved agreement between both L- and Ka-band products as measured emissions both illustrate more well-defined frozen signatures. Parts of the FT domain deemed to have >700 freezing days during the study period include the Arctic, Greenland, and parts of High Mountain Asia. Even as these regions are dominated by sub-freezing temperatures for the vast majority of the study period, FT products are shown to agree at a comparatively reduced rate in these regions (62.6-78.6%), falling significantly below the global mean. This result underlines the variability in measured P-MW emissions and their relationship to FT state, likely observing FT transitions in different parts of the land surface and algorithm deficiencies.

Regions such as these would be expected to have the clearest FT signal, though this is not the case. Such regions remain critical to examining the impacts of a changing climate, as arctic and sub-arctic regions experience rapid change. This highlights these areas as especially important, requiring improvement of current FT classifiers.

### *Land Cover*

The relationship between SMAP and SSM/I based FT products is then assessed in relation to MODIS IGBP land cover classifications. The increased uncertainty corresponding to the seasonal growth and ablation of frozen land surfaces is clear among most all cover classes. This is illustrated by decreases in agreement proportion in the shoulder seasons and during the winter period (**Figure 25**).



**Figure 25** Time series of product FT agreement separated by land cover (Table 7). Colors match Figure 19B

Distinctive signatures are prevalent when comparing forested regions to those with little to no forest cover and to land surfaces characterized by an increased prevalence of water. It is important to note, that land cover and climate have considerable overlap, as specific plant species tend to grow in certain climates with particular temperature and moisture regimes. Those classified as evergreen broadleaf are located predominantly in the southern hemisphere and over largely non-FT constrained land. Evergreen broadleaf forests were defined as frozen by SMAP and SSM/I products in only 6.6% and 1.1% of cells over the study period. Other forested classes, while having similar  $A_p$  to the global mean of 83.5% (**Table 10**), each display large variability through the September – June time period. The evergreen needleleaf and mixed forest classes are shown to have similar steep reductions in agreement during the NH freeze onset (~60%) before increasing in agreement in mid-winter (>80%) and falling during the melt period. Located predominantly in the eastern U.S. and Russia, deciduous forests follow a similar tendency but with improved  $A_p$  during freeze onset and comparatively reduced  $A_p$  during the thaw period. This indicates potential varied emissions across L- and Ka-band related to snow melt within these forested regions. Across land cover types characterized by less tree cover (e.g., shrublands, savannas, grasslands, croplands, and barren regions) less seasonal variability is observed compared to forests (**Table 10**). This decrease in variability supports a smaller seasonal effect on P-MW emissions in less forested regions. Notably, similar agreement signatures are observed between areas with the least dense vegetation cover including the barren/sparsely vegetated, cropland, and grassland classes. These

regions are also primarily located in the mid-latitudes, where freezing cycles can be increasingly ephemeral, and are more homogenous.

**Table 10 SMAP-FT to SSM/I-FT agreement statistics for  $A_p$  (mean, standard deviation, and confidence interval) separated by land cover. Percentage of class defined as frozen is computed over the entire study period**

Label	Cells	Mean	Standard deviation	99% confidence interval	Percentage of Class Defined as Frozen	
					SMAP	SSM/I
Evergreen needleleaf forest	2,214	0.847	0.155	.843 - .852	28%	32%
Evergreen broadleaf forest	277	0.928	0.038	.927 - .929	7%	1%
Deciduous forests	1,274	0.888	0.154	.883 - .894	35%	40%
Mixed forest	5,083	0.843	0.143	.840 - .846	21%	25%
Shrublands	9,559	0.880	0.100	.877 - .882	35%	37%
Savannas	2,381	0.905	0.104	.902 - .908	27%	29%
Grasslands	7,310	0.836	0.098	.834 - .839	28%	31%
Permanent wetlands	343	0.887	0.182	.881 - .892	40%	35%
Croplands	5,232	0.888	0.112	.885 - .890	13%	13%
Snow and ice	1,385	0.771	0.080	.768 - .774	68%	67%
Barren or sparsely vegetated	2,165	0.853	0.099	.850-.856	21%	23%
Water Dominated	3,642	0.562	0.118	.560-.564	19%	39%

Classes in which a higher density of both solid or liquid phases water is prevalent experience very distinctive variations between L- and Ka-band products. Snow and ice as well as cells deemed water dominated experienced the lowest  $A_p$  across all classes of 77.1% and 56.2% respectively. Even within these regions, products were able to agree on classifications during the mid-summer period near 80% of the time. However, as freeze onset begins over parts of the global domain, agreement in water dominated areas declines to below 40% emphasizing the variations in P-MW emissions over these largely coastal regions. Wetlands are generally characterized by clear thaw signals in summer (~100%) with sharp declines in agreement during the shoulder seasons (**Figure 25**). This variability during freezing periods underlines difficulties in classifying FT state over wetlands. This can be due to challenges in determining FT reference states and varied emission depths of P-MW frequencies in areas with heterogenous blends of vegetation,

soil, and water. High agreement mid-winter would suggest products agree when wetlands freeze more uniformly. While transitional periods present a significant challenge over wetland regions across products, in cold regions, there remains a clear FT signal due to the abundance of water.

The unique FT classification agreement trends across varied land types indicate the importance in accounting for and understanding the effects of land cover on FT signal. Most notably, challenges exist in FT classification near water, including over snow and ice, as well as in wetland and coastal regions. Other impacts of vegetation should also be considered when determining in which part of the landscape P-MW remote sensing techniques are capturing FT transitions whether it be the soil, snow/ice, surface vegetation, or canopy.

### *Topography*

Mountainous regions are characterized by highly variable conditions stemming from both the impedance of emitted P-MW energy and prevalence of highly varied climates which occur with large sub-grid scale changes in elevation. These higher elevations with increased topographic variability also experience an increased potential for more frequent freezing and thawing processes, even in the lower latitudes. Sub-grid variability coupled with challenges in collecting consistent radiometric measurements across complex terrain has resulted in difficulty capturing ephemeral FT events (*Dunbar et al. 2018*). **Figure 26** illustrates the agreement between SMAP (L-band) and SSM/I (Ka-band) FT products in relation to the standard deviation in elevation. A clear trend of decreasing product agreement with increasing topographic complexity emerges.

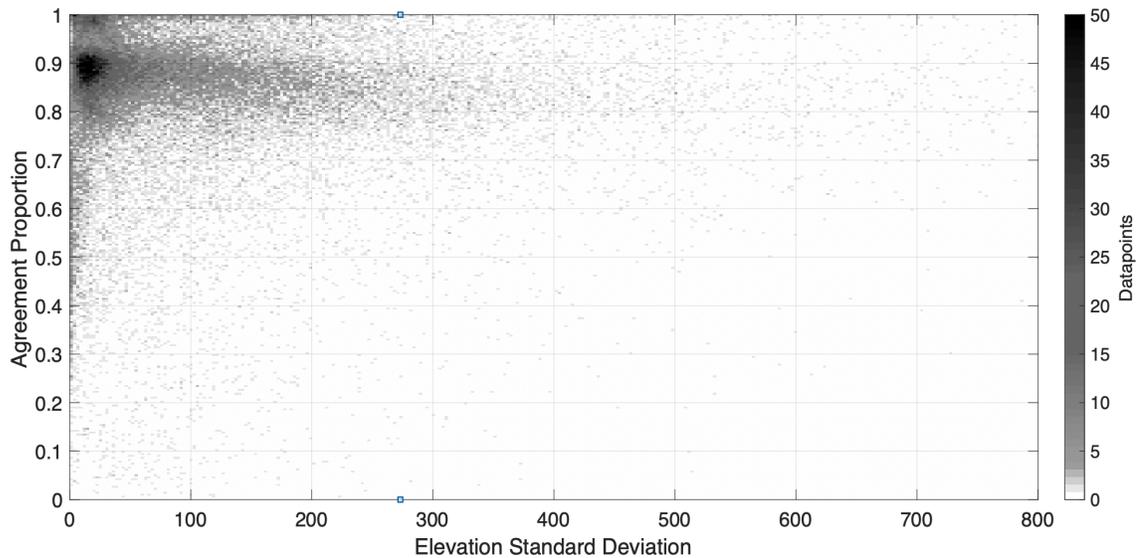


Figure 26 Density plot of product agreement proportion compared to sub-grid elevation variability

Reductions in agreement from regions with standard deviations of 10 meters to greater than 500 meters are shown to occur (**Table 11**). Similar to previous results, areas with lower standard deviations in elevations (0 - 10 m), which commonly occur in low lying coastal regions, are shown to have the low  $A_p$  (78.9%). Conversely the regions with the highest variations in elevation (> 400 m) exhibit the most variability in  $A_p$  through time. The vast majority of the FT constrained land mass is comprised of regions with elevation standard deviations ranging from 10 – 200 meters (74% of comparison cells). These parts of the domain have  $A_p$  near the global mean and are characterized by less seasonal variability in classification. Therefore, these results underline the difficulty in measuring FT especially in mountainous and relatively flat areas adjacent to large water bodies.

**Table 11 Product agreement proportion by elevation standard deviation groupings (in meters)**

Elevation Standard Deviation (m)		0-10	10-50	50-100	100-200	200-300	300-400	400-500	>500
Cells		4,577	15,513	7,030	7,778	3,236	1,372	616	769
Mean		0.789	0.856	0.836	0.832	0.827	0.812	0.817	0.811
Standard deviation		0.081	0.087	0.084	0.085	0.087	0.093	0.105	0.108
99% Confidence interval		.787 - .791	.854 - .857	.834 - .838	.830 - .834	.825 - .830	.809 - .814	.814 - .820	.807 - .814
Percent of Period Classified Frozen	SMAP-FT	21.2%	26.4%	30.2%	30.7%	27.7%	27.1%	23.4%	20.3%
	SSM/I-FT	30.1%	29.2%	33.2%	33.8%	30.4%	28.8%	25.4%	24.8%

***Assumptions and Limitations***

This study makes a number of assumptions in regard to the accuracy of input temperature (GLDAS), landcover (MODIS), elevation (GMTED), and climate classifications (Köppen-Geiger). All of these data products have some degree of error, either in the spatial representation of land cover and climate classifications or in temperature and elevation values, which may include spatial errors as well. When data is upscaled for this comparison the uncertainties in these datasets are amplified. Due to the large number of datapoints and the use of 99% confidence intervals for the mean across variables, the effect of error in these datasets is not expected to significantly alter the presented results. Additionally, the use of timeseries analysis has shown similar classes following analogous temporal patterns which helps corroborate general trends within these results. Another limitation is the use of binned freezing days and topographic variability data to present agreement proportion ( $A_p$ ). Changes in these boundaries may impact results, though the full distributions have also been presented to remedy this. It is also important to note that different seasonal thresholding algorithms (NPR and single

channel) will introduce some degree of classification variability, which is not comprehensively assessed in this study. In addition, the masking of coastal regions, removing  $T_b$  values impacted by water bodies before processing, and having consistent “never frozen/thawed” masks across product types is expected to enhance the agreement between global FT products.

### **2.3 Chapter Outcomes**

Proper understanding of strengths and weaknesses of remote sensing approaches remains essential to their effective and accurate application. Especially early in the lifetime of any new remote sensing product, performance investigations are critical to better understand limitations and identify opportunities for improvement. The efforts detailed within this chapter illustrate efforts along these lines, resulting in several conclusions regarding the current state of freeze/thaw remote sensing.

First, existing freeze/thaw records from the Earth System Data Record (Ka-band microwave) and SMAP (L-band microwave) are compared to surface skin, air, and soil temperatures. These comparisons were performed with the intent of better understanding how leading classifications relate to surface temperatures and to improve characterization of which surface components are represented. The foremost results are as follows:

- *ESDR and SMAP freeze/thaw products have notably different relationships to temperature variables and large uncertainties around the freezing point (0 °C)*
- *Air temperature is shown to have the closest relationship to all products, however SMAP was comparatively more representative of soil temperature*

- *When converted to binary freeze/thaw states using thresholding, skin temperatures are found to have the lowest agreement with FT products*
- *Products tended to define less frozen area compared to observations of surface skin temperature and modeled skin, air, and soil temperatures*
- *All products have improved performance in the afternoon relative to temperatures. Clear regional, seasonal, and band (Ka- vs. L-band) variability in these relationships is also observed*
- *False freezes are identified in SMAP-FT records and not in FT-ESDR which had more distinct class associated temperature distributions. MODIS may provide a method to mask regions, reducing these false classifications. Note: Recent updates to SMAP products (Version 2, 3) have largely removed this issue, improved resolution, and have extended records through 2021*

Next, a detailed assessment directly comparing FT records relative to climate, land cover, and elevation is completed. All surface variables assessed are shown to exhibit unique time-series variations in agreement between the FT datasets. As the only global comparison of its kind, the uncertainty in FT classifications is quantified to enable product improvements and to detail potential reasons for classification divergence. Additional outcomes from these investigations include:

- *Globally, current records are found to agree on 83.5% of freeze/thaw classifications from 2015 – 2017*

- *Information on climate variability can be used to help tune algorithms, especially across areas of varied moisture and those with significantly different freeze/thaw regimes (e.g., continental and near polar vs. temperate and transitional)*
- *Coastal regions and those dominated by liquid water, snow and ice are shown to have significantly lower agreement relative to the global mean (55.8% to 77.6%)*
- *Forested and non-forested lands have different uncertainty regimes, suggesting the use of vegetative cover may help to refine and better constrain freeze/thaw algorithms moving forward*
- *ESDR records are found to be very similar across comparisons*

Both studies highlight challenges over complex terrain, demonstrating the importance of both improving resolution and the potential to tune new algorithms with the consideration of topography. The implementation of fractional classification approaches may also help to improve representation of surface freeze/thaw states over these regions. Critically, band differences were also highlighted as Ka-band classifications experienced more variability between overpasses and better sensitivity to the surface layer. This is in contrast to L-band observations which remained more consistent and relatively more indicative of soil state. For example, ESDR (Ka-band) had increased frozen classifications in cold continental zones relative to L-band, suggesting that L-band may observe thawed soil under the snowpack when Ka does not. Different approaches thus have the potential to provide information on different components of the surface and fusing these records could provide enhanced and unified estimates of the surface states by utilizing their unique strengths and weaknesses.

While these are important studies for understanding current products, without the use of validation networks on the ground, we cannot determine which approaches may be performing better. Efforts to combine field work and in-situ observations provide a logical next step to enhance freeze/thaw classification algorithms. The ability to improve our understanding of complex FT processes through the combination of satellite observations, modeling, and in-situ measurements appears promising.

### **3 EXPLORING VARIABLES FOR FREEZE/THAW ESTIMATION AT FINE SCALES**

This chapter delves into a deeper exploration of the variables which control freeze/thaw processes at finer (sub-grid) scales and how these can be adequately captured by remote sensing approaches. More specifically, efforts included herein explore how new variables such as land surface temperatures, cover, climate, and topography, among others, may be useful for better characterizing freeze/thaw states. These efforts also provide further exploration into how these factors relate to surface states, by attempting to de-aggregate the contributions to the freeze thaw signal into its component parts (i.e., soil, snow, water, vegetation) and the thresholds at which a frozen signal may begin to dominate the microwave and thermal emissions. Prior efforts have explored freeze detection over focused study regions. Here, we combine field work campaigns with modeling, ground-based sensing sites, unpiloted aerial vehicle (UAV) observations, and snow sampling to reinforce and expand upon these efforts. Without a clear understanding of the physics and sensitivities of existing approaches at fine scales, the value and applicability of these products for operational assessments (i.e., of the growing season, runoff potential) is greatly reduced. Outcomes illustrated herein remain critical to the development of next generation of enhanced freeze/thaw classifications and represent the next step in foundational work towards achieving this goal.

### **3.1 Improving Freeze/Thaw State Classification Using Sub-grid Temperature**

Accurate representations of surface freezes can be exceedingly important due to the significance of seasonal freeze/thaw cycles within the Earth system. However, consistent freeze definitions with clear physical meaning can be challenging to achieve. This section closely explores the utility of resolution enhancement, draws conclusions towards freeze/thaw variability within a given sensor footprint, as well as how this may affect classifications. This includes studying thresholds that define the point at which frozen classifications begin (i.e., 50% of cell area), which may vary with variables such as land cover or moisture content. The results provide insight to improving surface freeze/thaw states and their definitions by examining the characteristics governing existing freeze classifications.

#### **3.1.1 Background and Motivations**

Frozen conditions are defined by the state change of water from liquid to solid, generally occurring around 0 °C. While simple in theory, accurate FT characterization remains challenging. As discussed, FT observation across large areas is constrained by limited in-situ surface skin and soil temperature observations, especially at high latitudes, making the use of satellite products crucial to spatially distributed FT cycle monitoring. Though, due to coarse resolution satellite-observations, the heterogeneous nature of the land surface, microwave (MW) observing bands with fundamentally different emitting layers, and varied classification algorithms, binary FT classifications come with considerable uncertainty (*Rowlandson et al. 2018; Johnston et al. 2019; Johnston et al. 2020*). Grid spacing of existing FT classification records remain on the order of 25 km.

Even as enhanced products have been developed ( $< 10$  km), they rely on much the same input MW data as native resolution products which is limited by band properties and antenna geometry. At these resolutions, MW FT observations capture a diverse blend of emissions due to varied surface soil characteristics, moisture states, and microclimates resulting in the coincident the presence of both frozen and liquid water (*Rowlandson et al. 2018; Roy et al. 2017a; Roy et al. 2017b; Prince et al. 2019*).

We have previously highlighted frozen classification variations between lower frequency (SMAP, L-band 1.41 GHz) and higher frequency (FT-ESDR, Ka-band 36.5 GHz) radiometer observations. Compared to the FT-ESDR, SMAP-FT has been shown to observe FT transitions deeper into the soil, to be less impacted by vegetation and surface snowmelt, though also being increasingly affected by arid conditions and having limitations due to a larger sensor observational footprint (*Rautiainen et al. 2012; Johnston et al. 2019; Johnston et al. 2020; Wang et al. 2020a*). Additionally, regions with ill-defined FT seasonality (i.e., mid-, and lower latitudes), persistent snow cover, abundant surface water, varied topography, and complex vegetation/soil layers have been identified as challenging regions to accurately assess surface FT state generally due to presence of mixed FT conditions (*Ulaby et al. 1981; Kimball et al. 2009; Podest et al. 2014; Roy et al. 2015; Kraatz et al. 2018*). Varied threshold approaches considering only brightness temperatures (FT-ESDR), or polarization differences (SMAP-FT) can contribute to a fundamentally different FT classifier. As a result, frozen states can be poorly characterized, referring to the combined ‘landscape’ FT state of which the definition may vary between platforms. For these reasons, multi-sensor data fusion may

prove useful in clarifying the definition of frozen states, through leveraging this information into multiple freeze types, as well as fractional and enhanced resolution products. In addition to examining the relationship between temperature and FT state, this study thus aims to determine the utility of, and how, information at the sub pixel level can improve the characterization of freeze/thaw state and inform algorithm advancements.

To understand the observational challenges and improve FT state observation, higher resolution data provide valuable ancillary information. Sub-grid cell temperature, land cover, soil type, and even climate classifications have the potential to help characterize both the threshold at which FT products determine a pixel as frozen and the characteristics that may dominate the emitted MW signal and govern binary FT classifications. Algorithms, which are the first to use higher-resolution surface temperature and topographic data, have been previously implemented over China with promising results (*Zhao et al. 2017; Hu et al. 2017*), but have not been applied over North America. As such, we include hourly land surface temperature observations from the Geostationary Operational Environmental Satellite-East (GOES-13, ~5 km) to assess inter-FT-pixel temperature variations (*Freitas et al. 2013*). Additionally, in many cases where in-situ observations are scarce and satellite products are challenged, land surface models provide a valuable input to understanding global-scale and spatially distributed physical processes at enhanced temporal and spatial scales. Therefore, through a model ensemble approach, we assess the ability of operational land surface models to accurately simulate landscape temperatures and physical processes, while also illustrating the associated FT classification uncertainties. In this study, we attempt to further characterize

the relationship between FT classification and environmental variables by answering the following questions:

- 1) How do sub-grid scale surface temperature and land surface characteristics relate to microwave FT classifications, both at the regional scale and grid scale?*
- 2) Where are the greatest temporal and spatial uncertainties in MW-based FT classifications relative to each other, as well as to surface temperatures?*
- 3) What are the sources of these uncertainties?*
- 4) How can higher resolution land surface variables be applied for the improvement of FT classification algorithms?*

These efforts provide an extension to previous analyses over North America by introducing observations from ground sites and exploring characteristics over the extent of the continent.

### **3.1.2 Research Methods**

North America (25 °N – 72 °N, 52 °W – 169 °W) is selected as the study region, as it encompasses a wide range of land cover, topography, and climate types. In addition to satellite FT classifications, land surface model (LSM) outputs from NASA's Snow Uncertainty Ensemble Project (SEUP), as well as Copernicus derived hourly land surface temperatures (LST) from GOES-13 are available across the domain. The study period extends for 20 months, or two full seasonal FT periods from September 1 – June 30, 2015 - 2017.

### ***Freeze/Thaw Products and Algorithm Variations***

Freeze/thaw products as used in **Chapter 2.2** are utilized herein and consist of binary state estimates. Specifically, the MEaSURES Global Record of Daily Landscape Freeze/Thaw status Version 4 (or FT-ESDR, *Kim et al. 2017*) and the SMAP Version 2 products (*Xu et al. 2018*). In assessing how these approaches relate to partial freezes, recent SLAP experiments (*Rowlandson et al. 2018*) shed light on sub-pixel temperature variability, showing that generally >60% of in-situ surface temperature measurements tended to be below freezing before SMAP deemed the pixel as frozen. However, there were also instances in which SMAP indicated frozen conditions when no in-situ sites indicated 5 cm depth soil temperatures below 0 °C. In these events, the surface layer was observed to dominate the observed emission. Seasonal thresholding algorithms (STAs) as used by SMAP-FT and FT-ESDR are generally effective for FT classification but do come with limitations as underlined in previous sections. Still, these products also rely on fundamentally different classification approaches, including single channel  $T_b$  (FT-ESDR) and inter-band polarization ratio (SMAP-FT).

### ***Land Surface Models***

To capture a wide range of model physics and encompass several operational modeling suites, we use land surface model (LSM) outputs from the NASA SEUP project (*Kim et al. 2020*). This effort, originally focused on quantifying snow water equivalent and the associated uncertainties over North America, includes outputs of FT relevant variables related to energy exchange (shortwave/longwave radiation), snow depth/density, as well as surface radiative temperature and soil temperature of various

layers. Simulations were run in the NASA Land Information System (LIS) using 5 km grid cell size from 2009 – 2017 with four LSMs and three different meteorological forcing inputs. The LSM's include Noah version 2.7.1 (Noah 2.7.1) (*Ek et al. 2003*), Noah Multi-parameterization land surface model version 3.6 (Noah-MP) (*Niu et al. 2011; Yang et al. 2011*), Catchment Land Surface Model version Fortuna 2.5 (CLSMFv2.5) (*Ducharne et al. 2000; Koster et al. 2000*) and the Joint UK Land Environment Simulator (JULES) (*Best et al. 2011*). Each of these models was forced using input data from the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA2) (*Molod et al. 2015; Gelaro et al. 2017*), Global Data Assimilation System (GDAS) (*Derber et al. 1991*), and European Center for Medium-Range Weather Forecasts (ECMWF) (*Molteni et al. 1996*). The resulting 12-member ensemble provides a range of outputs from models used operationally at major modeling centers around the world (*Kim et al. 2020*).

The two variables of interest include:

- 1) Surface radiative temperature ( $T_{\text{rad}}/\text{RadT}$ ), representing the surface skin temperature, and
- 2) Top layer soil temperature ( $T_{\text{soil}}/\text{SoilT}$ ) which represents the mean temperature of the 0 – 10 cm soil layer (JULES, Noah-MP, and Noah 2.7.1) and the 0 – 2 cm soil layer for CLSMFv2.5

Whilst a shallow soil layer (0-2 cm) generally results in increased temperature variability compared to a thicker layer (0-10 cm), an investigation over the North American domain showed CLSMFv2.5  $T_{\text{soil}}$  falling within 1°C of the ensemble mean

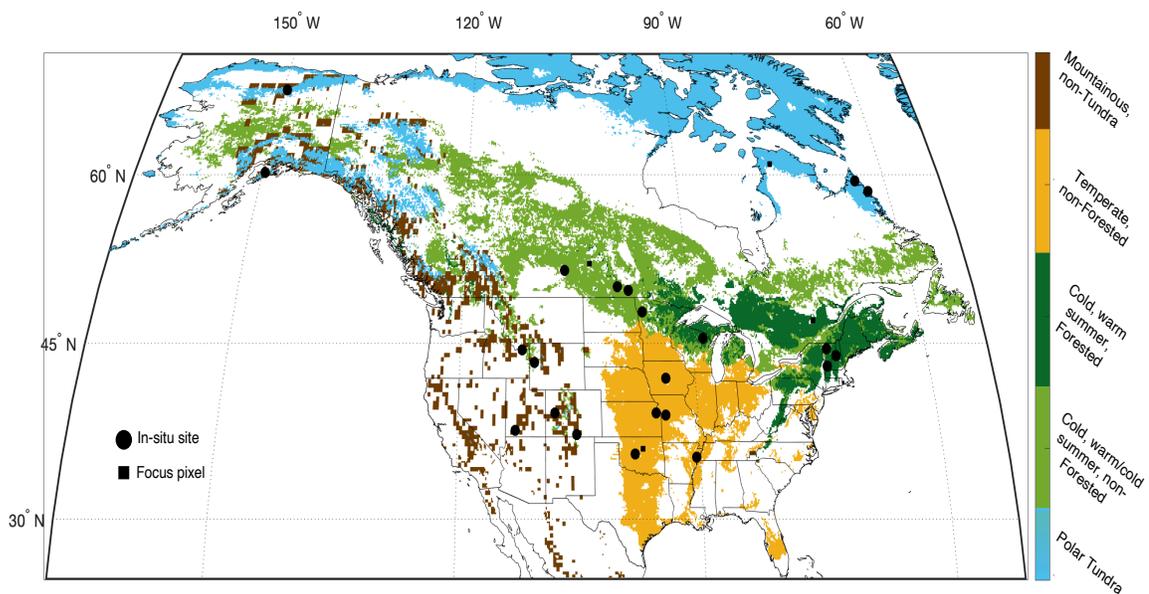
during the cold season and is not the warmest, nor the coldest model during the study period. Radiative surface temperature outputs from Noah models are described as the aggregate surface skin temperature which can represent a mixture of canopy, soil surface, and snow temperatures. Similarly, JULES  $T_{\text{rad}}$  represents the effective radiative temperature (surface skin) and cannot exceed 0 °C when snow remains on the ground. In CLSMFv2.5, over non-snow-covered locations,  $T_{\text{rad}}$  represents surface skin temperature as determined by an area-weighted average of the temperature across different moisture and vegetation states within the computed area (*De Lannoy and Reichle 2016*). In this study, we do not explicitly compare snow states as they relate to FT and include both snow and non-snow-covered pixels, illustrating a mixture of results from snow free and snow on conditions. Notably, for timeseries analysis, ensemble mean snow depth is also included.

**Table 12 Summary of ensemble validation results**

		Temperate, Non-Forested	Tundra	Cold, Warm Summer, Forested	Cold, Non-Forested	Mountainous, Non-Tundra
$T_{\text{rad}}$	Bias (°C)	1.34	-0.28	1.17	1.82	0.94
	RMSE (°C)	3.72	4.88	3.33	3.93	5.33
$T_{\text{soil}}$	Sites (N obs)	5 (10,721)	4 (9,203)	4 (9,033)	5 (11,391)	4 (8,651)
	Bias (°C)	-1.97	6.66	-2.94	-5.49	0.50
	RMSE (°C)	4.18	9.28	5.22	7.34	4.25

Ensemble mean  $T_{\text{rad}}$  and  $T_{\text{soil}}$  is compared against GOES land surface temperature observations and soil temperatures from ground observing sites across five unique regions of the domain (**Table 12**). Observations included are from various sites within the Soil Climate Analysis Network (SCAN, *USDA 2020a*), the SNOWpack TELemetry

network (SNOTEL, *USDA 2020b*), the Real-Time In-Situ Soil Monitoring for Agriculture network (RISMA, *Pacheco et al. 2019*), and observations made in the Canadian tundra (via SoilTemp, *Lembrechts et al. 2020*). Results indicated an ensemble that is slightly warmer than GOES LST observations ranging regionally from bias of -0.3 °C to 1.8 °C. This is most pronounced in the plains and mountainous regions in winter. Modeled soil temperatures present adequate performance with regionwide biases ranging from -5.5 °C to 6.7 °C (RMSE 4.2 °C – 9.3 °C). Diminished accuracy over the tundra and colder non-forested areas which represent more challenging regions to model. The location of validation sites and extent of focus regions is included in **Figure 27**. The derivation of these regions is based on climate, land cover, and elevation variability classes as described in subsequent sections.



**Figure 27** Focus pixels, representative regions, and in-situ site locations

The complexities of natural processes and meteorological input uncertainties lead to error and uncertainties in LSM simulations. As opposed to the use of individual models, an ensemble approach allows a quantification of these uncertainties, as the ensemble spread generally provides a measure of uncertainty across models and forcing data (*Dirmeyer et al. 2006; Guo et al. 2007; Bohn et al. 2010; Murphy et al. 2020*). Additionally, the combination of several model realizations can result in increased simulation accuracy by allowing model errors to negate each other (*Xia et al. 2012a*). Results presented in this study are computed through an aggregation of ensemble outputs.

#### ***Other Remote Sensing Datasets***

In addition to LSM outputs, satellite observed LST is used to examine sub-FT-pixel scale surface temperatures. Near global 5 km hourly resolution LST is derived by the Copernicus Global Land Surface using a constellation of the Geostationary Operational Environmental Satellites (GOES) and other geostationary satellites (Meteosat Second Generation, MTSAT/Himawari, *Freitas et al. 2013*). Specifically, retrievals from the GOES-13 (East) thermal band (band 4, 10.7  $\mu\text{m}$ ) are used in this study and cover a large portion of North America. Copernicus LST products do not include hourly land surface temperature from GOES-15 (West) during the period, therefore Alaska and northwestern sections of North America are not compared against GOES LST. Estimated temperature uncertainties of the Copernicus products are expected to be within 1°C during night-time and up to 3°C during the day.

**Table 13 Observational and satellite derived datasets**

Data Product	Product ID	Coverage	Resolution		Usage
			Temporal	Spatial	
MODIS/Terra+Aqua Land Cover Type IGBP	MCD12Q1	Global, granuals	2016	0.5° & 500m	Land cover classification
ASTER Global Terrain Model	ASTGTMV003	Global, granuals	2000 - 2013	0.083° & 30 m	Intercell topography and standard deviations
Köppen-Geiger climate classes	Beck_KG_V1_present	Global	1980 - 2016	0.083°	Climate classification
Interpolated mean annual temperature	WorldClim	Global	1970 - 2000	1 km	Mean annual temperature
Copernicus LST, GOES-13 band 4 (10.7 µm)	LST_GLOBE_GEO_V1.2.1	North America	Hourly	5 km	Land surface temperature (LST)
SMAP L3 Radiometer Global Freeze/Thaw State	SMAPL3FTP R16010	Global	~6 AM/PM local solar time	36 km	Freeze/thaw state
MEaSURES Global Record of Landscape Freeze/Thaw	SSMI_37V_v04	Global	~6:30 AM/PM local solar time	25 km	Freeze/thaw state
	AMSR_36V_v04		~1:30 AM/PM local solar time		

Additional input variables (**Table 13**) necessary to characterize land cover include 500 m granules (MCD12Q1) and global 0.05° (MCD12C1, 2016) land cover from the Moderate Resolution Imaging Spectrometer (MODIS) (*Friedl and Sulla-Menashe 2015*) using the International Geosphere–Biosphere Programme (IGBP) classification scheme. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) global terrain model provides high-resolution (30 m) and global elevation (1/12°) to assess the effect of complex terrain on FT classification and characterize focus pixel locations (*ASTER Science Team 2019*). Furthermore, information on climate types is sourced from 1/12° Köppen-Geiger climate classifications, using the updated and enhanced resolution products developed by *Beck et al. 2018* (and updated relative to *Kottek et al. 2006, Chapter 2*). Mean annual temperature climatology is obtained from WorldClim bioclimatic variables at 5 km resolution (*Hijmans et al. 2005*). Where available, higher resolution inputs are utilized on the pixel scale, whereas coarser resolution products are used in the regional components of this analysis.

## ***Comparison Metrics***

### ***Sub-grid Frozen Proportions***

To compare sub-grid temperatures to binary FT classifications we compute the proportion of sub-pixel temperatures below 0 °C, within a given FT product pixel ( $P_i$ ). Similar to **Chapter 2**, this measure determines the proportion of sub-cells below freezing when classified as frozen ( $FZ_p$ ) and thawed ( $THW_p$ ) by satellite FT products. The formulation of these metrics is described by **Equations 10 – 12**, providing indicators of sub-FT pixel temperature variability relative to FT classifications:

**Equation 10 Frozen proportion**

$$P_i(t) = \frac{\sum_{j=1}^k \begin{cases} 1, & T_j < 0^\circ C \\ 0, & T_j \geq 0^\circ C \end{cases}}{k}$$

$k$  = the total number of temperature values within a given FT pixel [ $i$ ] at a given time [ $t$ ]

$T_j$  = the temperature in sub-pixel [ $j$ ]

$P_i(t)$  = frozen proportion within a given FT pixel [ $i$ ] at a given time [ $t$ ]

**Equation 11 Sub-pixel frozen proportion when classified as frozen (FZp)**

$$FZp_i = \frac{\sum_{t=1}^n \begin{cases} P_i(t), & \text{if frozen} \\ 0, & \text{if thawed} \end{cases}}{\sum_1^n \begin{cases} 1, & \text{if frozen} \\ 0, & \text{if thawed} \end{cases}}$$

**Equation 12 Sub-pixel frozen proportion when classified as thawed (THWp)**

$$THWp_i = \frac{\sum_{t=1}^n \begin{cases} 0, & \text{if frozen} \\ P_i(t), & \text{if thawed} \end{cases}}{\sum_1^n \begin{cases} 0, & \text{if frozen} \\ 1, & \text{if thawed} \end{cases}}$$

$FZ_{pi}$  = the mean sub-pixel frozen proportion for a given FT pixel [ $i$ ] when classified frozen

$THW_{pi}$  = the mean sub-pixel frozen proportion for a given FT pixel [ $i$ ] when classified thawed

$n$  = total number of FT classifications within the study period

We employ various temperatures from the ensemble ( $T_{\text{rad}}$ ,  $T_{\text{soil}}$ ) and GOES (LST) as temperature inputs into **Equation 10**. Where high  $FZ_p$  (0.8-1) and low  $THW_p$  (0 - 0.5) generally characterize a pixel as having well defined frozen and thawed states.

### *Three-product Frozen Agreement*

To characterize the differences and associated FT uncertainty across products, FT-ESDR classifications are upscaled to the 36 km SMAP grid by averaging, and then rounding all values within a given SMAP pixel to ensure a binary classification. The three-product agreement proportion ( $Ap_{3FZ}$ ) is calculated across the 20-month study period for both morning and afternoon overpasses at each pixel using (**Equation 13**):

**Equation 13 Three-product agreement proportion**

$$Ap_{3FZi} = \frac{\sum_1^n \begin{cases} 1, & \text{if all 3 frozen} \\ 0, & \text{otherwise} \end{cases}}{\sum_1^n \begin{cases} 1, & \text{if any frozen} \\ 0, & \text{if all 3 thawed} \end{cases}}$$

where n refers to the total number of timesteps in which all products classify a given pixel's (i) FT state.

This method focuses only on periods when at least one product classifies a pixel as frozen to remove cases in which all products are deemed thawed, which can skew the product agreement. This has a considerable effect, especially in the mid- to low latitudes where thawed classifications are frequent in all seasons.

### *Regional and Grid-scale Timeseries Analysis*

To better understand the relationship between coarse resolution FT classifications and the underlying temperature states, the sub-grid frozen proportion ( $P_i$ ) is compared against binary FT classifications at the pixel scale. Five focus locations and larger

representative regions (**Figure 27**) are selected based on variable climate, topography, FT classification uncertainty, and land cover (**Table 14**).

**Table 14 Point location characteristics of product pixels nearest the indicated coordinates**

Focus Pixel (Nearest Coordinate)		Southern Plains (36° N, 97° W)	Tundra (61° N, 76° W)	Northern Forest (47° N, 74° W)	Northern Plains (52° N, 103° W)	Mountainous (39° N, 108° W)	
Köppen-Geiger Climate Classification(s)		Temperate, no dry season, hot summer	Polar Tundra	Cold, no dry season, warm summer	Cold, no dry season, cold summer	Cold, no dry season, cold summer	
Mean Annual Temperature			Cold, dry summer, cold summer	Cold, no dry season, cold summer		Cold, no dry season, warm summer	
Land Cover Description		grasslands (83%), savannas (10%), cropland (4%) & other (2%)	grassland (94%), water (3%), wetland (2%) & other (1%)	mixed forest (90%), woody savannas (4%), deciduous forest (4%) & other (2%)	croplands (38%), woody savannas (28%), grasslands (16%), mixed forests (7%), savannas (5%), other (6%)	grasslands (67%), savannas (18%), woody savannas (6%), croplands (5%) & other (4%)	
Elevation	mean	ESDR	287m	290m	450m	606m	2507m
		SMAP	278m	185m	484m	573m	2047m
	range	ESDR	182 - 359m	170 - 478m	324 - 611m	514 - 683m	1565 - 3410m
		SMAP	165 - 348m	38 - 319m	338 - 694m	482 - 670m	1450 - 3325m
	standard deviation	ESDR	20m	64m	44m	17m	564m
		SMAP	19m	42m	52m	32m	549m

We employ timeseries to investigate the seasonal relationship between sub-grid surface states and their relationship to the observed FT state. This provides insight into the timing of freeze onsets, as defined by satellite-based products, along with information on pixel scale processes that may dominate the MW response over various regions. These regions include non-forested temperate regions (southern plains), forested cold climates with warm summers (northern forest), non-forested cold climates with warm/cold summers (northern plains), polar tundra, and non-tundra mountainous pixels (**Table 14**). Regions are classified using information shown in **Figure 28**.

For comparison, all values for each ensemble member and LST observation within a given focus pixel (and surrounding region) are extracted at each FT product acquisition timestep. For the larger SMAP pixels,  $P_i$  is calculated with 672 – 1056 individual ensemble temperature values and 56 – 104 from GOES LST. The mean

temperature from each dataset is computed at focus pixel locations and bounded by two standard deviations. This enables an assessment of sub-pixel modeled temperature uncertainty and observed temperature variability. For timeseries, SMAP pixel footprints are used to summarize the data and compute mean temperatures, snow depth, and  $P_i$ . In cases when multiple observations are present from FT-ESDR, frozen classifications are provided as a fraction. In contrast, for the focus pixel descriptions (**Table 14**), product specific grids are used.

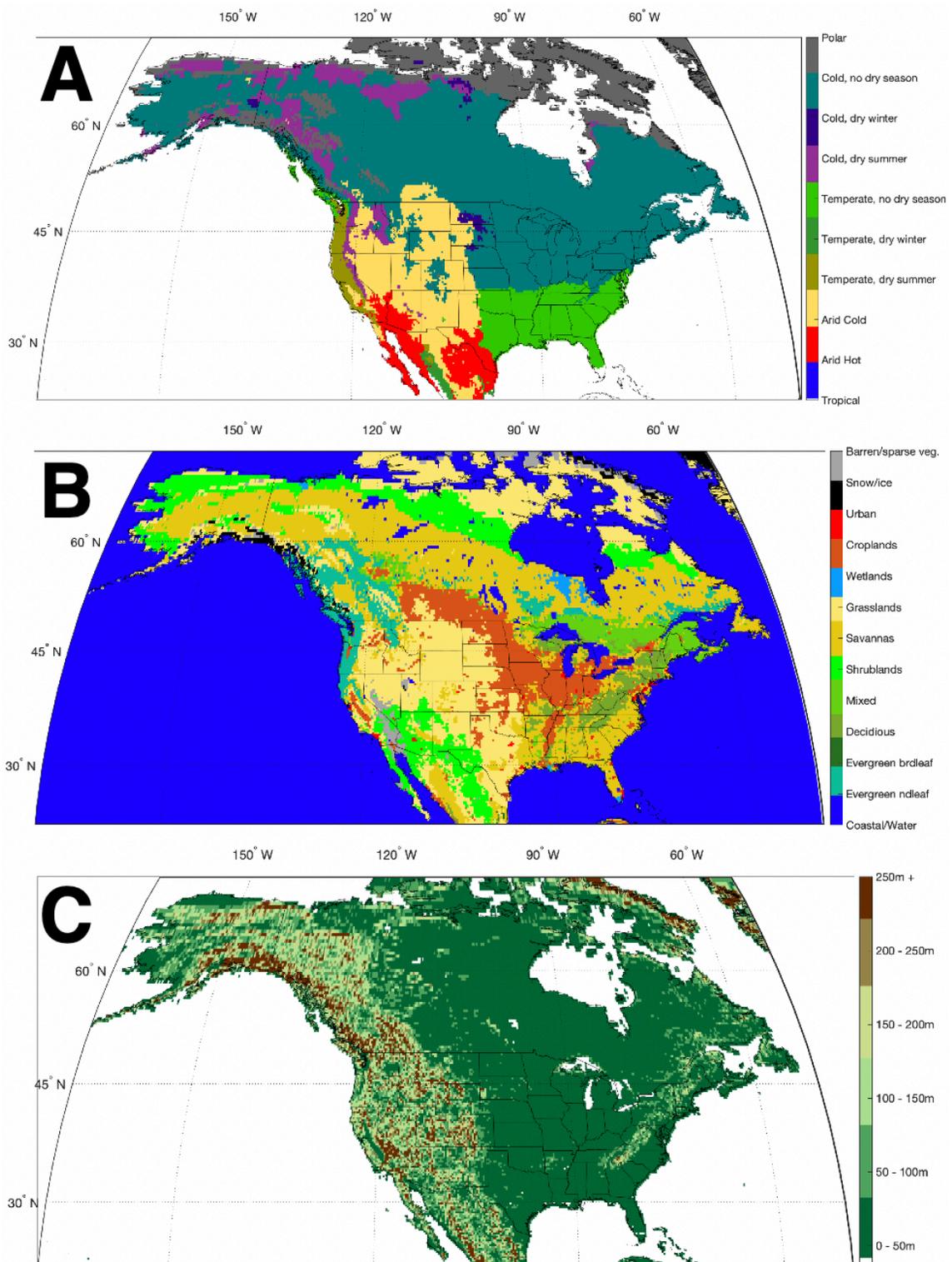


Figure 28 North American (A) climate, (B) land cover, and (C) elevation standard deviations classes

### 3.1.3 Exploring Sub-grid Temperature and Freeze/Thaw

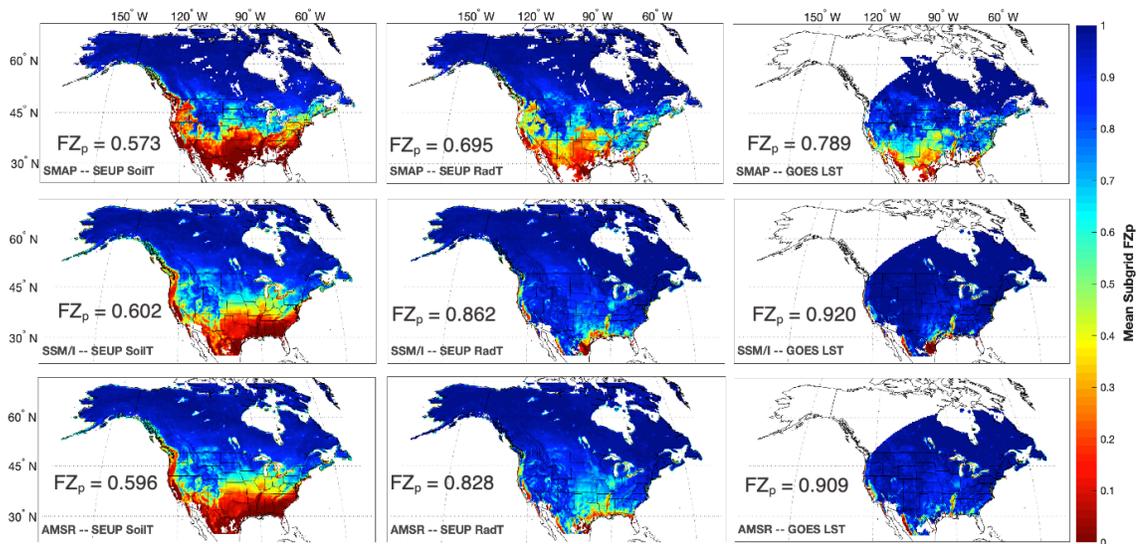
#### *Spatial Results for North America*

##### *Relationship of Freeze/Thaw to Sub-grid Temperature*

Existing FT products have been shown to be particularly sensitive to the near surface temperature (i.e., 2 m air) and moisture states (Rowlandson et al. 2018; Kraatz et al. 2018; Wang et al. 2020a). The results in **Figure 29** reiterate this finding, as observations of frozen conditions over much CONUS occur when significantly less than 50% (yellows/reds) of ensemble sub-pixel soil temperature ( $T_{\text{soil}}/\text{SoilT}$ ) values are modeled to be frozen. In contrast, in northern portions of the domain, FT states are accurately captured by both surface temperature and top layer soil temperature as  $FZ_p$  frequently exceeds 0.95. Over these regions, this indicates that on average nearly all modeled and observed near surface temperatures are below 0 °C at the sub-pixel scale when FT algorithms detect frozen conditions. While remaining largely above 0.80, this is slightly reduced in drier regions downwind of the Rocky Mountains.

For all FT records, comparison against GOES observed LST results in the highest  $FZ_p$  values on average (0.79 – 0.92), indicating that the proportion of sub-grid scale LST below freezing is generally more than 80% when classified as frozen by FT products. Only a few regions with less than 50% of sub-pixels below freezing (when frozen) exist and include the southern Mississippi River basin, California’s Central Valley, and drier regions in the plains and southwest (SMAP only). The magnitude of the difference between  $FZ_p$  values associated skin as compared to soil temperatures (0.32, ESDR to 0.22, SMAP) is considerably higher. This suggests that in many regions, frozen soil does

not provide the dominant part of the observed MW signal. Even though we expect a portion of the observed MW energy to correlate with soil temperatures (especially with SMAP), current algorithm thresholds have mostly been calibrated using synoptic air and surface temperature observations. This may explain the closer relationship of FT products to skin temperature as opposed to soil over many regions.

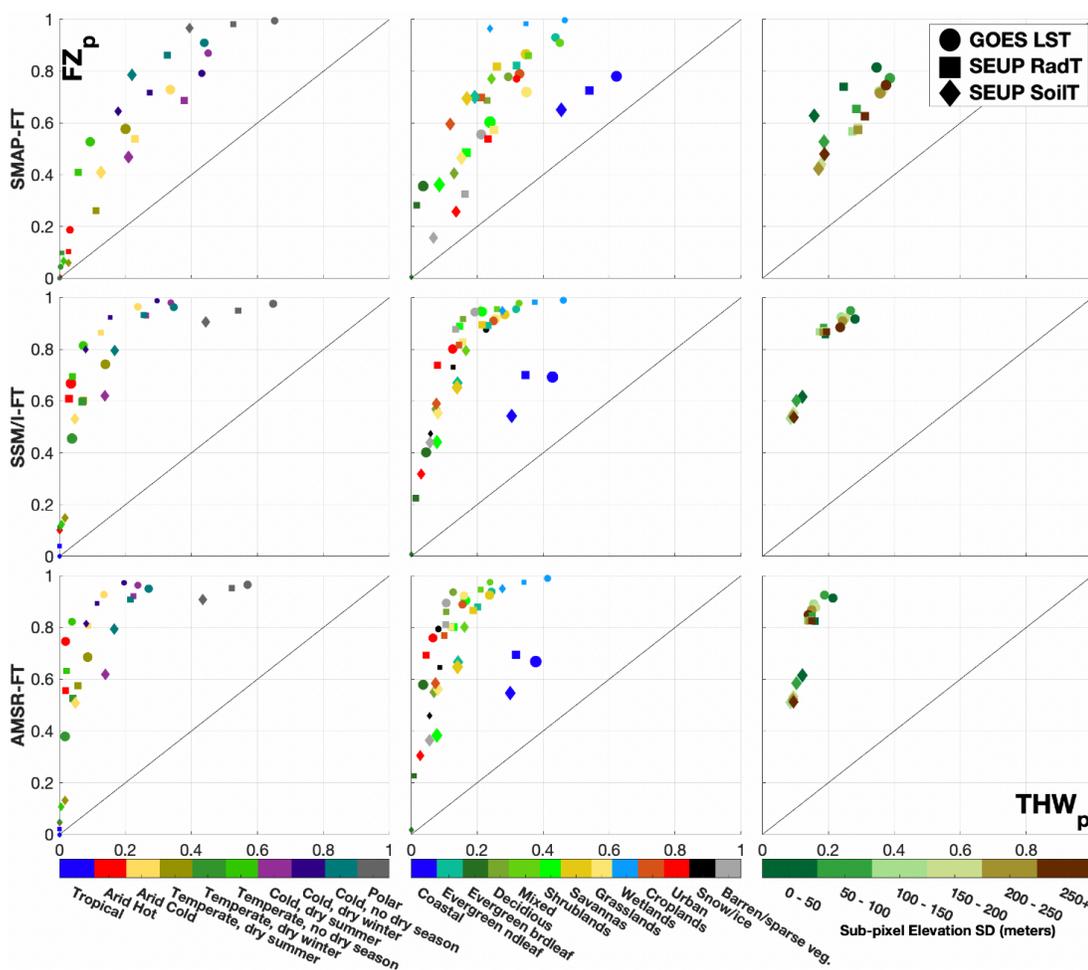


**Figure 29** Spatial mean sub-grid frozen proportion when classified as frozen ( $FZ_p$ ) for soil and surface skin temperatures. Dark blue regions indicate areas in which nearly all sub-pixels have a frozen signature when classified as frozen by FT products. North America domain averaged  $FZ_p$  shown on each sub-plot

Notably, results in **Figure 29** do not indicate the point at which FT algorithms begin to observe frozen conditions, but rather the average  $FZ_p$  across the full study period. Extended cold periods in northern regions can skew the overall mean higher, the  $FZ_p$  at which frozen classifications begin is likely much lower than indicated values. Still, these efforts provide insight into sub-pixel temperature states during classified ‘frozen’ periods and how this can vary substantially across North America.

## Comparisons Across Climate and Land Surface Variables

As an extension to work regarding freeze/thaw classification across varied land surface characteristics (**Chapter 2**), we compare the mean sub-grid frozen proportions associated with frozen ( $FZ_p$ ) and thawed ( $THW_p$ ) classified states to land cover, climate, and derived topographic variability to examine product relationships to these variables directly.



**Figure 30** Sub-grid frozen proportion means when classified by MW FT-products as frozen ( $FZ_p$ , y-axis) and thawed ( $THW_p$ , x-axis). Averages shown for GOES-East footprint by column: climate (left), landcover (middle), topographic variability (right). Marker size indicates standard deviations (larger = higher standard deviation)

The results for each FT product are presented in **Figure 30** and illustrate the averages over the entire 20-month study period. Instances in which pixel values are located near the 45° line indicate poorly distinguished FT states, as the sub-grid proportion of cells below freezing is nearly the same when classified as frozen or thawed by the indicated product. In contrast, points located towards the top left indicate well characterized FT states. These cases suggest that temperatures at the sub-grid level are nearly all frozen when classified as frozen and few instances exist in which thawed classifications occur when most sub-pixels are below freezing ( $THW_p < 0.5$ ).

SMAP is shown to distinguish most poorly between FT states, with less separation between sub-pixel characteristics of frozen and thawed classifications compared to both FT-ESDR records. Across climates, SMAP-FT shows a clear contrast to FT-ESDR, in which  $THW_p$  is frequently less than 0.20 (73% of classes) and  $FZ_p$  greater than 0.50 (73%), whereas only 57% and 50% of classes fall within these ranges using SMAP-FT (**Figure 30, Column 1**). There remains a notable distinction between sub-pixel temperature profiles dependent on climate, with polar/cold regions being well characterized (high  $FZ_p$ ) compared to temperate and warm zones. Still, the high proportion of  $THW_p$  on average in these zones ( $>0.40$  for all temperatures) suggests a considerable number of thawed classifications when temperatures continue to support an at least partially frozen landscape. The reduction in the ability of SMAP-FT to clearly distinguish states is also illustrated when comparing across land cover (**Figure 30, Column 2**) and sub-pixel elevation standard deviations (**Figure 30, Column 3**).

Different from soil temperatures, surface temperatures have a considerably higher proportion of below-freezing sub-pixels when classified as frozen. Generally, lower skin temperatures and increased  $FZ_p$  can be expected due to the increased susceptibility of surface skin temperature to change rapidly with air temperature as compared to soil. This again illustrates that the LST observations are currently the best proxy to FT classifications derived from passive MW-based observations if binary frozen classifications are taken at face value (indicating frozen conditions throughout the entire pixel). This means that the vast majority of LSTs at the sub-pixel scale are below freezing before being classified as frozen by current products. While this remains true when compared to SMAP-FT,  $FZ_p$  derived from ensemble soil temperatures are comparatively closer to skin temperature derived  $FZ_p$  relative to FT-ESDR, having more similar  $FZ_p$  values. Even as  $FZ_p$  values are the highest when computed using GOES LST, this does not necessarily indicate that surface skin temperatures govern passive microwave FT classifications. In most cases while  $FZ_p$  is reduced (RadT to SoilT), so is  $THW_p$ , resulting in minimal change in the difference between temperature profiles associated with frozen and thawed states across layers.

Different sub-pixel temperature signatures associated with FT states are also shown across various land cover types. Over North America, wetland regions are very well characterized whereas coastal regions are not. This may be due to the clear FT signal of Canadian wetlands as they transition from liquid water dominant to ice and snow covered compared to mixed land/water pixels. Additionally, there is also an enhanced variability across coastal regions indicating a wide range of  $FZ_p$  and  $THW_p$  values and

high uncertainty, as captured by the marker size in **Figure 30**. Variability within other classes is relatively small in comparison.

In southerly regions with vegetative cover, most often found in sub-tropical climates (i.e., evergreen broadleaf),  $FZ_p$  and  $THW_p$  values are very low. This shows that in most cases in which a pixel is deemed frozen, only a modest portion of the pixel is likely to be below freezing, whereas soil temperatures are unlikely to indicate frozen conditions at all. Another notable result is that SMAP classifications occur with significantly lower sub-pixel frozen proportion over sparsely vegetated areas and shrublands (0.16 – 0.60; SoilT - GOES LST) compared to FT-ESDR (0.36 - 0.95). The arid nature of these lands suggest that this may be a result of poorly distinguished FT reference states. Thus, sub-grid proportions shown here indicate that over particular regions the proportion of cells actually frozen when deemed as such by P-MW techniques varies widely.

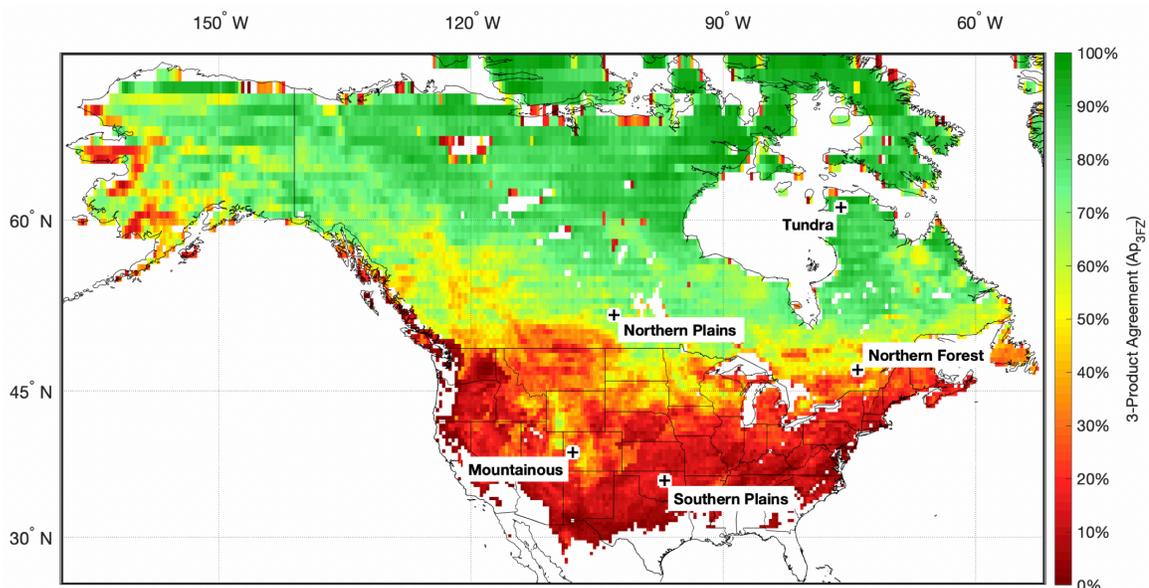
Finally, increased sub-grid heterogeneity in elevation is shown to reduce the average proportion of frozen sub-pixels required for a frozen classification (**Figure 30, Column 3**). Though this is especially true for SMAP-FT, all products show a general decrease in mean  $FZ_p$  from areas with more homogenous topography (<50 m inter-pixel standard deviation) to more complex and mountainous regions (50 - 250 m +).

### ***Freeze/Thaw Sub-regions and Timeseries Results***

#### *Focus Pixel Selection and Product Comparison*

In regions in which several different products commonly agree, we have enhanced confidence in frozen soil and surface conditions compared to regions in which

classifications range widely. Especially in regions with ephemeral freezes, differences are expected in FT state classifications due to varied satellite observation timings. For example, observations around 6 p.m. local time from SMAP and SSM/I have an increased probability of being frozen compared to AMSR observations occurring near the solar maximum (~1:30 pm local time). However, these effects do not account for product agreement of less than 30% over most of the land area south of 50 °N (**Figure 31**). Such results indicate inconsistent frozen classification thresholds across much of North America, especially over CONUS.



**Figure 31** Percentage of agreement on frozen classifications only for all three of SMAP-FT, SSM/I-FT, and AMSR-FT products. Includes September – June for 2015 – 2016 and 2016 – 2017. Focus pixel locations shown

To examine the temporal characteristics between FT products regarding freeze onset conditions, we identify five representative SMAP pixels (36 km) over the study region that encompass major climate regions in North America (**Figure 31**). These

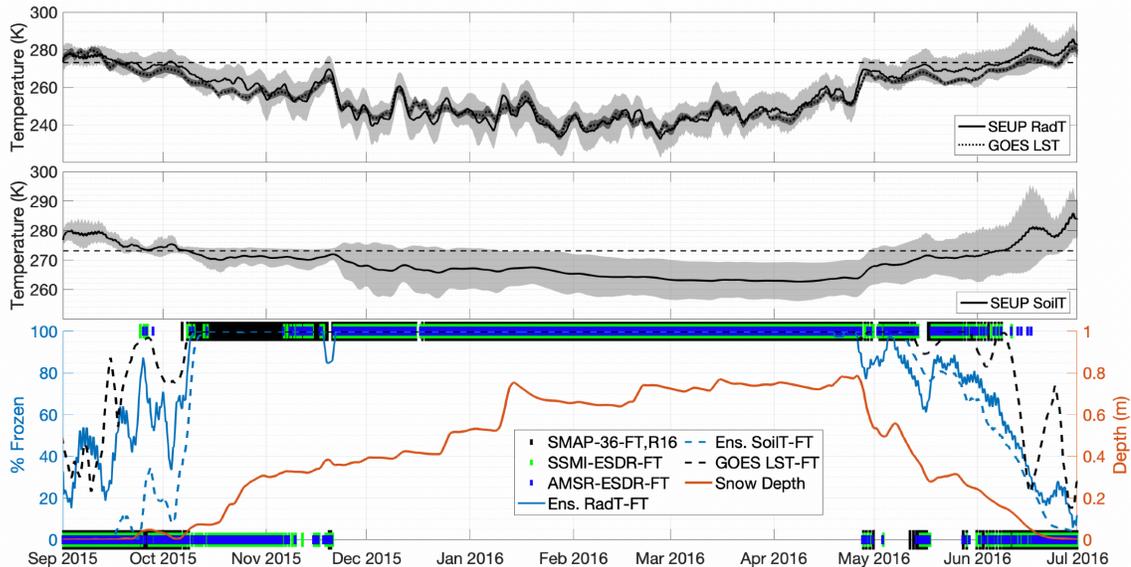
include polar/tundra, cold continental, and temperate climates. In addition, a pixel with considerable inter-cell elevation variability is included to assess the ability of FT products to distinguish FT state under conditions of high sub-pixel temperature heterogeneity. Several landcover classes are also encompassed by the selected regions and include forest dominant, grassland dominant, croplands, and mixed pixels. In addition, focus locations are also selected to represent a wide range of FT classification uncertainty ranging from the three-product agreement proportion ( $A_{p3FZ}$ ) of around 16% in the southern plains to over 82% in tundra. The remaining focus pixels, which include the mountainous, northern forest, and northern-plains locations, have agreement of 20%, 56%, and 68%, respectively.

#### *Grid-scale and Focus Region Results*

**Figures 32 – 36** illustrate a single 10-month period of temperature characteristics and FT classifications, which is long enough to capture the full frozen season even at high-latitudes. Whereas the numerical analysis is based on two of these periods from September through June (2015-2016, 2016-2017), the inter-annual FT onset and temperature characteristics are very similar. Therefore, we present a single FT season (2015-2016), though summary statistics (**Figure 37**) include both periods across all regions with similar climate and land cover (**Figure 27**).

**Figure 32** presents the seasonal FT characteristics over a pixel representative of a sub-arctic or tundra climate. Both temperature and FT products indicate an extended frozen season stretching from early October into June with continuous snowpack building through April before melting into the end of this period. This provides an example of a

well characterized site, as frozen classifications begin as soon as all sub-pixel soil and surface temperatures (GOES and SEUP) drop below the freezing point.



**Figure 32** Freeze/thaw state to temperature comparison for Tundra pixel. Shaded area denotes two standard deviations above and below the mean, timeseries are smoothed using 3-day averages, and dashed lines in top two plots indicate the freezing point

In contrast, **Figure 33** illustrates the sub-pixel characteristics across a temperate pixel located in Oklahoma, U.S.A. At this location, there are few clearly discernable points at which all products indicate frozen conditions outside of the period between late December and early January. Soil temperatures are never modeled to be frozen dominant during the entire period. This is in clear contrast to the results over the tundra, in which frozen classifications were not prevalent until most ensemble members indicated frozen surface and top layer soil temperatures. In the temperate non-forested regions, the SEUP ensemble also shows much stronger agreement with observed temperatures and reduced variability as compared to other focus locations (tundra, mountainous), indicating a

reduction in model temperature uncertainty. Still, frozen classifications appear to be relatively ambiguous, occurring during diurnal FT events in which less than half of sub-pixel temperatures fall below freezing while also not-occurring during periods at which soil temperatures are modeled the coldest (e.g., mid-January).

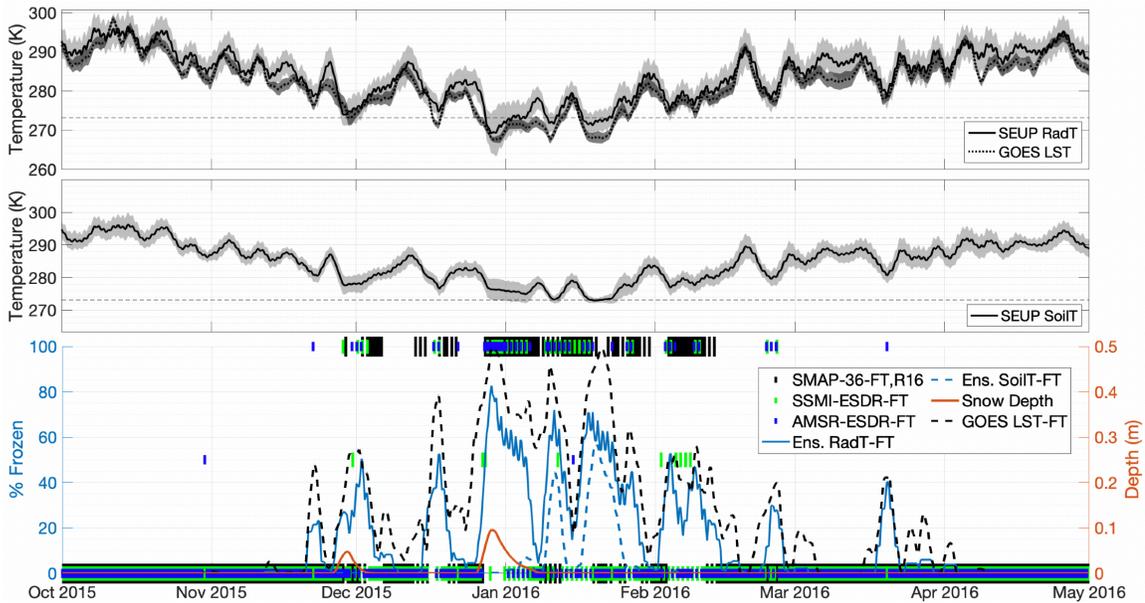


Figure 33 Same as Figure 32, but for Southern Plains pixel

Mountainous pixels are characterized by high inter-grid cell topographic variability and seasonal snowpack, which result in a considerable range of both skin and soil temperatures within the footprint of a SMAP pixel (**Figure 34**). Interestingly, as sub-freezing surface skin temperatures begin to encompass near 40% of the pixel area and FT-ESDR records begin to indicate diurnal freeze events, SMAP-FT no longer detects frozen conditions. Only when nearly all sub-pixels (>80%) are shown to be below freezing does SMAP-FT again classify the pixel as frozen. This also coincides with the

only period in which FT-ESDR records observe frozen conditions during both morning and afternoon overpasses. Finally, during the snowmelt onset period in mid-February, FT-ESDR observes diurnal FT cycling, while SMAP frozen classifications cease all together suggesting considerable differences in the surface components observed by FT classifications over such regions. Uncertainty in early and late season freeze classifications directly also correspond to the snow accumulation and ablation periods.

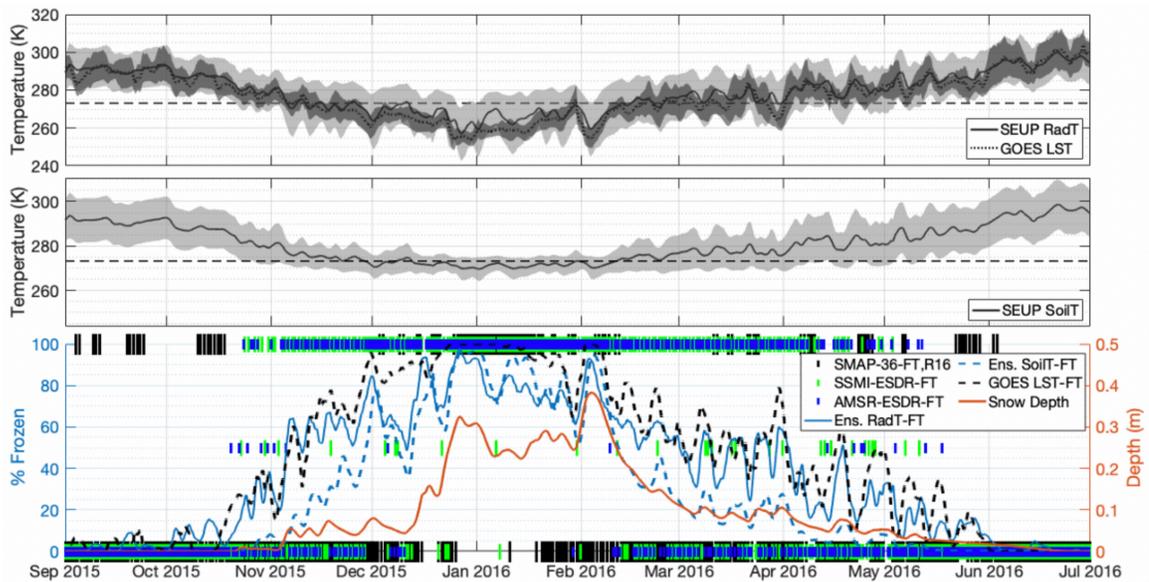
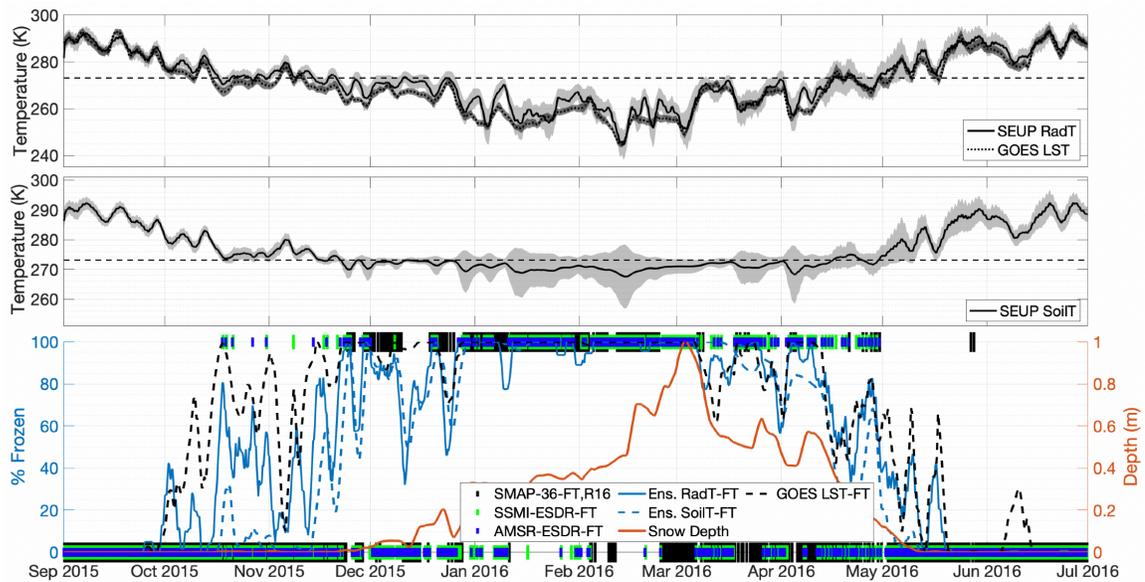


Figure 34 Same as Figure 32, but for Mountainous pixel

**Figures 35 and 36** capture FT classifications over pixels including a forested region north of Montreal, Canada (northern forest) and within the grasslands of the Canadian province of Saskatchewan (northern plains), respectively. These pixels have similar temperature profiles characterized by large temperature fluctuations in the shoulder seasons, before having consistently sub-freezing temperatures from December

into March. Mean average annual temperatures range from -5 to 5°C indicating regions with extended frozen periods and clear seasonality, an assertion supported by the timeseries of FT classifications. Generally, seasonality is well captured by FT products within both regions as frozen classifications rarely occur with less than 80% of sub-pixels below the freezing point. Over the northern plains, this is especially true and illustrated by strong agreement in FT classifications from December into the spring (**Figure 36**). In the northern forest, SMAP classifications are clearly linked to the snow accumulation and ablation periods. This contrasts with FT-ESDR products in which frozen classifications were observed earlier in the season (and later), before (after) modeled SoilT are shown to fall below (increase above) freezing (**Figure 35**). FT-ESDR records appear to be strongly influenced by daytime melt occurring above the snow in the canopy, or at the air-snow interface.



**Figure 35** Same as Figure 32, but for Northern Forest pixel

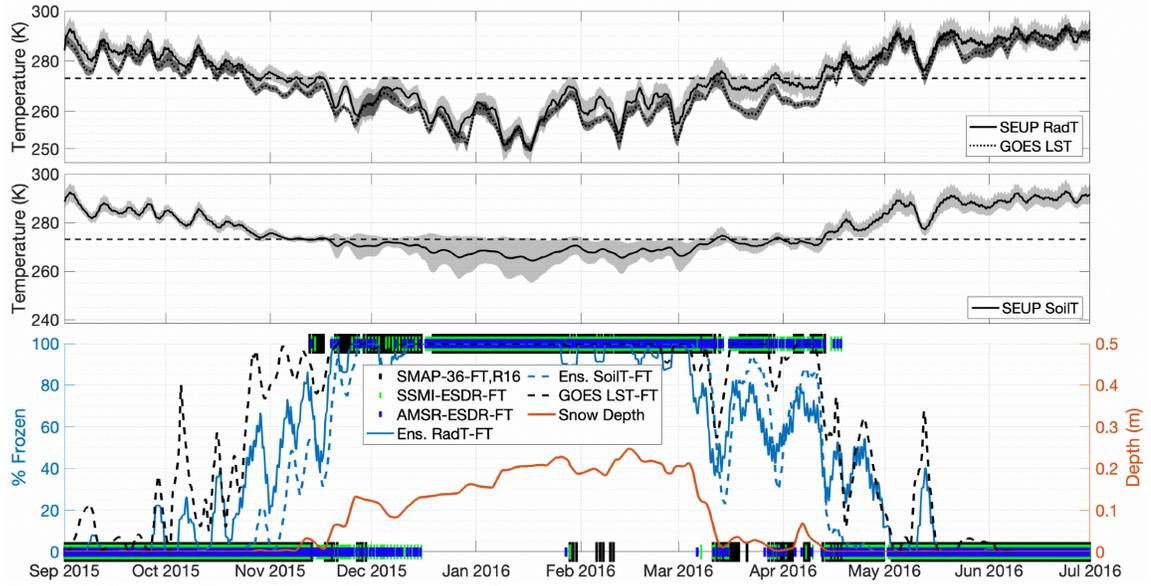


Figure 36 Same as Figure 32, but for Northern Plains pixel

Finally, to quantify the sub-pixel characteristics during the onset of MW frozen classifications, we examine the distribution of  $FZ_p$  within each characteristic region.

Figure 37A presents these results as averages for the lowest 1%, 5%, 10%, 25%, and 50% of  $FZ_p$  values, as well as the overall regional averages (100%).

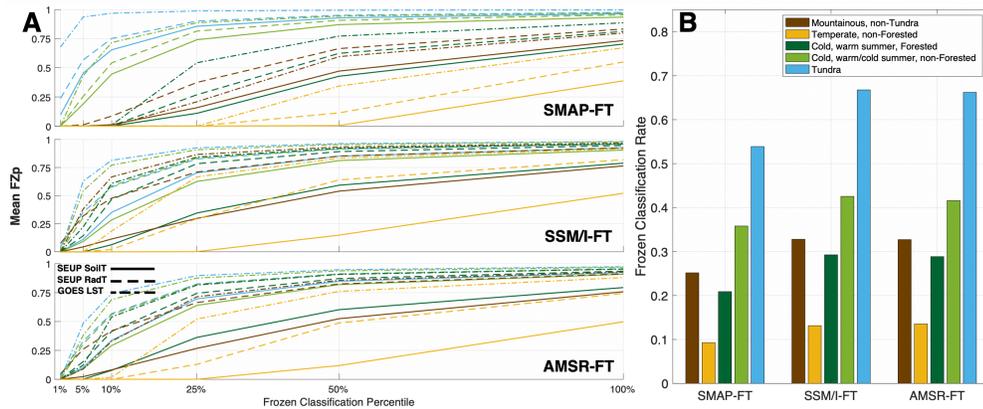


Figure 37 Summary of results for each focus region (shown in Figure 27). (A) Mean sub-grid frozen proportion when classified as frozen for the 1,5,10,25,50, and 100% of the lowest  $FZ_p$  values and (B) the proportion of frozen classifications to total classifications during the study period

Excluding far northern portions of the domain, the proportion of sub-pixels below freezing around onset remains below 10%, indicating that FT classifications may be dominated by the presence of proportionally small frozen areas. These differences between SEUP ensemble outputs and GOES LST observations may also suggest model biases. Still, while results support generally good performance in the tundra, shoulder season freezes are still characterized by uncertain temperature states and low  $FZ_p$  when using modeled soil and surface temperatures ( $FZ_p < 0.5$ ). This can be explained in part by a considerable model warm bias in soil temperatures across the tundra region ( $\sim 6.5$  °C). Particularly over the forested, mountainous, and temperate regions, a clear difference between skin and soil derived  $FZ_p$  values is prevalent. Temperate pixels are shown to be the most poorly characterized, having very low  $FZ_p$  values for all observed and modeled temperatures.

**Figure 37B** presents a comparison of the frozen classification rate during the period. Very similar frozen periods across ESDR products are shown, suggesting that varied retrieval timings have little effect on Ka-band classifications. However, differences between SMAP and FT-ESDR frozen classification rates are considerable across all regions. Indicating a shortened frozen period across all regions, as L-band NPR derived FT classifications are fewer by 4% (temperate, non-forested) to 13% (tundra).

### ***Discussion***

Each region detailed presents different FT detection challenges. While records diverge around freeze onset (October – November) in the case of the tundra and northern portion of the domain, the accuracy and definition of frozen classifications is clear,

indicating near surface soil and skin temperatures uniformly below freezing even at finer scales. However, with SMAP-FT especially, and over the more temperate and non-forested regions (i.e., southern plains), FT classifications remain imprecise and have extended frozen periods during times in which frozen conditions are not supported by either ensemble or GOES observed temperatures (e.g., early December, **Figure 33**). The results suggest that improvement in FT references will be critical to introduce more consistency into FT classification in regions that are characterized by transient and diurnal freeze events.

Over mountainous terrain (**Figure 34**), frozen conditions in a small portion of the mountainous grid cell can induce a frozen classification. This presents a case when enhanced resolution or fractional approaches would be useful to improve FT characterization. Also, in locations with persistent seasonal snowpack, depending on the onset of insulating snow cover, near surface soil temperatures have commonly been observed to gradually approach 0 °C, before remaining at the freezing point until the snowpack has melted. This means that even across many cold regions, soil temperatures can remain marginal making FT characterization increasingly challenging. We hypothesize that snow melt coupled with the insulating effect of the snow results in a layer of unfrozen soil that is observed by longer wavelength SMAP L-band, but not by FT-ESDR Ka-band products.

In cold forest and northern plain regions there are considerably different vegetative cover and precipitation patterns. The forested regions are characterized by higher annual snowfall and dense vegetation compared to the relatively arid and sparsely

vegetated northern plains. This leads to a considerable divergence in FT classifications between products during the melt season (March – May) over the forested region compared to that of the northern plains. FT uncertainty in both focus pixels is shown be related to snow accumulation. A shortened FT season for SMAP (especially in forested regions), may indicate different emission characteristics observed by SMAP, compared to Ka-band. Thus, illustrating the increased sensitivity of shorter wavelength derived FT-ESDR records, which continue to observe diurnal changes in the FT state of the surface (while SMAP tends to follow soil temperatures). These complex interactions between the vegetation, snow, and soil surface present challenges at explicitly detecting soil FT states as opposed to a general ‘landscape’ FT status.

Over many regions existing FT algorithms are shown to identify freeze onset under inconsistent temperature conditions. Meaning, surface temperature states that may indicate a frozen classification in one region may not result in a frozen classification in another, even within FT products derived using the same observing MW band. Regions with less clear FT seasonality (e.g., southern plains) and more sub-pixel temperature variability (e.g., mountainous) experience increased FT uncertainty and require a significantly smaller proportion of sub-pixels to be below freezing for algorithms to indicate frozen conditions.

### ***Study Limitations***

Several study limitations are important to be aware of when interpreting these results. Factors such as varied sensor retrieval timings and radiometer footprints can induce errors between FT products and temperature data, especially in regions

characterized by ephemeral freeze events and topographic variability. Over mixed pixels, the smaller radiometer footprint of ESDR products can contribute to improved characterization and is not explicitly considered. Also, by examining a limited number of timeseries, incorrect generalizations can be made about larger classes of pixels sharing similar characteristics. Though, by examining spatial characteristics as well, we hope to minimize the impact of this on the study conclusions. Most critically, it is important to note that models used in this study are not observations nor considered as such. Instead, they are used to inform a reasonable range of sub-pixel temperature states. Whereas SEUP ensemble members are forced with reanalysis products that do use observations such as GDAS, MERRA2, and ECMWF, outliers in the data can impact the proportional comparison methodologies employed in this study. As such single biased model or forcing input can influence the presented results.

Surface temperatures and microwave derived FT states utilizing changes in surface dielectric properties are also fundamentally different observations. While imperfect, the use of kinetic temperatures to infer surface FT states can help fill gaps in resolution (of P-MW) as well as directly relate these products to observed surface temperature states. In our estimation of FT using kinetic temperatures, we assume that modeled or observed temperatures below the freezing point indicate frozen conditions. While this assumption will not always be true, it provides a reasonable FT proxy. Ideally, direct satellite observation of soil temperatures would be preferable, but is not realistic, thus integrating LSMs and other satellite observations provides critical inputs to better defining frozen and thawed classifications.

### ***Algorithm Improvements***

These results draw into question not only the consistency of FT definitions across different products, but within the same product spatially. FT uncertainty illustrated through analysis into various sub-regions asks several questions about what current frozen classifications represent, such as:

- *Is it the point at which frozen water exists in any part of the landscape or the majority of parts?*
- *When plant productivity is inhibited by frozen soils and persistent sub-freezing temperatures?*
- *Or, when the seasonal snowpack sets in?*

Among other possible definitions, it seems likely that it is a combination of these, depending on location, sensing band, and time. As such, a priority of these efforts includes asking questions regarding what frozen classifications currently reveal and what they should reveal, then tailoring remote sensing and data fusion strategies to capture relevant information. This work proposes the implementation of multiple data streams to achieve this goal, as information from high resolution kinetic surface temperatures can be combined with P-MW derived snow and soil properties to infer the specific frozen components of the surface (i.e., snow, soil, surface water). Integration of multiple MW bands such as L- and Ka-band, may allow freeze types to be distinguished between the surface and soil by combining information from fundamentally different emitting layers. New algorithms should rely on quality ground observations of surface and soil temperature as well as snow cover to help validate classifications. A transition from

binary representations of FT is also expected to contribute to improvements in classification by providing more realistic representations of freeze conditions.

Existing FT classifications are valuable; however, these results suggest region-specific improvements to these products that could expand their relevant applications moving forward. The importance of developing regionally variable algorithms is underlined, as similar observed MW responses in different regions may indicate completely different surface FT states. For example, current binary classifications can indicate frozen conditions both in portions of the sub-arctic (e.g., with soil frost depths >1 meter) and lower latitudes (e.g., with discontinuous surface frost). States across which associated energy, hydrologic, and biogeochemical implications are substantially different. This work outlines the need for a unified definition of what a frozen classification indicates, such as soil state, and proposes that the implementation of multiple sensing bands will be critical in achieving this goal.

### **3.2 Exploring Field-scale Thermal, Soil Freezing, and Melt Processes**

While many of seasonal freezing processes are examined on a macro-scale (i.e., 10's of kilometers) over global or continental domains, it is important to understand the driving factors which control melt and refreeze at fine scales. These fine, or 'field' scales, represent sub-basin conditions in which agricultural activities and spatial variability in topography, moisture, snow re-distribution, and land cover can be significant. These scales are what contain habitat (*Parkin 1993*), dominate watershed response (*Ala-aho et al. 2021*), have agricultural implications (*Rowlandson et al. 2018*), and are generally relevant to capturing realistic representations of surface freezing processes. In a way,

these explorations focus on the microclimates experienced by life on Earth, including ourselves. This includes the complex interplay between components like snow, vegetation, moisture, soil, solar exposure, and the associated energy balances. Satellite observations cannot explicitly capture these components and processes either due to spatial resolution limitations (e.g., km-scale microwave observations) or infrequent revisit times (e.g., 16-day cycles) for higher resolution imagers.

Observations from point sites and aerial imagery provide a new perspective. The popularization of Unpiloted Aerial Vehicles (UAVs) or ‘drones’ have filled the observational gap between ground studies, airborne campaigns (e.g., SLAPEX, SMAPVEX), and orbiting sensors by enabling on-demand capture of surface thermal characteristics at unparalleled centimeter-scale resolutions. The growth in the use of UAVs for scientific applications has been fueled by improvements in drone platforms (i.e., batteries, automated flight), availability of a variety of light weight sensors, flexible deployment capabilities (e.g., over dangerous terrain and/or with frequent repeat), high resolutions, minimal atmospheric influence, and affordability when compared to traditional in-situ, airborne, or satellite remote sensing methods (*Whitehead et al. 2014; Turner et al. 2014; Harvey et al. 2016; Manfreda et al. 2019*). A wide range of sensors have been used on board UAVs in addition to thermal imagers, including multi-spectral, LiDAR, and visible band cameras with a myriad of applications within hydrology (*Resop et al. 2019; Wigmore et al. 2019*) and vegetation monitoring (*Berni et al. 2009; Turner et al. 2014; Espinoza et al. 2017; Lin and Habib 2021*). For these reasons, UAVs have

become a highly effective method of which to explore hydrology, micro-climates, and surface thermodynamics.

The primary objectives of this project are to:

- 1) *Explore soil and snow temperature dynamics, their controls on freeze depth, and satellite freeze/thaw observations using in-situ observations from Grand Mesa, Colorado*
- 2) *Explore surface temperature variability and its distribution using aerial imaging from UAVs at fine scales (<1 meter)*
- 3) *Modeling and characterizing land surface features which drive temperature variability at the field-scale*

### **3.2.1 Study Methods**

#### ***Grand Mesa Study Site***

Located in western Colorado (39.0° N, 108.1° W), Grand Mesa spans over 1,400 km<sup>2</sup> and lies predominantly above 3,000 meters in elevation (**Figure 38**). Formed through volcanic activity and erosion, it is the largest Mesa in the world. This region was used for several UAV flight campaigns over the period from 2019 into 2020, indicated in **Figure 38**. The area is ideal for field investigations, as it is well characterized as part of the NASA Snow Experiment (**SnowEx 2017, 2019-20**, <https://snow.nasa.gov/campaigns/snowex>) activities, has a considerable winter snowpack, relatively flat topography, and a mixture of forested and unforested areas. In terms of vegetation, Grand Mesa consists of primarily Spruce, Fir, Aspen, Pinyon Pine, and Juniper trees as well as open shrub and grasslands (*USGS 2017*). Four (4) long term meteorological stations (two shown in **Figure 38**) are

maintained across the Mesa as part of SnowEx and George Mason University field activities (2016 – present), along with a two (2) SNOTEL observing sites. UAV flights were flown primarily in November 2019, over the western portion of the Mesa, and in February 2020 in the northeastern portion near the long-term observing sites indicated. The sites record standard meteorological variables such as air temperature, wind speed, humidity, and precipitation in addition to being fully instrumented for radiation balance and multi-layer soil and snow temperature observations.

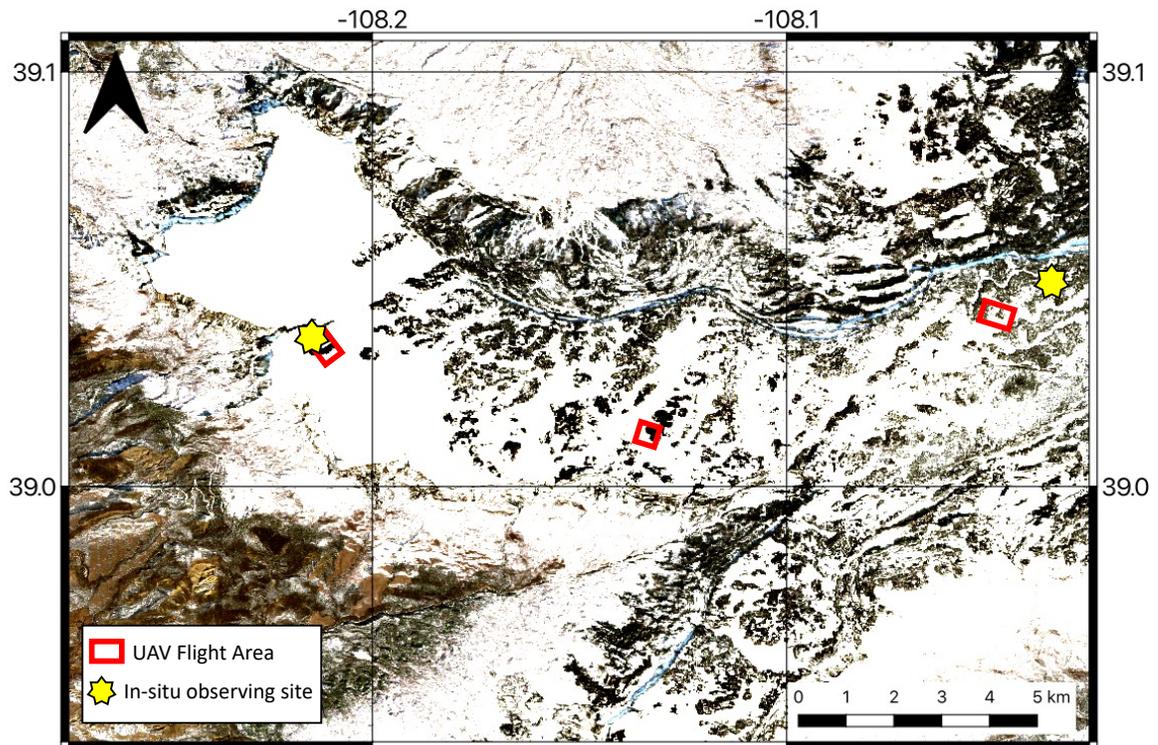


Figure 38 Grand Mesa, CO Study area. Base map: Landsat 8 imagery (February 14, 2020)

## ***Datasets and Processing***

### *UAV Collections*

A DJI Inspire 2 quadcopter is used to carry the sensor payload, which includes an integrated visible-band (RGB) camera (Zenmuse X4S, 20 megapixels) and a FLIR Vue Pro R radiometric thermal imager (7.5 – 13.5 mm, 640 x 512 pixels). Both cameras are gimbal mounted to reduce blur in imagery and to ensure consistent incidence angles during flight (i.e., NADIR). The FLIR sensor consists of an uncooled microbolometer, which allows for low power consumption and remains light weight for UAV integration. However, compared to temperature controlled TIR imagers, these are less well characterized and forfeit accuracy. These sensors collect TIR energy emitted, reflected, or transmitted by an object, thus if emissivity and surface properties are known, they can remotely estimate target skin temperatures.

Though, these values can be difficult to estimate, and observations remain affected by humidity, incidence angle, proximity to the target, camera temperatures, flight speed, and other sources of emitted or reflected energy (*Sugiura et al. 2007; Mesas-Carrascosa et al. 2018; Kelly et al. 2019*). Thus, these sensors require regular flight to flight calibration. While TIR sensors are typically calibrated using a temperature controlled black body target, this is not realistic in the field, thus self-recording calibration targets and in-situ surface snow temperature observations are used to develop an empirical relationship relating the observed camera digital numbers to the surface kinetic temperature. We employ a linear regression fit calibration using observed target temperatures and digital values in each individual image. This approach is similar to

efforts by *Sheng et al. (2010)*, *Harvey et al. 2016*, and *Kelly et al. (2019)* which relied on various ground targets (e.g., water, painted control targets) to perform radiometric corrections. A global correction function is derived for each flight, in order to limit errors from changes in between flight environmental conditions (*Sagan et al. 2019*). Product validation efforts using outputs from this approach have shown image bias on the order of 0.5 °C and RMSE < 2 °C, with very high correlations to surface temperature observations.

#### *Satellite Temperature Data*

Observations from the Advanced Baseline Imager (ABI) on board the GOES-16 (GOES-East) geostationary satellite provide sub-hourly observations in 16 spectral bands (0.45 – 13.6 μm) from across the majority of North and South America. These observations, among others from geostationary orbiting platforms (Meteosat Second Generation and the Multifunction Transport Satellite), have been utilized to derive a near global hourly land surface temperature (LST) product (*Freitas et al. 2013; Martins et al. 2019*) at approximately 5 km resolution. Furthermore, observations from the Moderate resolution Imaging Spectroradiometer (MODIS, MOD/MYD11 Version 6.1, *Wan et al. 2021*) are also leveraged for LST observation with twice-daily global coverage across two satellite platforms (Terra/Aqua). Together, these instruments provide four-times daily coverage with equatorial crossings at approximately 10:30 AM/PM (Terra) and 1:30 AM/PM (Aqua) local solar time with approximately 1 km spatial resolution. Products are derived from seven MODIS TIR bands using the generalized split window approach with stated accuracy on the order of 1 °K (*Wan et al. 1999*). While other high-

resolution imagers such as Sentinel-3 (SLSTR, 1 km), the Landsat 8 Operational Land Imager (Landsat 8 OLI, 120 m), the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER, 90 m), and the Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS, 70 m) have also been leveraged for LST estimation, these data products have much longer global repeat cycles (>14-days) and were not available during this particular study period.

### ***Surface Temperature Modeling and Feature Importance***

#### *Gaussian Process Regression Modeling*

Supervised machine learning (ML) approaches for regression problems remain valuable tools for capturing complex non-linear relationships and thus have strong predictive capacity. ML approaches using supervised learning have been previously applied to the prediction of land surface temperature and shown the ability to downscale satellite estimates using predictors such as topography (*Wang et al. 2020b; Xu et al. 2021*). For the prediction of fine-scale variability in LST, we apply a Gaussian Process Regression (GPR) framework (*Rasmussen 2004*). GPR models are a sub-set of probabilistic ML models which define a prior distribution (i.e., a Gaussian distribution) of a random variable, then update this distribution depending on evidence (i.e., predictors) forming a posterior distribution. While named after the Gaussian, or normal, distribution, this non-parametric approach is effective for capturing complex non-linear relationships through the combination of several Gaussian models and is effective for predicting unseen data (*Barkan et al. 2016*). As a probabilistic approach, GPR models

perform well at avoiding overfitting as they can generalize patterns in a Bayesian space and represent the uncertainty associated with a given prediction.

For our application, GPR was used to estimate the magnitude of temperature changes relative to scene averages in UAV collected thermal imagery across mixed snow and vegetated areas with varied topography. This was done using various predictors detailed in the following section. This approach performed notably better at recreating surface temperature observations as compared support vector machines (SVMs), multiple linear regression, decision trees (DT), decision tree bagging (or random forests, RF), and simple neural networks (NN).

The feature importance or ‘weighting’ of inputs to a LST model are of the utmost importance when assessing the physical drivers of temperature variability. Herein we estimate the importance of various predictors by computing their value within derived GPR models. It is of note, that the importance of a given predictor can change depending on the model type, and metric used. As such, values computed here specifically refer to the importance of various predictors within the GPR framework. To compute the relative feature importance, we evaluate model performance by iteratively training models without a given predictor and computing model RMSE and  $R^2$ . This is repeated  $N$  times ( $N = 100$ ) across all predictors. The average difference between the models including a given predictor to those excluding the same predictor is taken as a metric of feature importance. Specifically, the change in the  $R^2$  between model predictions is taken to quantify importance.

### *Predictive Features*

To assess drivers of observed surface temperature variability at study plots on Grand Mesa, several land surface and topographically derived predictors are used. Specifically, 1-meter resolution digital surface models (DSMs) of Grand Mesa from the Airborne Snow Observatory (ASO, February 2, 2020; *Painter et al. 2016*), land cover, and solar position. Land cover was partitioned into 2-classes using UAV collected visible imagery over the study site, representing vegetation and snow. Estimates of solar position are computed using the National Renewable Energy Laboratory's Solar Position Algorithm (*Reda and Andreas 2008*). This algorithm relies on time, date, elevation, latitude, and longitude to estimate solar azimuth and zenith angles at a given point of interest (center of a grid cell). Using information regarding solar position and topography, several additional indices are computed as model predictors. Surface roughness ( $R_{\text{srf}}$ ,  $R_{\text{tpi}}$ ), topographic prominence (localtopo), slope, curvature, aspect, shading and hill shade (*Schwanghart 2021*), as well as reflected and incident shortwave radiation are used in this context. Herein, these variables are assumed as the dominant drivers of small-scale surface temperature variability. Other features such as soil moisture, snow density, soil type and density, as well as vegetation type and health may also affect small scale variability in surface temperatures but are not considered in the context of this work.

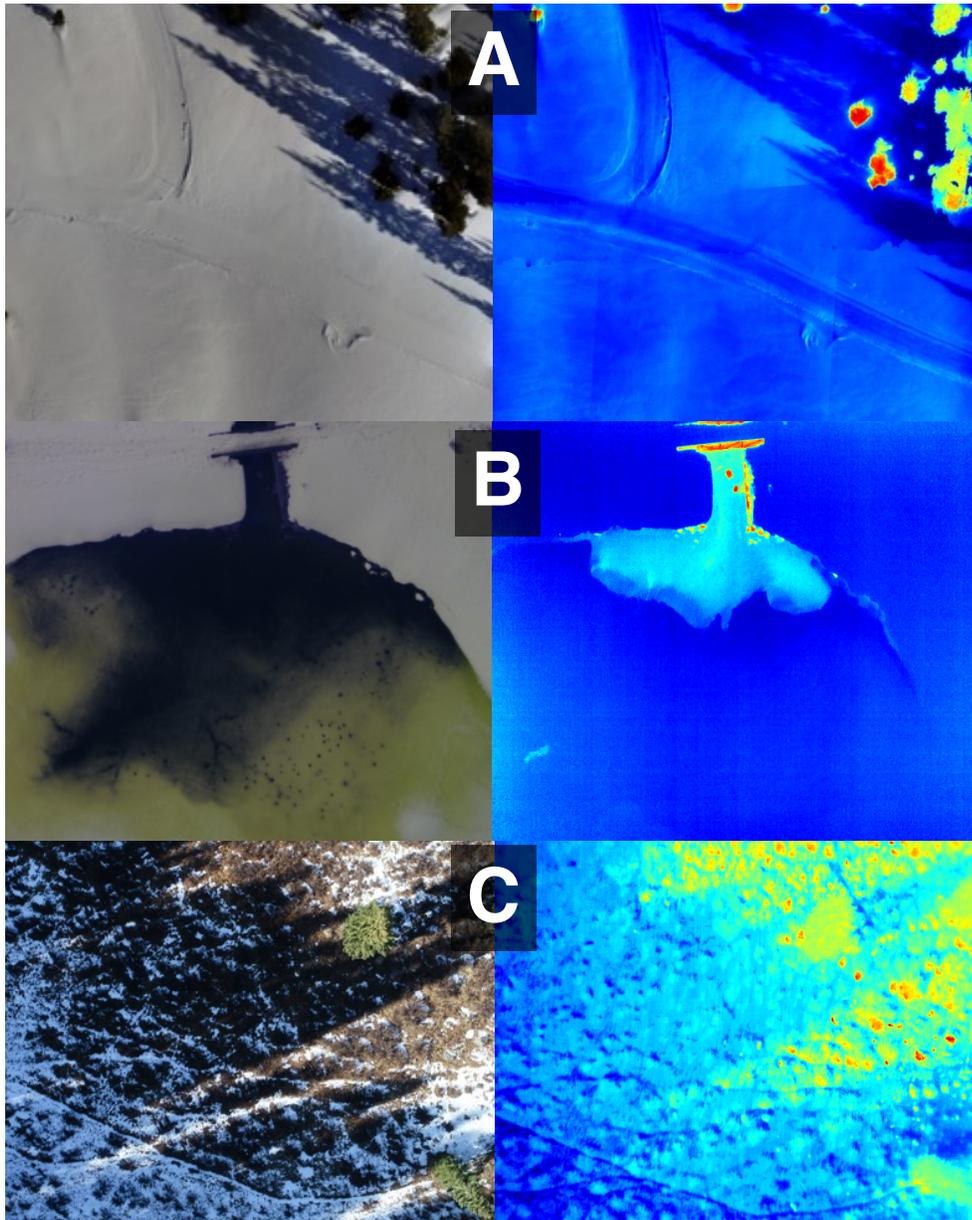
### **3.2.2 Results, Analysis, and Project Outcomes**

#### *Freeze/Thaw Dynamics and Classification Estimates at Grand Mesa*

The land surface in cold constrained regions often consists of a blend of snow, shrubs, bare soil, and tree cover. Surface water in the forms of streams, lakes, and

impoundments are also exceedingly common, particularly at Grand Mesa. To explore the various thermal signatures during cold constrained periods, coincident visible and thermal imagery is collected across various conditions (**Figure 39**).

The first collection (**Figure 39A**) illustrates the effect of shading on snow surface temperatures in a mixed snow-vegetated environment. In this instance, vegetation remains considerably warmer than surface snow temperatures and plays a role in melt processes through limiting solar exposure in forested zones or at the forests edge. As a result, these regions tend to be the last to melt out during thaw events. The reflective properties of snow (i.e., high albedo) and its reduced thermal conductivity, especially of fresh snow, also help to limit surface warming and slow thaw even in periods when temperatures exceed the freezing point. **Figure 39B** presents the TIR profile of a pond outflow. In the thermal band, the difference between open surface water and the surrounding ice and snowpack is clearly discerned. While a much larger portion of the ponds edge appears thawed in the visible band, there remained a relatively thin layer of surface ice. Thermal band sensitivity to a very small surface skin layer (LST) is also illustrated in this case. In mixed snow ice conditions such as this, regions with liquid water remain at 0 °C, whereas snow and ice temperatures can drop much lower. This leads to notably different thermal signatures. Similarly, the emissive properties of water, ice, and snow are known to vary widely, driving the difference in visible bands (dark water vs. bright white snow and ice) and thermal conductivity.



**Figure 39** UAV thermal and visible image pairs captured in varied conditions (A) Mixed snow-vegetated, (B) Ice-open water, and (C) bare ground-snow-vegetation. Blues indicate sub- or at-freezing temperatures

In the shoulder seasons over most FT constrained land area, conditions tend to resemble the mixture of shrubs, soil, snow, and tree cover shown in **Figure 39C**. The presence of several thermal processes dominates freeze/thaw regimes at these scales. This

includes complex shading patterns from canopy and understory vegetation, which can create microclimates. Bare soil and rock can prove especially absorptive to solar radiation compared to snow; however, vegetation is observed to warm at an even faster relative rate and is elevated from the soil surface increasing its exposure to solar radiation.

Temperature variability present at these scales further complicates the interpretation of remote sensing signals which aggregate emissions from large regions (~625 km<sup>2</sup>, for 25 km pixel size) into single observations. These variations have implications for changes in the underlying soil freeze/thaw state.

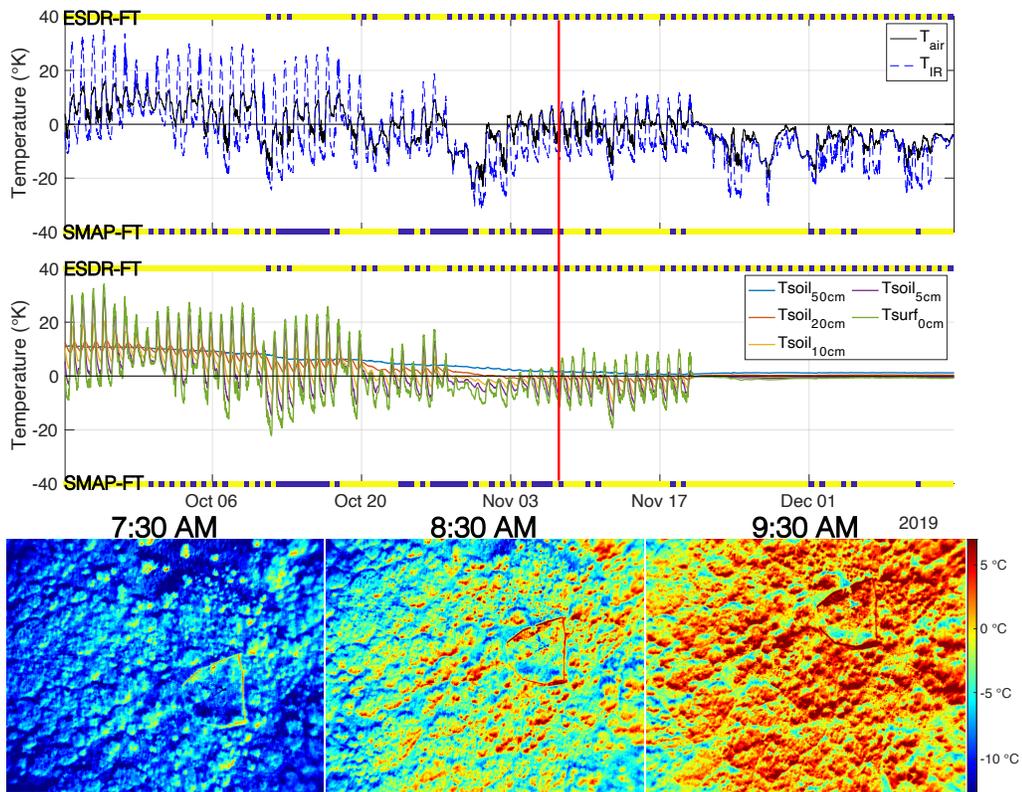


Figure 40 (Top) Air and skin temperature, (Middle) multi-layer soil temperatures, and (Bottom) UAV TIR imagery over observing site November 8, 2019 (red line). FT classifications indicated frozen (blue), thawed (yellow)

**Figure 40** illustrates the comparison between point observations and the corresponding satellite freeze thaw classifications for the western portion of Grand Mesa (at the observing site indicated in **Figure 38**). Notably, coarse satellite-based classifications are representative of the land surface atop Grand Mesa and parts of the surrounding lowlands. FT-ESDR records accurately capture diurnal transitions in the region beginning in October and continuing through the snow accumulation period beginning in late November. SMAP-FT provides significantly different estimates of FT states over the same region. First, by indicating continuous frozen conditions in early October when soil temperatures to 10 cm depth begin to fall below freezing at night. These estimates while accurate in the nighttime hours are invalid during the day, when soil, surface, and air temperatures climb well above the freezing point ( $> 10\text{ }^{\circ}\text{C}$ ). Similar to conclusions in **Chapter 2**, the presence of snowpack is shown to significantly alter the between band FT response. With Ka-band capturing daily melt and refreeze cycles in early winter, while the SMAP algorithm suggests primarily thawed ground over this site until the end of the year. In this case, SMAP may observe a thin, thawed layer of soil at the snow air interface, as soil temperatures remain around  $0\text{ }^{\circ}\text{C}$  deep into winter. Near surface soil temperatures to around 10 cm depth are shown to mirror daily air temperature cycles prior to snow accumulation (**Figure 40**). While deeper into the soil layer, these temperature changes occur more slowly on seasonal scales (50 cm depth). Generally, phase changes in the top few centimeters of the soil are expected drive changes in the surface emissions and brightness temperatures observed by satellite instruments.

Based on observations on the morning of November 8, 2019, when air temperatures remained below freezing prior to the final flight, soil surface temperatures are shown to vary considerably in space, depending on variability in surface cover and shading (**Figure 40**). The rapid warming occurring from observations taken earlier in the morning (7:30 a.m.) to those taken two hours later (increases of  $>10$  °C) can illustrate rapid FT transitions which can occur in the shoulder seasons and even during cold periods. Solar loading provides an abundance of energy to facilitate phase changes that can lead to frequent FT cycling as observed by both satellite products. Even so, aerial imagery suggests that certain portions of the landscape may remain frozen throughout the day, as other ‘hot-spots’ such as vegetation may appear to retain above freezing temperatures through the night. UAV observations and multi-layer soil temperature observations provide a good illustration of how FT representation using a binary indicator remains a vast oversimplification. The potential differences between 2-meter air temperature and phase changes, especially in the soil, remains clear (**Figure 40**). The FT signal and soil temperatures can largely depend on when snow cover arrives. For example, snowfall following an anomalously cold and dry period soil tends to result in persistent frozen soil throughout the winter beneath the snowpack. In contrast, if the snowpack sets in when the soil is relatively warm, it can remain thawed or at freezing through the duration of the cold season. Changes in snow cover extent, snow depths, and solar exposure regimes contribute to process complexity even across such fine, field-scales.

### *Exploring Fine-scale Surface Temperature Variability with UAVs over Snow*

Compared to imagery in **Figures 39 and 40**, UAV imagery in **Figure 41** show considerably less variation. Even so, there remain several fascinating ongoing processes. The relatively few land cover classes allow for further quantitative analysis into terrain parameters which control LST variability. In this case, multi-temporal surface temperature changes over snow are also characterized over a larger area (~6 acres) by combining many individual calibrated thermal images.

Shading, differences in cover between the roadway, snow, and forest dominate the observed variability. Colder relative temperatures due to increased snow density over a packed trail also illustrate the effect snow density can have on surface temperature. Even as **Figure 41** presents observations of snow during a period when air temperatures are observed to be above freezing (**Figure 42A**), the amount of energy required to drive large scale phase changes in a deep and non iso-thermal snowpack is large. As a result, no notable melt is observed over the study plot, though near tree crowns and the road surface snow surface temperatures are shown to approach 0 °C during the solar maximum (**Figure 41C, Figure 42B**).

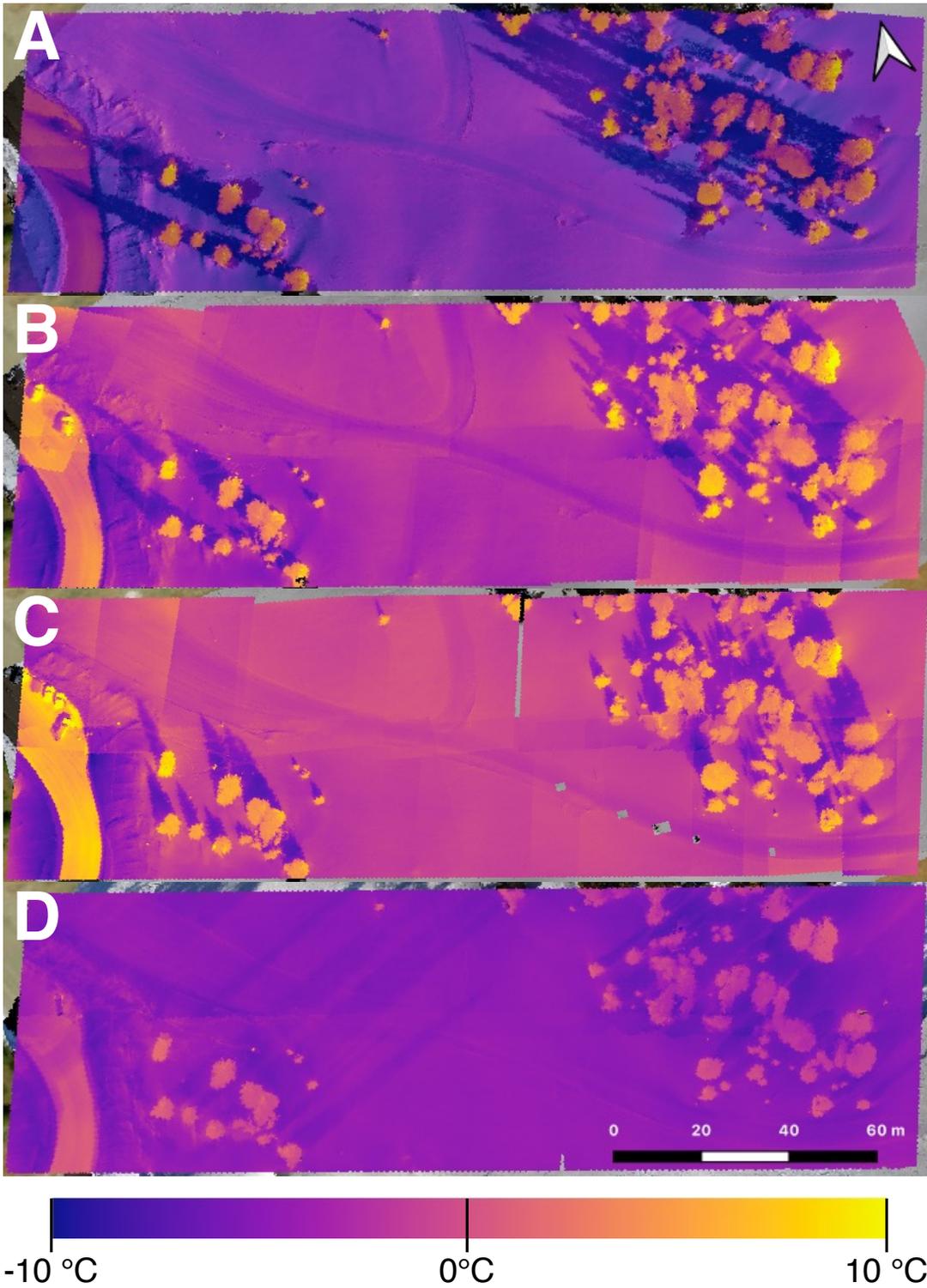


Figure 41 Surface temperature imagery from UAV collections from February 2, 2020: (A) 9:33 a.m., (B) 10:42 a.m., (C) 12:03 p.m., and (D) 4:23 p.m. All times are in Mountain Standard Time (MST, UTC - 7)

Vegetation and snow classes are partitioned to develop class specific temperature distributions throughout the study period (**Figure 43C**). While changing at similar rates throughout the day and into the afternoon, each class has a distinct thermal signature. Expressly, snow surface temperature ranged from 262 °K (-11 °C) to around 273 °K (0 °C), and 270 °K (-3 °C) to approximately 283 °K (10 °C) for vegetation. These values are shown to fall within the range of satellite LST observations from both GOES and MODIS for the study site. The tendency of a satellite observation in the region to be relatively colder or warmer is largely determined by the dominant local cover class of either snow (colder) or forested areas (warmer). These comparisons provide perspective on how satellite imagery remains an aggregation of many complex emissions from the surface, tending towards a blended average of all sub-pixel components. Future efforts to utilize modeling such as GPR, other ML approaches, and physically based modeling offers potential for de-aggregating or downscaling these observed signals to derive improved estimates of temperature and energy exchanges at the sub-pixel scale.

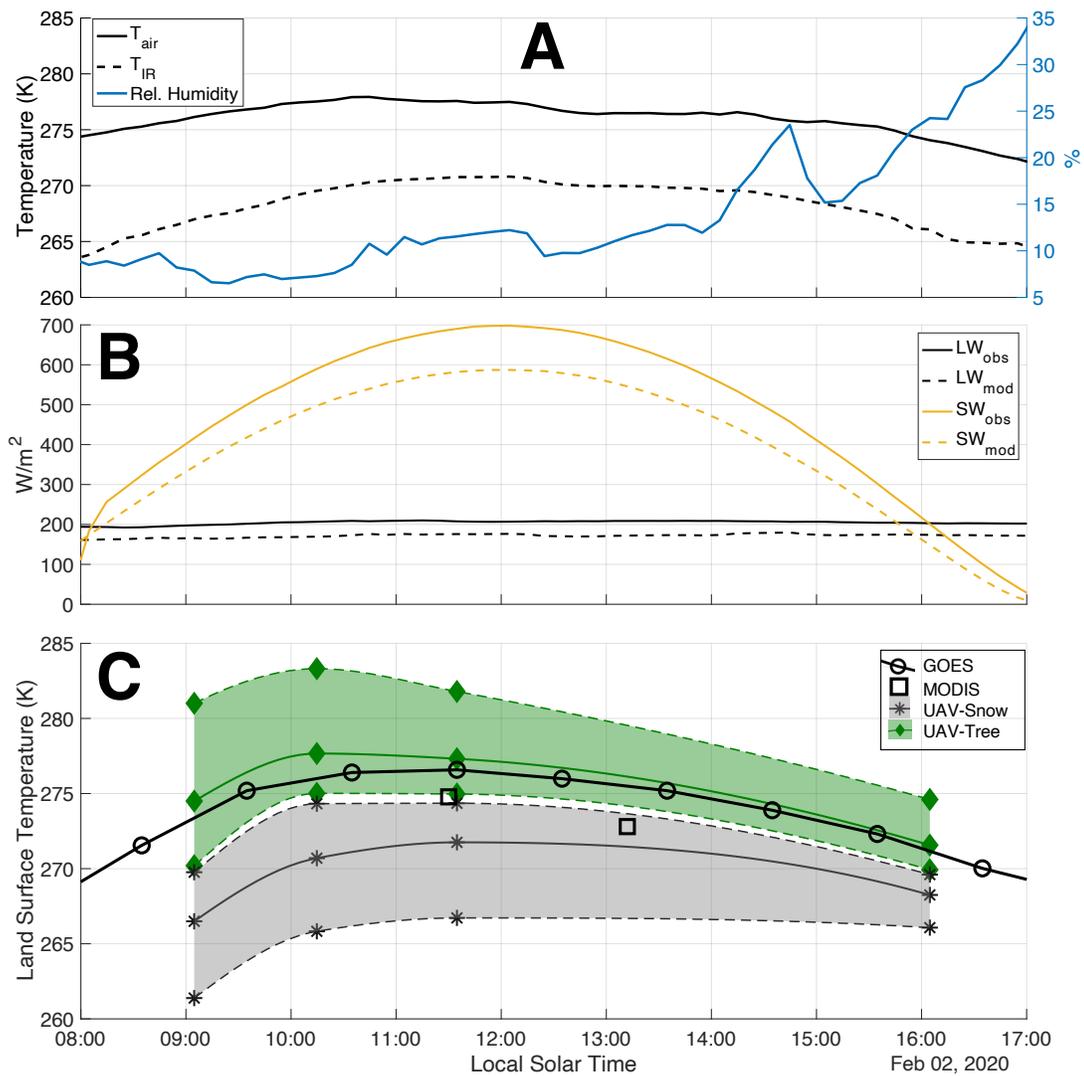


Figure 42 Timeseries of near site observations and energy balance: (A) In-situ air temperature ( $T_{air}$ ), surface skin (snow) temperature ( $T_{IR}$ ), and humidity; (B) Longwave and shortwave radiation (observed and modeled); (C) Satellite LST observations and UAV partitioned snow and vegetation specific surface temperatures. Means (solid lines) and 1st – 99th percentile (dashed)

### *Field Work Learning Outcomes and Feature Importance*

Using collected thermal imagery detailed in **Figures 41 and 42**, the characteristics driving large temperature ranges within each image are explored using GPR models. The relative difference in temperatures to the scene average is used as the

target variable, predicted through a combination of features such as land cover type (snow or vegetation), terrain, and derived predictors relevant to solar exposure such as hill shade, topographic prominence, and shading. Relative feature importance is estimated (**Figure 43**) to provide insight into which characteristics impact surface temperatures and the associated radiation balance.

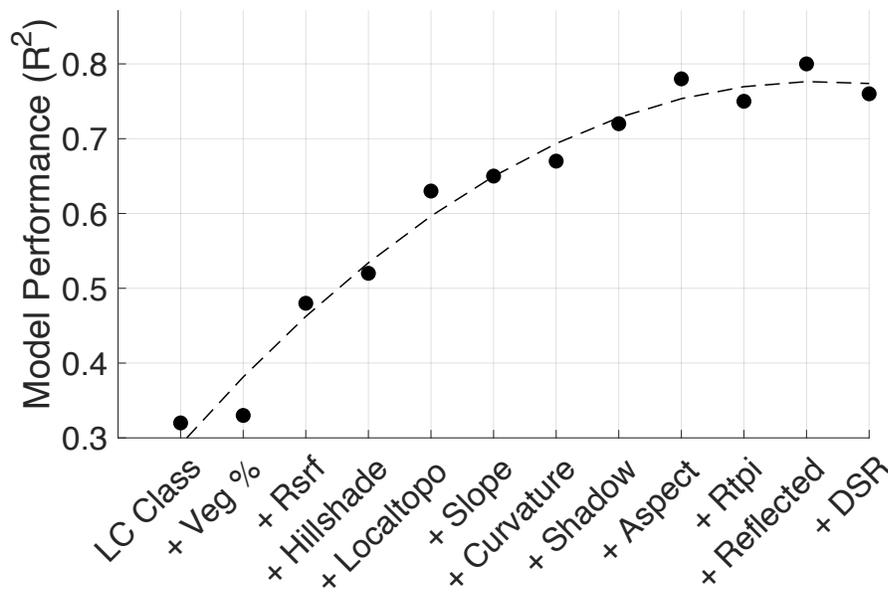


Figure 43 R<sup>2</sup> improvement by adding additional predictors to land surface temperature model

Using land cover alone to estimate temperature relationships provides a reasonable model ( $R^2 > 0.3$ ). In this case, binary cover classes of vegetation and snow are found to be the most important of all model predictors. Followed by the exposure of slopes (i.e., hillshade, shading, slope, and aspect), and terrain influences like surface roughness ( $R_{srf}$ ) and topographic prominence (Localtopo). Differences among such variables are known to create microclimates which can have considerable implications

for snow melt and freeze/thaw processes at fine scales (*Charrier et al. 2015*).

Remarkably, the final trained GPR model showed exceptional predictive skill at de-aggregating UAV imagery, even on holdout validation sets and between all four image collection periods ( $R^2 = 0.84$ ). This approach effectively reproduced the temperature spatial distribution at the study site, suggesting the predictive value of the presented variables at de-aggregating scene averaged surface temperatures.

Extensive differences are observed in the thermal profiles of various different parts of the same landscape. Even in periods of stable temperatures, and relatively simple land cover (e.g., vegetation and snow), variations in the energy exchange between various surfaces and solar loading, can lead to a wide range of temperatures. Drivers of such differences can be challenging to determine through traditional regression and statistical approaches. Machine learning enables the learning of multi-dimensional relationships between air temperature, surface topography, solar loading, and land cover providing insights to modeling surface temperatures and associate melt freeze cycles at ever finer resolutions. Similar UAV investigations are expected to provide valuable observational data over additional land cover and climate regimes to aid in the development of new models relying on basic land surface parameters to predict surface and near surface soil temperatures at the field-scale. Such models may be useful for decision makers for prediction of freeze events for agricultural, hydrology, and numerous research applications. Applications merging coarse resolution satellite data with such models using a data assimilation framework also present a viable opportunity to downscale satellite temperature observations to any desired resolution.

### **3.3 Chapter Outcomes**

Introducing comparisons of known surface conditions (i.e., detailed ground-based observations) to remote sensing datasets (i.e., P-MW, thermal infrared, land cover) provides a valuable step towards deriving improved representations of surface freezing and thawing processes. These efforts have identified new variables which may be useful for freeze/thaw classification and also presents a new physically based understanding of variability in surface freeze/thaw conditions at sub-pixel and sub-watershed scales.

Initially, through these comparisons we identify thresholds at which FT classifications begin, move to improve frozen classification definitions, and specifically examine seasonal changes and surface characteristics governing FT observations over a diverse range of climate regions. Specific conclusions and suggestions stemming from this work include:

- *Products perform well in far northern regions with clear freeze/thaw seasonality*
- *Elsewhere, frozen classifications are shown to be uncertain and representative of a wide range of temperature states which underlines the weaknesses of seasonal threshold algorithms*
- *Vegetation type, snow cover, and depth are identified as valuable contributors to varied FT signatures. As is varied topography, surface water, and climate*
- *There is a considerable need for classifications beyond binary states, such as fractional products or those with several freeze classes to improve sub-grid and*

*regional representations. Especially in regions with mixed topography and abundant surface water*

- *Fusion of multiple datasets such as surface temperature, land surface variables, ground observations, and multi-band microwave (L- and Ka-band) observations can provide different information towards achieving a classifier that captures specific surface components*
- *Regionally varied classification algorithms should be implemented to ensure globally consistent definitions of surface freeze states*
- *Surface skin temperature-based rules could be useful for reducing false freeze classifications and deriving improved resolution classifications*

Primarily, these results will inform new FT classification algorithms, in which well characterized FT definitions are the primary objective. For example, identifying the difference between frozen soil and surface frost using a combination of data inputs in the microwave bands (L- and Ka-band). The use of fractional and freeze type classifiers is expected to prove valuable in achieving this goal by better characterizing frozen states and thus the region-specific implications for water, carbon, and energy cycling.

Other efforts focused on field-scale temperature characteristics have identified several factors driving differences in surface temperature and potential factors contributing to variability in FT cycling across relatively short distances. While some of these results support the above findings, detailed examinations of observational sites and UAV temperature surveys have supported additional conclusions:

- *Existing freeze/thaw products capture different freeze signatures compared to an observing site at Grand Mesa. UAV observations illustrate how temperature variability at fine scales and the use of point observations may contribute to these uncertainties*
- *Comparisons of UAV imagery to satellite observations support the assertion that coarse-scale observations represent the average emissions of many different fine-scale processes*
- *Land cover type, vegetation, topographic prominence, surface roughness, aspect, and shading provide strong controls on surface temperature variability and offer potential for developing hyper-resolution predictive models*
- *While land surface temperature does not explicitly observe soil freeze/thaw states, the captured patterns provide insight on how soil temperatures may be evolving beneath the surface skin layer*

Advanced computational approaches (i.e., machine learning) presents the capability to merge multiple data streams including in-situ and satellite information to tune and validate such algorithms. These approaches are being used increasingly to efficiently leverage all available information to optimize new classifiers.

#### **4 GMU-FREEZE/THAW: REMOTE SENSING DATA FUSION FOR IMPROVED FREEZE/THAW ESTIMATES**

Preceding analyses ranging from local to global scales provide a comprehensive assessment of the current observational capabilities for detecting freeze/thaw transitions. These works also underline the potential of new approaches for improving such classifications by identifying limitations and new relevant variables that can be leveraged for their improvement. Within this chapter, the culmination of these lessons is applied to the development of a new freeze/thaw classification approach.

Soil temperature observations remain the best method of which to accurately capture surface freezes since air temperatures have been found to vary considerably from actual surface states. Using soil temperatures specifically also provides a clear physical basis for the estimation of freezing and thawing transitions, which can allow for direct comparisons of surface freezes between different regions by moving beyond the somewhat arbitrary nature of current landscape classifiers. It has become clear, that the differences between regions in terms of climate, land cover, and topographic variability introduce significant variations in observed remote sensing signals across bands (i.e., L-, Ka-band, thermal infrared) and thus completely different relative meanings of what defines a 'frozen' or 'thawed' state. This work illustrates the value of merging multi-dimensional and markedly different observations for optimizing freeze/thaw classifications. This is achieved by introducing new regional variables into the classification scheme, training algorithms based upon globally diverse networks of soil

observations, and by integrating machine learning approaches such as decision trees and random forests into global freeze/thaw classification.

#### **4.1 Motivations and Approaches for the Enhancement of Freeze/Thaw Observation**

Global and continental scale assessments of current freeze/thaw detection methods have identified several weaknesses of existing classifiers (*Xu et al. 2018; Lyu et al. 2018; Johnston et al. 2020; Johnston et al. 2021*). Current methodologies have exhibited reduced performance due to factors such low moisture content (i.e., arid regions), transitional areas (i.e., temperate regions), and over complex topography (*Kim et al. 2011; Derksen et al. 2017*). Similarly, existing approaches to FT detection have remained highly variable in their outputs and the associated physical meaning of a frozen classification within these areas. This highlights a need for improved classification approaches. Tailoring classification algorithms to explicitly target freeze detection for the near surface soil layer helps to improve model characterization and indicate periods in which runoff, plant productivity, and land atmosphere exchanges are most constrained.

In-situ observations provide the best measure of realistic surface soil states, by avoiding large uncertainties in global reanalysis data or modeled temperatures to parameterize a classification model (*Kim et al. 2017; Dunbar et al. 2014-2020*). However, challenges at acquiring a regionally diverse dataset of soil temperatures, at least compared to the breadth of air temperature observations, can limit its applications. This is because regions with sparse data tend to have increased uncertainty from lack of an observational record and thus limited validation. Also, physical differences in soil properties such as increased salt content can cause depressed freezing points in which soil

temperatures can fall below 0 C and still remain thawed (*Suzuki 2004*). Still, with a sufficiently large and spatially diverse dataset, capturing the wide range of conditions can be achieved. The use of in-situ meteorological observations to train machine learning models does not represent a new paradigm (*Du et al. 2019; Chen et al. 2020*). The approach has even been applied on smaller sub-continental scales to FT classification with success (*Zhong et al. 2022*). Even still, its benefits have yet to be applied exhaustively towards the modeling of soil FT states on a global scale. Notably, approaches using spatially distributed variables such as microwave brightness temperatures for FT prediction tend to also be validated using in-situ observations. Meaning, that observational uncertainties are not well constrained over data sparse regions anyway.

To compliment information provided by in-situ observations and in order to derive spatially relevant information at a global scale, remote sensing provides crucial inputs. Various bands present different observational capabilities. Namely, microwave-band observations can provide information on water conditions in various layers of the surface and have direct sensitivity to FT transitions (*Royer et al. 2010; Mavrovic et al. 2021*). Still, observational resolutions on the order of 5 to 10's of kilometers, relative to shorter wavelength capabilities (< 100 m), can provide a limitation in microwave bands. Observed signatures can also be complex to de-aggregate and highly dependent on temperature and surface moisture conditions (*Zheng et al. 2020a*). Land surface temperatures (LST) have also shown relevance to the underlying freeze condition of the surface by providing a boundary condition and control on freeze depth (*Zheng et al.*

2020b). Relevant remote sensing LST products also present improved resolution (100 m – 5km) and have a clear relationship to surface physical temperatures. These capabilities have enabled resolution improvements for near-surface FT state estimates (*Zhao et al. 2017*). Still, these observations are not directly sensitive to the surface FT state and can be regularly impeded by cloud cover. Snow cover presence and related snow indices, while usually binary, are also known to have relevance to surface FT conditions, providing additional information to better characterize surface states (*Iwata et al. 2011*; **Chapter 3**).

In another vein, other surface features and relevant regionally variables that persist over extended periods of time (herein referred to as ‘static’ variables) also display value for achieving robust FT classification methods. Features such as topographic complexity can impact the net radiation balance (*Yan 2020*), the presence of vegetative cover can impair shorter wavelength bands, and climate variations can have direct implications moisture and surface temperature conditions (*Ni et al. 2019*). Such variables provide additional potential for regional varied classification, similar to existing approaches (*Jia and Richards 1999*; *dos Santos et al. 2012*), but instead by capturing regional drivers of soil FT states in place of tuned pixel-by-pixel thresholds (*Kimball et al. 2009*; *Derksen et al. 2017*).

Data fusion through machine learning allows for the incorporation of many pieces of information and can selectively draw from various strengths or weaknesses of input data (*Castanedo 2013*). These data driven models have the capability to learn complex patterns and relationships within N-dimensional datasets, making them especially

attractive for research applications and in optimizing systems in our everyday lives. Compared to some classification tasks, which can have hundreds if not thousands of potential outputs to distinguish between (*Pappu and Pardalos 2014*), the task of modeling FT states is a relatively simple one. Still, single observational bands remain incapable of observing surface conditions across all components of the surface. The determination of physical states within a complex and everchanging environment also makes the application of such methods non-trivial.

While binary FT representations are valuable, they greatly simplify true surface states, which can vary on extremely fine scales (*Johnston et al. 2021, Chapter 3*). Machine learning approaches enable probabilistic predictions that can directly give an indication of model confidence (*Li 2013*). This can be used to indicate uncertainty in transitional or other climatologically complex regions and to present intermediate or partial freeze states. The importance of product characterization cannot be understated. Differences between a 1-meter depth soil freeze, surface snow cover with thawed underlying soil, or diurnally cycling surface frost can have considerable and varied implications for global energy balances, nutrient exchange, biodiversity, and carbon uptake, especially when considered at continental scales (**see Chapter 1**).

The underlying physics and detection methods for the remote sensing of freeze/thaw processes have been reviewed in previous chapters. From these efforts, we identify opportunities to fill observational gaps by drawing on the benefits of data fusion techniques (*Johnston et al. 2021*). Specifically, this work details the combination of an exhaustive global network of soil temperature observations, static variables (i.e.,

topography, climate, land cover), and multiple daily satellite observations both in passive microwave bands (SMAP, SSMIS) as well as thermal infrared bands (MODIS, land surface temperature) for deriving an enhanced classifier. The details of both selecting and tuning an appropriate machine learning model to achieve this are also included herein.

Project specific objectives for FT enhancement include:

- 1) *Development of a freeze/thaw classifier with explicit physical meaning (i.e., soil states)*
- 2) *Inclusion of associated fractional and probabilistic representations of the freeze likelihood*
- 3) *Applying (1) and (2) for global-scale classification using regionally varied modeling, with associated model validation*

#### **4.2 Deriving an Improved Classifier**

The successful implementation of any machine learning-based classification problem relies on an extensive training data set. This can be defined as a large and diverse collection of variables relevant to the classification problem of interest. This ensures model robustness by capturing a large set of possible conditions. In the case of supervised classification problems, it also relies on accurate labels to be associated with such variables. In this case, worldwide soil temperature observations serve as the ground truth to provide model labels by thresholding temperatures to define binary soil FT states. The selection of various predictors such as remote sensing observations or other static land surface information provide observational inputs enabling FT classification when soil states are unknown. This section details the development of a large soil temperature-

based training set, the selection of model predictors, as well as the use of random forest models for FT classification.

#### **4.2.1 Data Inputs**

##### ***In-situ Observations***

A large quality-controlled ground truth dataset provides the backbone for supervised learning models and other data driven approaches. To achieve this, large observing networks containing observations of soil temperatures are identified globally. Namely, these consist of large, distributed networks within several regions of the United States and North America:

- 1) The Snow Telemetry Network (SNOTEL, *USDA 2020b*) – 442 sites
- 2) The Soil Climate Analysis Network (SCAN, *USDA 2020a*) – 201 sites
- 3) The United States Climate Reference Network (USCRN, *Bell et al. 2013*) – 115 sites
- 4) The Real-Time In-situ Soil Monitoring for Agriculture Network (RISMA, *Pacheco et al. 2019*) – 7 sites
- 5) The SnowEx Grand Mesa In-situ Network (*Houser and Johnston 2021*) – 5 sites

This network provides impressive coverage (770 long-term sites) over much of CONUS, Alaska, and southern Canada with particular site density in hydrologically relevant mountainous headwater basins. These records extend back decades in many cases, and all

capture the focus study period which uses observations spanning from SMAP launch (April 2015) through 2019, and in some cases extending into 2021.

While these provide a large training set over North America, these networks provide no information outside of the continent. Thus, the global SoilTemp (*Lembrechts et al. 2020*) database, developed specifically for providing soil and near surface temperatures, is also used. This network includes over 3,000 unique observational datasets provided by the research community from all seven continents and major climate regions (<https://soiltemp.weebly.com/>). Additional observations (57 sites) from in-situ networks are drawn from the data sparse High Mountain Asia (HMA) region (*Yang 2021*). Considered the Earth's third pole, this region remains particularly important for water availability in China and the surrounding regions. Improving FT classification here has implications for improving melt detection and runoff timing estimates.

All datasets are carefully cleaned removing erroneous values. Sites with poor performance (i.e., frequent errant values) or limited data availability (i.e., large periods of missing data) are also not included in the training set. To be valid for use in the training set, near surface soil temperature observations between 1 cm and 10 cm depths are required. These define soil freeze/thaw state through classifying thawed soil when soil temperatures are below 0 °C, and thawed when  $\geq 0$  °C. Where available, air temperature is also used to derive a more complex FT classifier, considering the potential of additional transitional states, however, these multi-class models are undergoing development.

The combination of all sites achieves an exceptionally large and dispersed global network of sites (>3,500). To our knowledge, this is the largest known aggregation of soil temperature information available at global scale. As such, it provides the best opportunity for training and validating soil FT state classifiers.

### ***Global-scale Static Variables***

Several inputs used to characterize climate or land surface variability are used throughout this project. Those of which that remain relatively constant over annual time periods are considered to be static variables. These static variables are included below as well as being detailed in **Appendix A & B**:

- 1) Köppen-Geiger climate classifications for 1980-2016 (*Beck et al. 2018*).

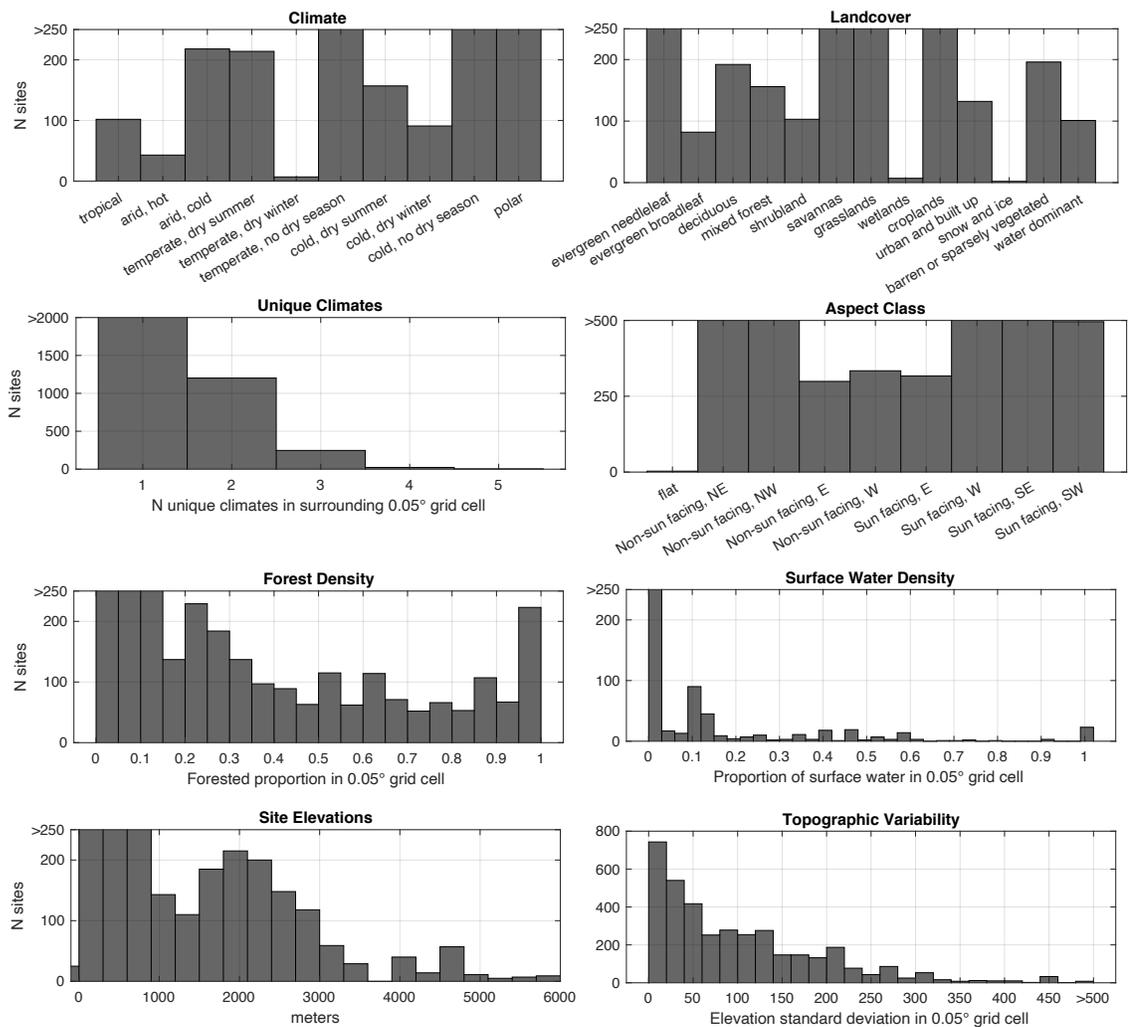
These define 30 major climate regions on an approximately 1 km resolution global grid (aggregated to 10 classes). These observations are also used to derive a metric capturing the number of unique classes within each 0.05° grid cell. Providing an estimate of potentially challenging classification regions, which frequently occur in areas with large elevation gradients.

- 2) Global land cover from MODIS using a 17-class scheme (MCD12C1, *Friedl and Sulla-Menashe 2015*). This is aggregated to 13 by combining similar classes. Estimates of forest density and surface water coverage proportions are also derived from MODIS land cover products (0.05°).

- 3) Elevation data from the ETOPO1 Global Relief Model (*Amante et al. 2009*). These 1/60° (~2km) grid of elevations are used to derive elevation standard

deviations, a topographic prominence index (TPI), and aspect, using a scheme similar to (*Gómez-Plaza et al. 2001*) at 0.05° resolution.

The distribution of in-situ sites across these various classes is detailed in **Figure 44**. All climate regions besides those in the comparatively small temperate, dry winter climate class are represented by over 30 observing sites. The distribution of sites across varied landcovers is also sufficient, outside of wetlands or areas with persistent snow and ice (i.e., Antarctica and Greenland). These areas are thus masked out in the soil FT classifier. Variability in the number of climates surrounding each site is shown to range from 1 to 5, with over 80% of the 0.05° grids containing each site being characterized by only one or two climate regions. Finally, topographic profiles suggest the majority of sites fall in relatively flat areas, but also illustrate a diverse range of topographic conditions with hundreds of sites having highly varied surrounding terrain (elevation standard deviations >200 meters).



**Figure 44** Distribution of global network of in-situ soil temperature sites ( $N = 3757$  sites) across various static variables. Global maps of these variables included in Appendix B

### *Remote Sensing Observations*

Observations from various remote sensing platforms provide the backbone of FT detection algorithms. Here, enhanced passive microwave observations of brightness temperatures ( $T_b$ ) from the Soil Moisture Active Passive (SMAP, *Brodzik et al. 2021*) and the Special Sensor Microwave Imager/Sounder (SSMIS, *Brodzik et al. 2016*) provide multi-band, twice daily retrievals with near global coverage. Both observing platforms

have similar equatorial overpass timings, allowing for direct intercomparison between these observations and improved model applicability by reducing uncertainties imposed due to varied retrieval times. Observational bands include:

- 1) SMAP 1.41 GHz Vertical (V) and Horizontal (H) polarizations [L-band]
- 2) SSMIS 19 GHz V/H-polarizations, 22 V-polarization [K-band], and 37 GHz H-polarization [Ka-band]

All microwave bands between L-band and Ka-bands used herein respond to variability in near surface soil temperature, moisture, and water phase changes. Recent efforts combining many overlapping radiometer observations from these platforms have allowed for higher resolution  $T_b$  estimates approaching 3 km. The radiometer version of Scatterometer Image Reconstruction (rSIR) algorithm uses complex antenna specific gain and measurement response functions to estimate enhanced resolution brightness temperature observations at these resolutions (*Brodzik et al. 2016*).

Other remote sensing inputs include land surface temperature observations from MODIS (11C1 v061, *Wan et al. 2021*) which use both Terra/Aqua platforms for 4-times daily global observation ( $0.05^\circ$ ). Finally, as snow cover is found to control FT processes in the underlying soil layers, snow cover extent from the Rutgers Global Snow Lab (GSL, *Robinson and Estilow 2021*) and MODIS Normalized Difference Snow Index (NDSI, *Hall and Riggs 2021*) observations are also included. Notably, snow cover extent from GSL products only provide coverage over the northern hemisphere with weekly temporal resolution.

## 4.2.2 Decision Trees and Random Forests for Freeze/Thaw Classification

### *Decision Trees, Random Forests, and Model Selection*

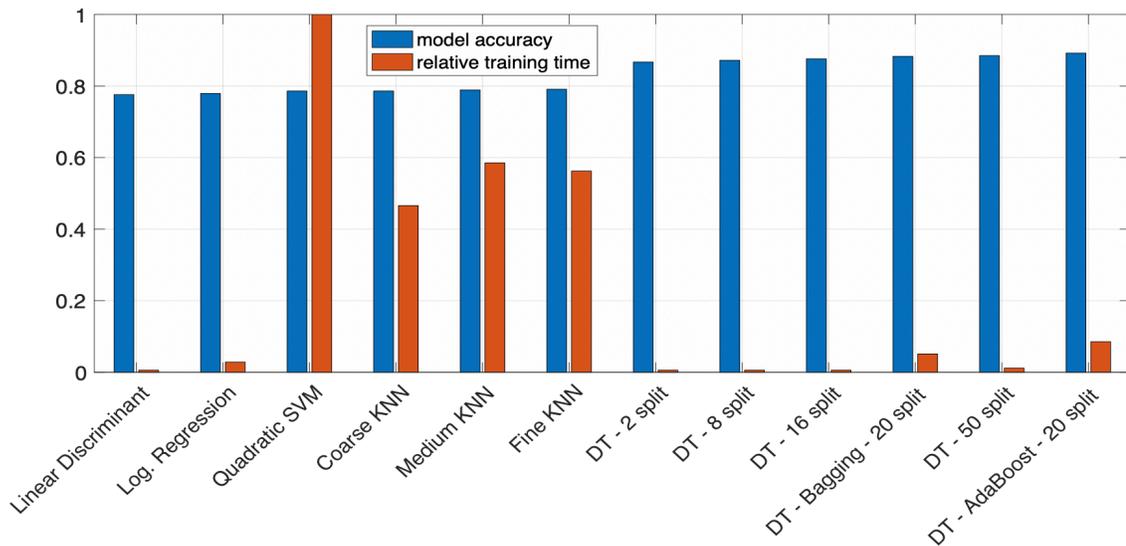
Machine learning has permeated the disciplines of data science, academic research, artificial intelligence, and beyond. Continued advancements in computational capabilities and the underlying algorithms have enabled the implementation of such data driven approaches for a multitude of predictive tasks. Essentially, machine learning provides an effective framework for automated model building using mathematical principles. These models can take many forms, and generally fall into categories of either supervised or un-supervised learning. Supervised models rely on the modeler to provide estimates of true class labels or known continuous outcomes (i.e., in regression problems). Whereas their unsupervised counterparts are used to examine the presence of potentially unknown patterns or clusters within a set of information.

Decision trees (DTs) make up a specific sub-discipline of supervised machine learning methods. They come with a multitude of benefits such as their wide applicability as a non-parametric model for both classification and regression tasks. In classification problems, given labeled training set of features, a tree is grown by iteratively partitioning the data set through split rules into leaf nodes (*Breiman et al. 1984*). These rules come in many forms, but generally include measures of node purity (*Song and Lu 2015*). Purity metrics are used to identify the locations at which the dataset can be most closely partitioned into unique classes using a single predictor. This process is performed iteratively until a stop criterion is reached, which can be in the form of maximum number of splits, model performance thresholds, or the point when minimum entropy is reached.

Particularly in binary classification problems, these trees remain relatively non-complex while being highly effective (*Bahel et al. 2020*). Pruning can be employed to reduce the total number of leaf nodes and simplify the final model. In FT classification, *Jin et al., (2009)* previously used decision tree models to estimate soil FT states using microwave observations with success (87% accuracy) in China.

Random forests (RFs) and bagging approaches leverage this predictive capacity by creating ensembles of unique decision trees. In bagging, random subsets of a training set are sampled, and models grown through the standard method by using splits across all predictors ( $M$ ). RF approaches follow a similar methodology; however, they induce additional randomness by only allowing splits across a random subset of predictors ( $m$ ) in which  $m < M$ . These approaches have been shown to result in more robust models for unseen data when compared to a single DT (*Breiman 2001*).

Random forest classifiers in particular have grown in popularity within the remote sensing community due to their effectiveness across a multitude of classification tasks and ability to learn complex multi-dimensional relationships (*Belgiu and Dragut 2016*). For example, RFs have been used for remote sensing-based land cover classification (*Pal 2003*) and in spatially downscaling spaceborne estimates of precipitation (*Mei et al. 2020*). Similar approaches have even been recently applied in classifying lake phenology and its relationship to FT cycling (*Han et al. 2020*). Such approaches enable seamless combination of multiple data sets by allowing prediction even in the case of missing data while providing predictions at the highest resolution of any single input variable.



**Figure 45 Model accuracy and relative training time of various supervised machine learning methods**

In assessing the applicability of tree-based models for FT classification, test data including several static and  $T_b$  observations as predictors over North America are used. **Figure 45** shows the slower relative training time and reduced performance of various classifiers for FT detection tasks compared to DTs. Relative to linear discriminant, logistic regression, K-nearest neighbors (KNN), and support vector machine (SVM) approaches, DTs have proven the most effective with prediction accuracy approaching 90%. The relative simplicity of DTs also ensures rapid training, applicability, and in many cases enhanced model interpretability. Alone, single decision trees have impressive predictive capacity, however, the integration of random forests specifically allows for probabilistic model estimates and improves model robustness. These models can allow for improved uncertainty estimates and provide valuable information beyond that of binary classifiers.

### ***Model Parameters and Tuning***

Several parameters selection and tuning decisions must be made when optimizing decision tree and random forest models. These provide valuable limits on the model framework to avoid overfitting and increase generalization capacity. Metrics such as the number of trees within a random forest, maximum splits and minimum node size, feature weighting, and sampling techniques, when properly selected, can improve model performance and generalization substantially. Other additional modeling considerations during the training phase of this project include normalization, tree pruning, split criterion, and number of predictors to randomly sample in the case of random forests. However, these have been applied consistently across various regional models due to minimal effect on model performance. Further sensitivity analyses are included within **Section 4.3.**

#### *Number of Learners/Trees*

This metric largely defines the complexity of the model. Single decision trees provide the baseline of tree-based learners (1 learner) but are inflexible. While these models can be particularly effective, the introduction of many trees is applied to improve model generalization through creating an ensemble of random learners. More trees also allow for improved estimates of class membership probabilities associated with frozen or thawed soil states. This value is set at an optimized level in which additional model improvement was minimal while also avoiding overfitting. Generally, this value was set to 100 learners (200 in the case of very few freezes) as the gain in predictive capacity after this point is observed to be near zero.

### *Maximum Allowable Splits and Minimum Leaf Size*

Minimum leaf size and maximum allowable splits provide very similar influence on a given RF model. Each leaf represents a terminal node in which model classification occurs and presents considerable influence on model complexity and thus overfitting potential. For example, a model derived from large datasets such as within this project (~50 million feature sets), allowing splits that contain a single data point is unwise. This approach can lead to very high model performance scores but provides no predictive capacity in cases of unseen data (overfitting). In this project, the maximum number of splits is set as a function of the size of a given training set, ranging from 1% of training set size to as low as 1/1000<sup>th</sup> of a percent in the case of classes with very few freeze classifications. This value is tuned in the final algorithms, to optimize the performance and robustness trade-off. In other words, it is kept as large as possible without significantly reducing model performance. In model training the maximum number of possible splits is set at 200. In most cases, the leaf size parameter provides the primary limit on model complexity and the actual number of splits within each learner remains below 50. In deriving clearly traceable results such as when using a single decision tree, fewer splits are ideal.

### *Weighting of Input Features*

The weight given to a particular correctly classified feature provides a method to effectively tune model performance in unbalanced datasets (e.g., many more thawed than frozen observations). In this project, we optimize weighting by climate type to ensure that there are not excessive frozen or thawed classifications and frozen events are consistently

captured. Weights are varied depending on the number of frozen classifications relative to thawed classifications, to ensure one dominant class doesn't become the only one predicted. For example, in regions with <1% frozen classifications, the model can achieve 99% accuracy by predicting only thawed states. However, this approach will omit all cases in which freezes did occur. Notably, these weights do not play a role in the trained model, only in the optimization stage.

### *In Bag Fraction*

In bag fraction defines the size of the sub-sample used to select data for each tree in the random forest. When doing a single DT, the data sampled includes all of the training data. However, in random forests, this value controls how the data is subset. The sample size in a random forest act as a control of randomness. When this value is set higher, there is less randomness, and more of the full set is represented in each tree. Conversely, a value that is too small may result in a subset that inadequately represents the data. When using a smaller subset, the trees tend to vary more widely which can lead to a more generalizable model but may result in reduced model performance due to training the model with small sub-samples of the full feature space. In practice, this value is selected using bootstrapping in which the sample size is equivalent to the training set size (In Bag Fraction = 1), though sampling is applied with replacement so certain features are not sampled while others are sampled twice. This leads to around 60-70% of the training dataset being used in each tree.

For this application and in the case of large and diverse data inputs, this sampling proportion can be reduced to as low as 10% (0.1) of the training set. Using lower values

results in considerable reductions in training time and reduced overfitting potential. For most large feature sets in cold climates with abundant frozen classifications, we set this value to 20% (0.2). This value was increased (0.4 – 0.6) for smaller datasets with few frozen classifications in order to ensure sampling of feature sets classified as frozen. Even so, the impact to model performance was relatively small and no significant drop off is observed in performance until the In Bag Fraction was reduced to below 1% (0.01) of the training set.

#### *Split Criteria and Split Selection Method*

To quantify value of splits at any tree node, the Gini-Simpson Diversity Index (GDI) is employed (*Simpson 1949; Sen 2005*). This commonly used approach provides a measure of node purity, in which a fully pure node represents all features are labeled with the same class (i.e., frozen or thawed). GDI measures the probability of a particular variable being wrongly classified if randomly chosen. In a fully pure node, this probability is equal to zero, as is GDI. In cases in which elements are randomly distributed across various classes, GDI approaches unity (1). This provides a satisfactory metric of which to estimate optimal splits. Using computed GDI values, optimal splits are determined by minimizing GDI over all possible splits (*Breiman et al. 1984*). This approach was tested and shown to provide comparable accuracy to other split criteria metrics including maximum entropy and the twoing rule.

#### *Number of Predictors to Sample*

This hyperparameter is specific to random forest models as classical DTs use all input features to determine optimal splits. This metric determines the number of

predictors to randomly select and define the next split within each node of the tree. A standard rule of thumb involves estimating this value by rounding the square root of total predictors. For example, in the case of 20 ( $N_{all}$ ) predictive features,  $N_{sample} = round(\sqrt{N_{all}})$ , or 4 features. This parameter was varied in training to find minimal model improvement with more than 4 features, as higher  $N_{sample}$  values also result in less ‘random’ or generalized model.

### *Tree Pruning*

Tree pruning is a valuable approach to reduce the number of branches and nodes in the tree to improve its interpretability and avoiding overfitting. This is done in a way to reduce model complexity without forfeiting much accuracy. In the case of decision trees, this method is applied to classification branches contributing minimal model improvement until the desired number of splits is met. In the case of random forests and ensemble learners, pruning is generally thought to be unnecessary and does not affect classification estimates.

### *Normalization*

Normalization remains an important consideration in the training of ML models. Normalization prevents any single predictor from dominating model performance, simply by having larger values. To be effectively applied, the valid range of values need to be well understood. In the case of decision tree classification problems, the problem of normalization is less critical, as splits are determined dependent on partitioning of class labels, not input predictor values. As such MW products are not normalized to retain

actual  $T_b$  observations and to retain interpretability. For detailed description of each predictor used in these models, see **Appendix A**.

### 4.2.3 Feature Creation and Model Development

#### *Feature Engineering and Training Set Creation*

In addition to satellite  $T_b$  observations, two calculated freeze/thaw relevant predictors are incorporated into the model. The first, Normalized Polarization Ratio (NPR, *Dunbar et al. 2018*), has been widely used for FT detection in SMAP-FT products with good performance in northern regions. NPR approaches have shown the potential to detect dielectric changes in water properties which, due to the strong polar nature of water molecules, can result in different behaviors across bands. While this approach is generally applied to L-band observations (1.41 GHz), we also derive this metric using K-band (19 GHz) cross polar observations from SSMIS using the approach detailed in **Equation 3 (Chapter 1)**.

The second metric, spectral gradient (SG) is computed as the change in  $T_b$  over the change in frequency ( $^{\circ}\text{K}/\text{GHz}$ ). *Zuerndorfer et al. (1990)* were the first to use this metric to discriminate FT boundaries. Negative spectral gradients between lower frequency to higher frequency bands are shown to correlate well to frozen classes as volume scatter darkening in the lower frequencies dominates the response. More specifically, *Judge et al. (1997)* demonstrated the sensitivity of the Ka-band to surface dielectric properties and FT state using the 19 to 37 GHz gradient. In this work, we utilize the same approach with SSMIS observations, as well as testing a new gradient using

SMAP observations between the 1.41 GHz and 37 GHz bands. In theory, this should exaggerate the band darkening response observed between 19 and 37 GHz.

Any model derived through supervised learning techniques is only as accurate as the data it is trained on. For this task, a large and quality-controlled set of labeled training data is required. In this case, labels are defined using a binary metric representing frozen ( $T_{\text{soil}} < 0 \text{ } ^\circ\text{C}$ ) or thawed ( $T_{\text{soil}} \geq 0 \text{ } ^\circ\text{C}$ ) soil states. Following the identification and selection of the aforementioned FT classification predictors, these metrics must then be matched to in-situ observations and soil FT classifications both in space and time. In the case of remote sensing observations, classifications were matched using the nearest-neighbor approach, in which observational data from the grid cell center nearest a given in-situ site is appended to the feature set. However, variability in the temporal component may also induce errors as different satellite products may be representative of different times. To achieve this, product observational times included in product metadata are recorded for each specific site. These timeseries are then matched to the corresponding in-situ timeseries, with a 3-to-6-hour observational tolerance, depending on the observed rate of change of a given observation (i.e., LST was set to 3 hours, as it generally changes more rapidly than MW  $T_b$  observations). This allows for increased overlap between data originating from different observational platforms and thus a larger set of training features. Descriptions of all features in the final training set are included in Appendix A.

### ***Feature Importance Estimates***

Exploring feature importance provides a method to understand which variables provide the most information to a classification or regression task. In this case,

understanding how this varies spatially enables improved understanding of which remote sensing and land surface variables are most predictive of frozen soil. Random forest models are trained using a random sub-set of the training set, to ensure diversity in observations used to train each individual tree. The features not selected for a given tree are considered to be the out-of-bag features (OOB). Using these OOB features, the error of the originally trained tree is computed. Then, values of each predictor variable used to define splits are randomly permuted (while others are held constant) and error is again calculated on the dataset. The difference in error between the original model minus the permuted one provides a measure of feature importance. This approach is generally referred to as predictor importance by permutation (*Breiman et al. 1984; Breiman et al. 2001; La Cava et al. 2020*).

If the permutation error is large, it indicates that a predictor was of particular importance to the trained model, and vice versa. In other words, if it is high, it means the model estimates change significantly when a given variable is randomly re-ordered. If it is lower, it means the variable is less so important and thus as its values change, the model shows minimal loss in predictive capability. These metrics are then normalized within each trained RF (or DT) model by averaging scores across all trees (or a single tree) and predictors. Then values are normalized 0 – 1, in which the most important predictors achieve higher scores to allow for intercomparisons between regional models.

### ***Model Performance Metrics***

The effectiveness of a given model and the importance of various predictors is captured through the use several metrics. Namely, these include classification accuracy

(**Equation 7**) as well as precision (**Equation 14**) and recall (**Equation 15**). These metrics enable thorough assessments classifier performance in binary classification problems. Precision details the ability of a product to avoid false positive (frozen) classifications, whereas recall measures the ability of a classifier to avoid false negatives (missed frozen classifications). In training effective models, we attempt to balance these metrics, with increased focus on improving recall scores.

**Equation 14 Precision**

$$Precision = \frac{N_{FZfZ}}{N_{FZfZ} + N_{FZtHW}}$$

**Equation 15 Recall**

$$Recall = \frac{N_{FZfZ}}{N_{FZfZ} + N_{tHWFZ}}$$

$N_{FZfZ}$  = the total number of instances where both the model and observation are frozen

$N_{FZtHW}$  = the total number of instances where the model classifies frozen but is thawed

$N_{tHWFZ}$  = the total number of instances where the model classifies thawed but is frozen

Similarly, the F1-score (**Equation 16**) provides a metric which balances both precision and recall giving a more wholistic understanding of classification accuracy (*Sasaki 2007*). It has been used regularly to characterize model performance in land surface classification problems for soil (*Abraham et al. 2020*) and anomaly detection tasks (*Kulanuwat et al. 2021*). This metric places particular importance on the value of frozen classifications. For example, with classification accuracy alone, it is very difficult to assess model skill in a scenario in which freeze classifications are few (i.e., temperate regions). In such areas, a model can have very high accuracy scores by always classifying the region as thawed, however, this approach misses instances in which freezes do occur resulting in a very poor F1-score.

**Equation 16 F1-Score**

$$F_1 score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

All metrics are computed for models using holdout validation in which 75% of a given training set is used for training and the remaining 25% is held out of the model and used for validation. This approach is used to prevent over fitting. A spatially weighted predictor selection approach is also employed, to avoid regions with very large data records from being weighted too heavily in the model. Specifically, limits are set on the number of soil observations that can be drawn from any single 1° x 1° grid cell depending on the median number values across all 1° grids containing data. Values are then randomly extracted for each region using these conditions. In the cases of very large sets of training data > 5 million features, even smaller random data subsets are used (<25% of features). Even so, model performance is shown to remain high (**Section 4.3**) when validated on the large holdout feature sets.

## **4.3 Product Performance and Results**

### **4.3.1 Freeze/Thaw Modeling at Point Locations**

#### ***Sites for Point Modeling***

The effectiveness of decision trees and random forests, as well as the various regional differences in FT signatures is explored at five (5) unique observational sites, representative of different freeze/thaw regimes. This effort presents an investigation into the capabilities of DT-based modeling approaches and their accuracy at reproducing site-specific observations of soil FT states. This work also introduces new probability metrics

to incorporate model confidence into classification estimates. **Figure 46** details the locations of focus sites (numbered 1 – 5) as well as the distribution and features of soil temperature observing sites within the complete training set.



**Figure 46** Soil temperature observational sites. Record duration indicated by marker size, and relative percentage of the record with sub-freezing temperatures indicated by color. Red markers numbered 1 to 5 indicate sites where decision tree classification models were trained

Observational sites selected herein exhibit multi-year and continuous records across different climate regions and land covers. The selected sites are detailed in **Table 15** and are ordered from most soil freezes (Alaska, ~60% of the multi-year period) to least (Texas, <1%). For each site, both DT and RF models are trained, using 75% training and 25% validation splits. In the case of DTs, models are limited to 10 splits for interpretability, then pruned accordingly to remove such splits which add minimal improvement in classification performance. For RF models, 100 unique trees were

trained using a maximum of 20 splits. Only non-static features, including MW observations, LST, snow cover, and derived metrics are included in model training. In **Figures 47-51**, classifications from the single DTs are shown along with the corresponding probabilistic estimates from the RF classifier. This provides an opportunity to compare predictor importance, model performance, and uncertainty across each unique site.

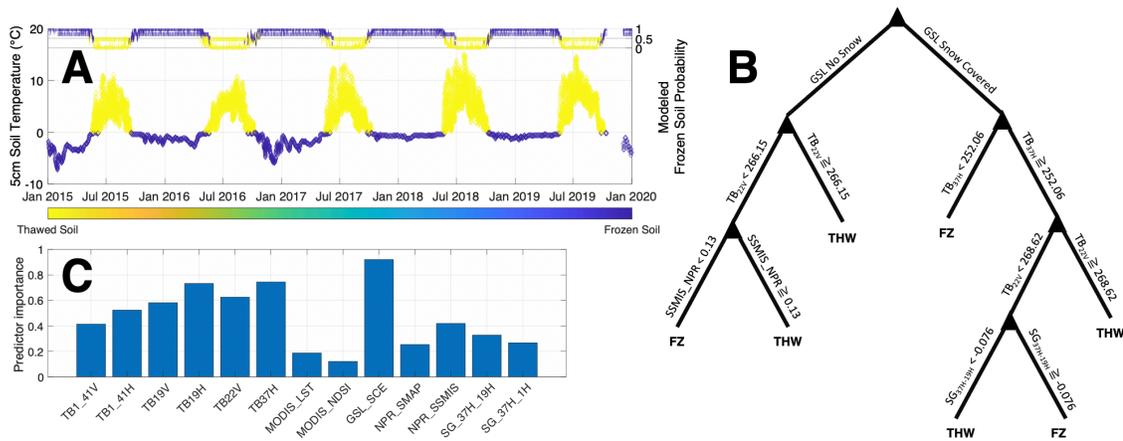
**Table 15 Focus site descriptions. Numbers correspond to sites in Figure 46**

Site (ID)	Location	Network	Climate	Frozen %	Land cover	Terrain (elevation)
Alaska, USA (1)	67.9° N, 162.3° W	SNOTEL	Cold climate, no dry season	60%	shrubland	Flat (70 m)
High Mountain Asia (2)	31.6° N, 91.7° E	HMA	Polar/tundra, dry	40%	grasslands	Flat, highlands (4550 m)
Northern Italy (3)	44.3° N, 10.4° E	SoilTemp	Cold climate, no dry season	25%	63% deciduous forest	Mountainous (1815 m)
Illinois, USA (4)	40.0° N, 88.4° W	USCRN	Cold climate, no dry season	10%	croplands, 13% forest	Flat (212 m)
Texas, USA (5)	33.4° N, 99.9° W	SCAN	Arid, cold	0%	croplands, 10% forest	Flat (441 m)

### *Decision Tree and Random Forest Model Results*

At the Alaska observational site, characteristic of a sub-polar region with minimal vegetative cover, both RF and DTs are shown to be extremely effective (>95% accuracy, **Figure 47**). The coincident timing of snow cover and frozen soil result in snow cover presence acting as a strong proxy to soil FT state. Thus, snow cover is selected as the feature with the most predictive value (**Figure 47C**). Microwave classifications in the higher frequency bands (19 GHz – 37 GHz) also show increased model value relative to SMAP observations in L-band (1.41 GHz). The DT model effectively identifies site specific thresholds across bands that optimize performance (**Figure 47B**). This allows the

model to achieve exceptional accuracy, even in cases of minimal splits (6). Probabilistic estimates also clearly identify reduced freeze confidence during diurnal cycles when particular predictors are missing and in the transitional periods (i.e., spring thaw, fall freeze-up).

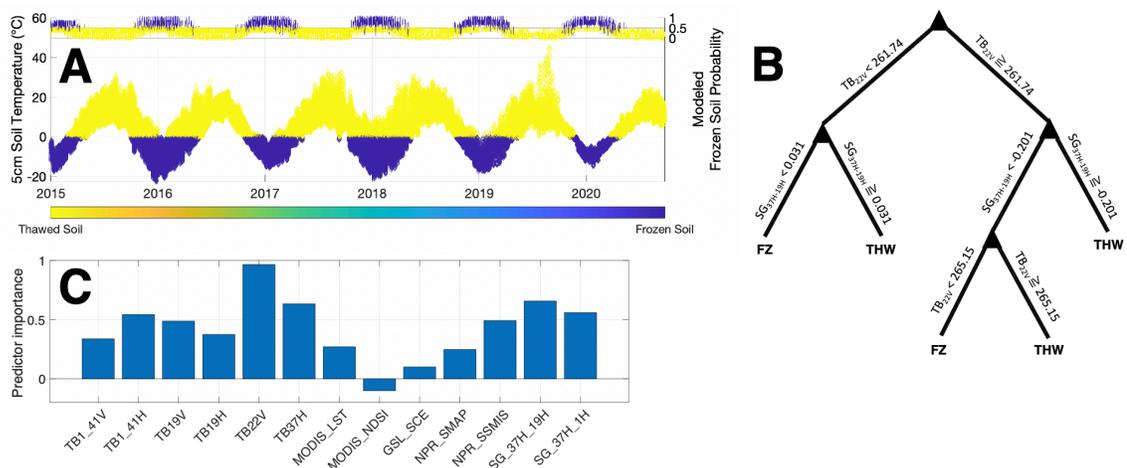


**Figure 47 Decision tree modeling for Alaska, USA site (1).** (A) In-situ soil temperature observations with soil FT classification and model estimated FT class (from DT) and probability (from RF) (top), (B) DT model framework, (C) Predictor importance from RF training

Over High Mountain Asia and the Tibetan Plateau, which is characteristic of a cold dry climate, model performance is reduced (~80%). Minimal snow, cloud, and vegetative cover as well as low soil moisture led to large daily swings in soil temperature across this region, even in in mid-winter (**Figure 48A**). This, coupled with a reduced freeze signal across dry soils result in higher frequency bands (i.e., 22 GHz) being more indicative of soil temperatures (**Figure 48C**). Spectral gradients are also shown to be of particular importance to classification in this region, as indicated by previous efforts of freeze detection in similar areas (*Zhao et al. 2011*). Diurnal temperature swings of as

much as 20 °C even in soil temperature, result in a challenging classification.

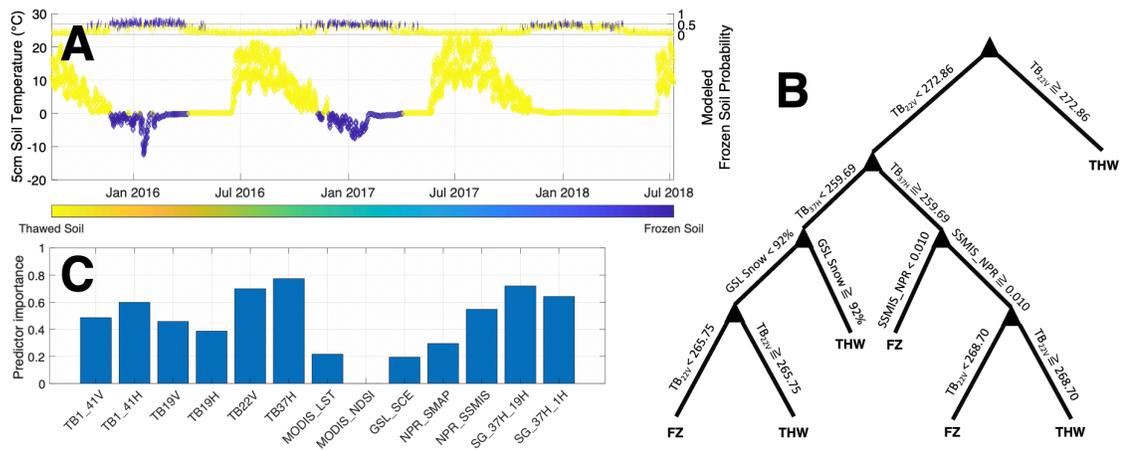
Importantly, this uncertainty is captured by frozen soil probabilities around 0.5 even in mid-winter by the RF classifier. Still, a simple 4-split DT model (**Figure 48B**) is shown to achieve nearly 80% classification accuracy even across the multi-year period, supporting its value as a classifier.



**Figure 48** Decision tree modeling for High Mountain Asia (HMA) site (2). See Figure 47 for description

In the mountains of Northern Italy, there exists clear FT seasonality and notable soil insulation by the snowpack in winter (**Figure 49**). In some cases, soil temperature is shown to fall below freezing and remain frozen for an extended period due to snow insulation (winter 2016, 2017). In contrast, snow is shown to have the opposite effect in the winter of 2018 as the soil temperature remains near, but above freezing for the duration of winter (**Figure 49A**). This adds considerable uncertainty to classifications during periods with snow cover as soil temperatures tend to remain close to the freezing

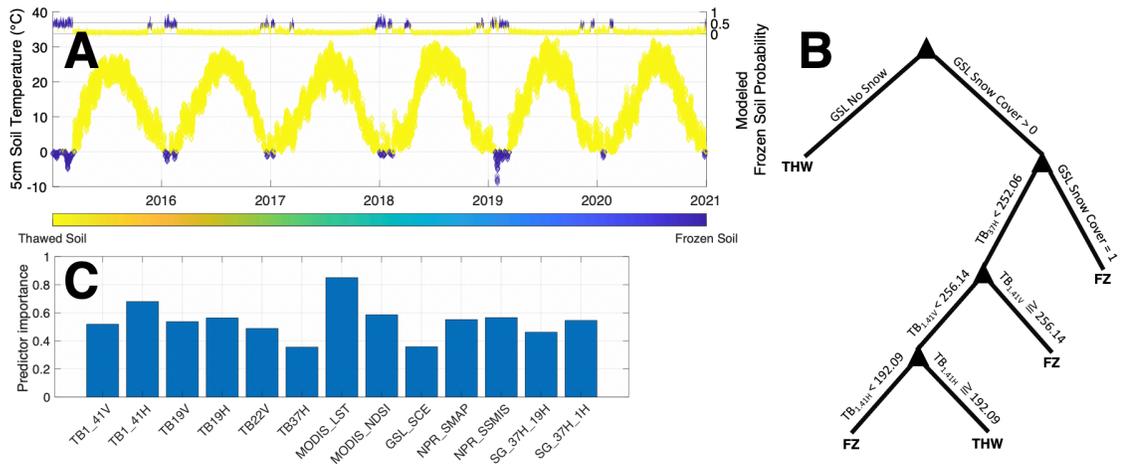
point and water in its liquid and frozen phases coexist. Due to these challenges, classification performance is somewhat reduced, approximately 82%. Relative to other locations, 1.41 GHz band observations have increased value to the RF model, suggesting their use for detecting thawed soil beneath snow. Even so, higher frequency observations more sensitive to surface layer and air temperatures (22 GHz, 37 GHz) are shown to have the highest predictive value (**Figure 49C**).



**Figure 49** Decision tree modeling for Northern Italy site (3). See Figure 47 for description

Lower latitude cold constrained regions are characterized by short annual periods of frozen ground (i.e., Illinois, USA). Overall model performance remains high (>90%), due to extended unfrozen periods. When freezes occur, they are frequently captured, however precision scores of the DT remain low (**Figure 50**). This suggests over classification of frozen ground across the region. The importance of snow cover presence is a critical predictor in the simple DT model; however, RF modeling suggest both SMAP  $T_b$  observations and LST provide the most predictive value (**Figure 50B**). Over such

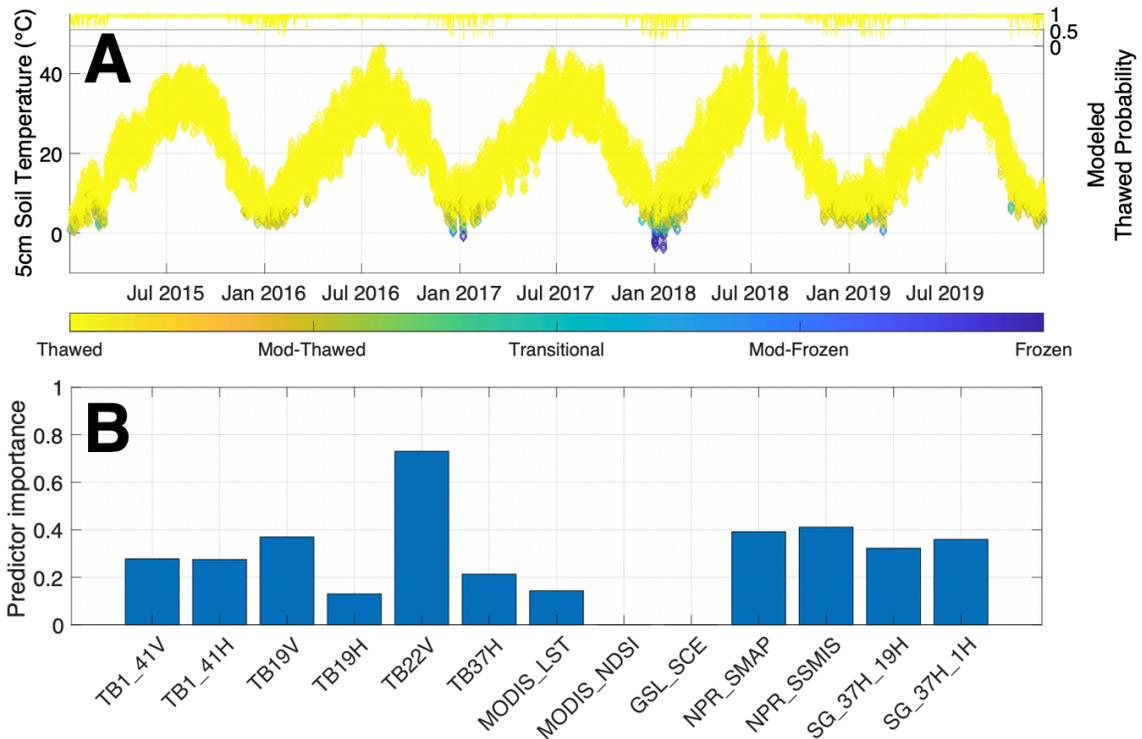
regions complex FT dynamics lead to very low L-band  $T_b$  observations ( $< 192$  °K) as well as elevated values ( $> 256$  °K) both resulting in freeze classifications. Lack of clear FT seasonality make the Illinois site a challenging case study. Here, the inclusion of the freeze probability metric is particularly valuable for capturing this uncertainty.



**Figure 50** Decision tree modeling for Illinois, USA site (4). See Figure 47 for description

Over drier temperate regions such as Texas, soils may not freeze at all in any given year (**Figure 51**). This makes it particularly challenging to train a single DT model to capture soil freeze events without introducing excessive model complexity and many splits. The optimized DT model in this case predicts no freezes (no splits), as instances with frozen soil occur less than 1% of the time. This results in an accurate model but one with no predictive value for freeze detection. For this reason, no decision tree is shown in **Figure 51**. However, while no periods are classified as frozen with the DT, the RF algorithm shows reductions in probability of thawed soil during cold periods (**Figure 51A**). Regions which experience rare freeze events are particularly challenged using a

binary classification for soil. Even though diurnal cycling in winter can result in sub-freezing temperatures, snow, and surface freeze events. Thus, the introduction of new multi-class rules could be considered in future efforts, which distinguish between sub-surface and air-skin freezes (i.e., frost).



**Figure 51** Decision tree modeling for Texas, USA site (5). (A) In-situ soil temperature observations with 5-class freeze category indicated with binary FT estimates (all thawed, DT) and associated thaw probability (RF) and (B) Predictor importance from 5-class RF model

All site performance summary statistics resulting from these comparisons are detailed in **Table 16**. Training at such point locations provides both a proof of concept and a method to examine how well DT-based models work for freeze classification. The accuracy of these predictive approaches is shown to vary depending on location. Introduced

confidence metrics and probabilistic classifiers also exhibit value for identifying challenging classification areas. This is illustrated by regularly capturing changes in model confidence and during cases when particular predictors are missing (e.g., diurnal differences in observations).

**Table 16 Summary results for freeze/thaw classification algorithms at point locations. Numbers correspond to locations in Figure 46**

Site (ID)	Binary FT Decision Tree				Binary FT Random Forest		
	N Splits	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Alaska, USA (1)	6	95%	0.949	0.972	96%	0.971	0.967
High Mountain Asia - HMA (2)	4	79%	0.859	0.561	80%	0.889	0.55
Northern Italy (3)	6	83%	0.623	0.683	82%	0.745	0.367
Illinois, USA (4)	4	93%	0.544	0.765	93%	0.806	0.232
Texas, USA (5)	0	100%	--	--	100%	--	--

Unique site signatures have suggested the value of regionally varied algorithms and the fusion of multiple remote sensing observations for improving FT classification. While feature importance varies between sites, generally snow cover and microwave  $T_b$  observations provide the best predictive value. Cloud obscuration of MODIS observations and different retrieval timings compared to the other bands may cause the reduced value of LST at most locations. The training of regionally specific models and the use of categorical variables to better constrain such models is shown to be paramount in applying such approaches globally. The incorporation of variables like climate, snow cover, topography, and land cover have relevance for improving model applicability at larger spatial scales. Overall, simple DT models and RF classifiers exhibit similar strong performance, but tuning hyperparameters is expected to have large implications for improving the balance

between model precision and recall performance. The use of RFs also has a particular advantage of outputting enhanced probabilistic classifications.

#### **4.3.2 Global Algorithm Training and Development**

##### ***Introduction and Modeling Approach***

The next stage in this project involves scaling up these methods to accurately capture FT transitions at continental scales. In doing so, several regions are identified through the use of both in-situ sites and long-term climatological records as being non-FT constrained. This includes those deemed tropical zones, arid hot regions, and temperate regions with dry winters. While the former two classes do not experience sub-freezing temperatures; the latter is known to experience such conditions in rare cases. Yet, due to the relatively few training sites ( $N = 10$ ) and no observed soil freeze events, this region is masked out. These temperate regions may be added into the FT constrained area in future efforts as additional training data becomes available.

The remaining regions contain nearly 50 million individual training features and include a high number of observed soil-freeze events ( $> 7$  million cases). The training set, derived from global site data between 2015 and 2019 (few sites extend to 2021), is then partitioned into each of the seven (7) remaining climate regions (see **Appendix B**). These include:

- 1) Arid, cold
- 2) Temperate, dry summer
- 3) Temperate, no dry season
- 4) Cold, dry summer

- 5) Cold, dry winter
- 6) Cold, no dry season
- 7) Polar/tundra

Thus, FT modeling presented herein details optimization and performance of algorithms for these regions.

### ***Parameter Sensitivity Analysis***

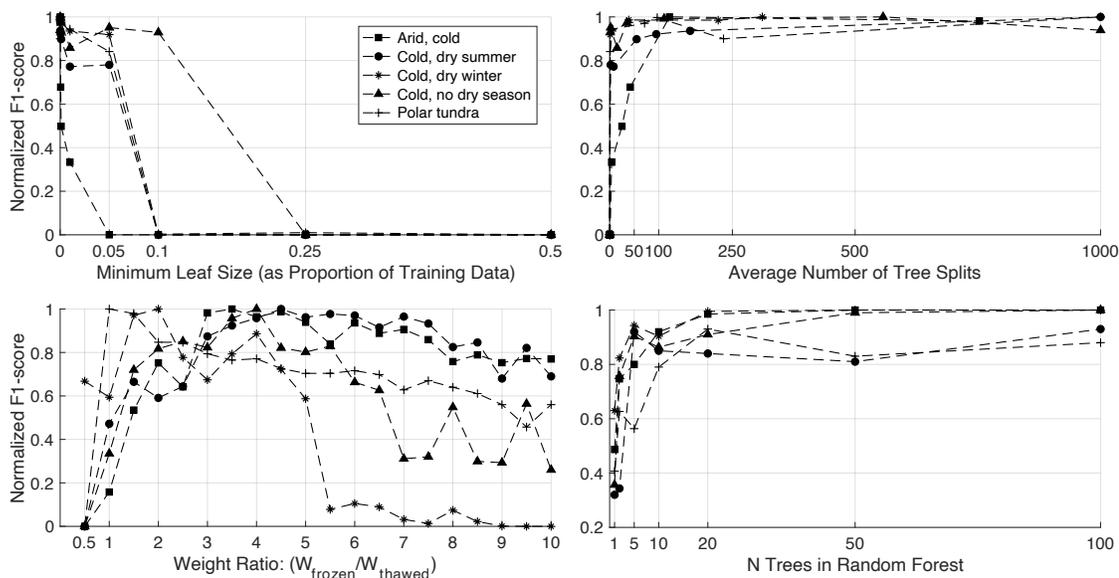
In order to balance the tradeoff between avoiding false freeze classifications (precision) and under freeze classification (recall) a normalized (by site) F1 metric is used to assess model performance. Three parameters are optimized for each region-specific classification model. These include:

- 1) Minimum Leaf Size. Taken as a proportion of the total size of the training set. This metric is directly linked to the number of allowable splits. For example, a minimum leaf size of  $> 0.5$  (50% of data) prevents any model splits, whereas small values (e.g.,  $< 0.01$ ) allow for many potential tree splits.
- 2) Label weights or weight ratio. This is taken as the relative value between weight parameters for frozen classifications ( $W_{\text{frozen}}$ ) to that of thawed classifications ( $W_{\text{thawed}}$ ). Tuning this value has particular importance in unbalanced datasets (many thaws, few freezes).
- 3) Number of trees. This metric defines the number of unique learners to use in creating the RF ensemble for each region.

The results for cold and polar climate regions are presented in **Figure 52**.

Weighting ratios are also tuned for both temperate classes but are not shown to the large imbalance in training set feature frequency (weight ratios > 100).

The top panels in **Figure 52** suggest that finer trees generally result in improved performance but may also reduce model robustness by overfitting. The degree to which this performance increased after approximately 100 splits is shown to be minimal. Thus, each ensemble member in the random forest is limited to a maximum of 100 splits. In most cases the actual number of splits in each member remained well below this level while still achieving strong performance.



**Figure 52** Sensitivity analysis for select random forest parameters

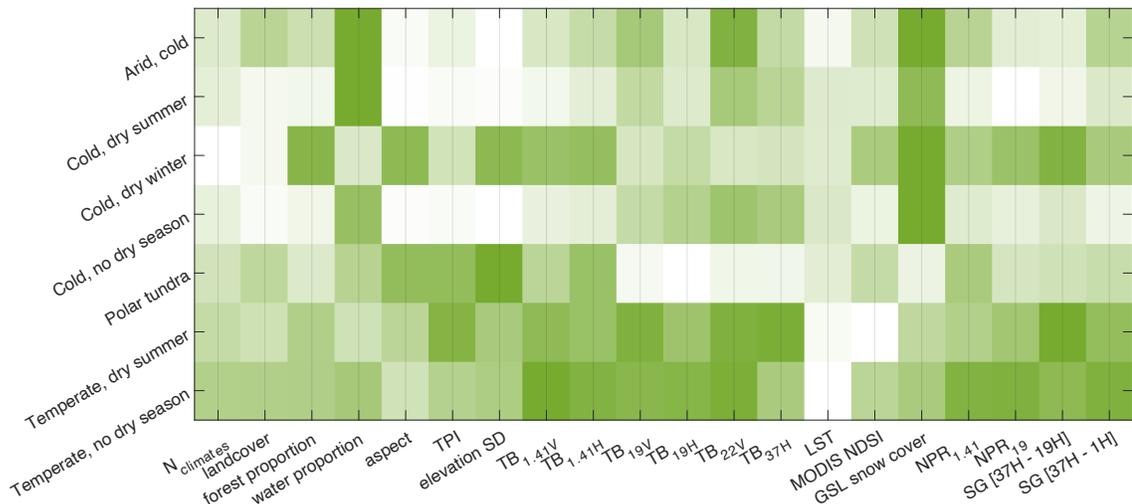
Next, the weight ratio is set to normally spaced values (0.5 intervals) from 0.5 to 10. Depending on the distribution of frozen to thaw classifications in the training set, this value was optimized. For instance, regions with similar numbers of frozen and thawed classes (i.e., polar, and cold, dry winters) are seen to have an optimal weight ratio near 1. Meaning the value of a correct frozen classification has the same influence in model training as a correct thawed classification. For other sites, model performance is optimized with weight ratios set near 4.

Lastly, another critical hyperparameter in tuning of RF models is the number of learners, or trees. This determines the number of unique trees that will make up the random forest. Results suggest that greater than 20 trees cause minimal improvement in the model. Still, more trees can result in a more ‘random’ classifier with improved generalization capacity by using more data subsets and through the creation of a larger ensemble. Thus, for most cold and near polar regions, 100 learners are selected. In minimally FT constrained regions (i.e., temperate areas) this was increased to 200 to improve the likelihood that numerous learners will include multiple freeze-labeled feature vectors in their training.

For models with larger training sets, around 20% of features were sampled for each random tree (In Bag Fraction = 0.2). For those with smaller datasets or fewer freezes, this value was set to around 50% (0.5). Additional sensitivity analysis identified that this parameter did not affect model accuracy much in the cases of large datasets but remained important in smaller datasets with few freezes.

### *Feature Importance and Climate Specific Model Results*

Using the final trained models (**Table 17**), feature importance estimates are computed by region using the permuted delta error approach (**Figure 53**). To assess prediction value across all predictors, each is kept in the model, and the relative importance is presented for each climate region. In cold, non-polar/tundra regions, snow cover and the amount of surface water adjacent to a given training set have particular importance to identifying soil FT states. In drier cold regions with little snow, additional variables relating to the topographic profile and SMAP  $T_b$  observations (1.41 GHz) are found to have increased model importance.



**Figure 53 Normalized feature importance by climate class. Darker green indicates variables that are more valuable to the random forest model. Based on permuted delta error method**

In temperate regions, microwave observations and metrics derived from such observations show similarly high value in delineating frozen soil. This outlines the distinct climatological and regional differences in the relative importance of predictors. Interestingly, the value of physical surface skin temperature (LST) provides little to no added predictive value within each regional model. This suggests that omission of LST may be made in future as it does not have explicit sensitivity to soil temperature and data availability from cloud obscuration limit its usefulness as a predictor.

**Table 17** presents final training results across all climates. When weighted by global FT constrained land area, the trained model achieves nearly 88% accuracy. This is considering the complete dataset, even as models are trained using only a small fraction of the total feature sets. Specifically, 8% to 33% of the available training data is used for training. This value depends both on the amount of available training data and the relative density of sites in each unique  $1^\circ \times 1^\circ$  pixel represented. This approach prevents any one region with high data density from dominating the model training phase. As such, all indications suggest a spatially robust model. Accuracy remains comparable if not better than the existing stated accuracy of current classifiers (~80-85%) and is soil specific. Even in regions with few freeze classifications, recall scores indicate soil freezes are captured over 50% of the time. Conversely, this results in reduced F1 scores through an increased relative rate of false freezes. Even so, models in these temperate regions achieve greater than 99% classification accuracy.

**Table 17 Model training summary and performance scores by climate region**

Climate Region	Training					Performance		
	Number of represented 1° x 1° pixel regions	Soil Freeze % in Training Data	Total Training Feature Sets	Used for Training	Ntrees	Classification Accuracy	Recall	F1-score
Polar	47	45.3%	6.9 M	17%	100	82.7%	0.875	0.818
Cold, Dry Winter	8	40.7%	0.6 M	33%	100	90.1%	0.907	0.861
Arid, Cold	104	14.2%	5.6 M	14%	100	86.9%	0.786	0.628
Cold, Dry Summer	7	12.0%	3.8 M	21%	100	83.0%	0.650	0.511
Cold, No Dry Season	275	11.6%	25.6 M	9%	100	83.5%	0.789	0.625
Temperate, No Dry Season	104	0.1%	3.9 M	8%	200	99.4%	0.503	0.249
Temperate, Dry Summer	31	0.1%	2.2 M	10%	100	99.0%	0.771	0.253
<b>Overall Performance in FT Constrained Regions, Weighted by Class Land Area</b>						<b>87.6%</b>	<b>0.760</b>	<b>0.590</b>

### 4.3.3 Global Application and Validation

#### *Model Framework*

As the final stage of product development, trained classifiers are applied globally for the 2020 validation period at 0.05° (~5 km) resolution. The general workflow is described in **Figure 54**. As a first stage, non-FT constrained or perennial snow- and ice-covered areas are masked as never frozen or never thawed, respectively. Snow and ice regions are classified according to MODIS IGBP land cover, whereas non-FT constrained lands include tropical, arid hot, and temperate, dry winter climates. Pre-trained models are then applied to specific relevant climate regions and output binary soil FT classifications and associated frozen probabilities (0 – 1) for both morning (a.m.) and evening (p.m.) overpass periods. These are representative of soil FT state conditions at approximately 6 a.m. and 6 p.m. local solar time. Using masking information and probabilistic outputs, a flagging procedure is also implemented as follows:

**Flag 0:** No flag (frozen probability < 25% or >75%)

**Flag 1:** Frozen probability between 40% and 60% (high uncertainty)

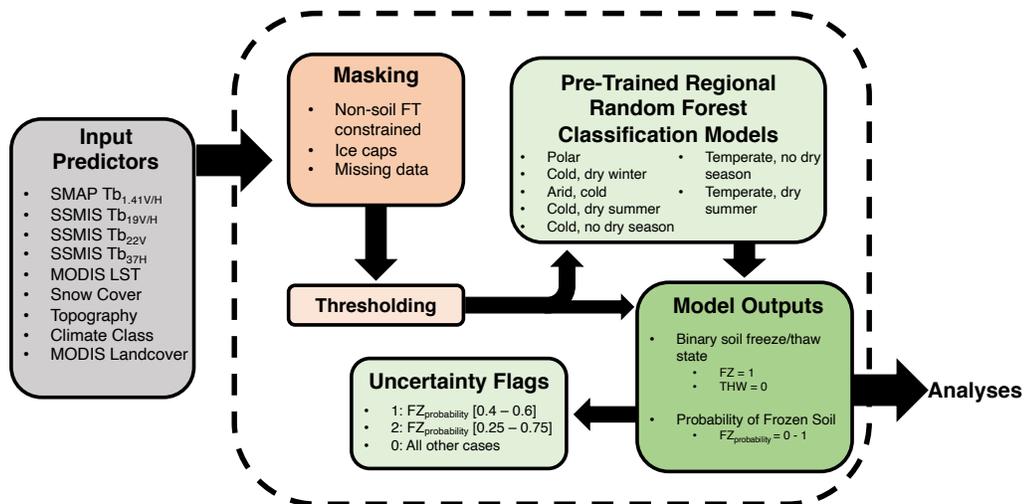
**Flag 2:** If not Flag 1, frozen probability between 25% and 75% (moderate uncertainty)

**Flag 3:** Always frozen, ice caps

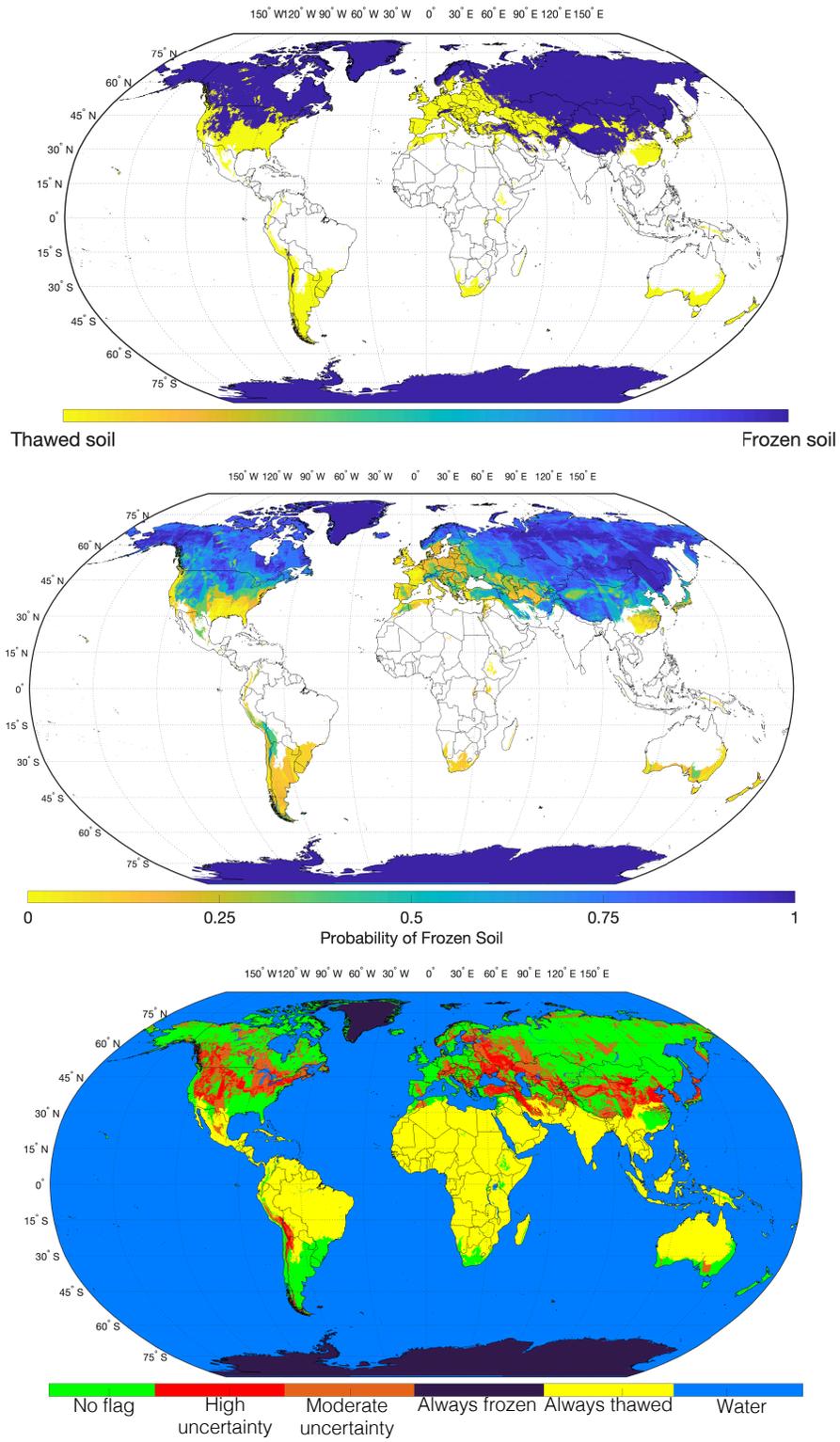
**Flag 4:** Non-freeze/thaw constrained and not classified

**Flag 5:** Masked water, defined as regions when >50% pixel is surface water

The introduction of an additional thresholding step allows for new rules to be implemented in the future to avoid false freezes (or thaws) which can be applied globally. For example, current SMAP products set areas with  $T_{bV}$  observations  $> 273$  °K as automatically thawed. While this is not applied herein, similar rules can be implemented to improve model performance. **Figure 55** details model output layers for the morning period on January 30, 2020. The top panel presents binary classifications, the second panel frozen soil probabilities, and the bottom presents the distribution of masked regions and uncertainty flags.



**Figure 54** Framework for global freeze/thaw classification using random forests



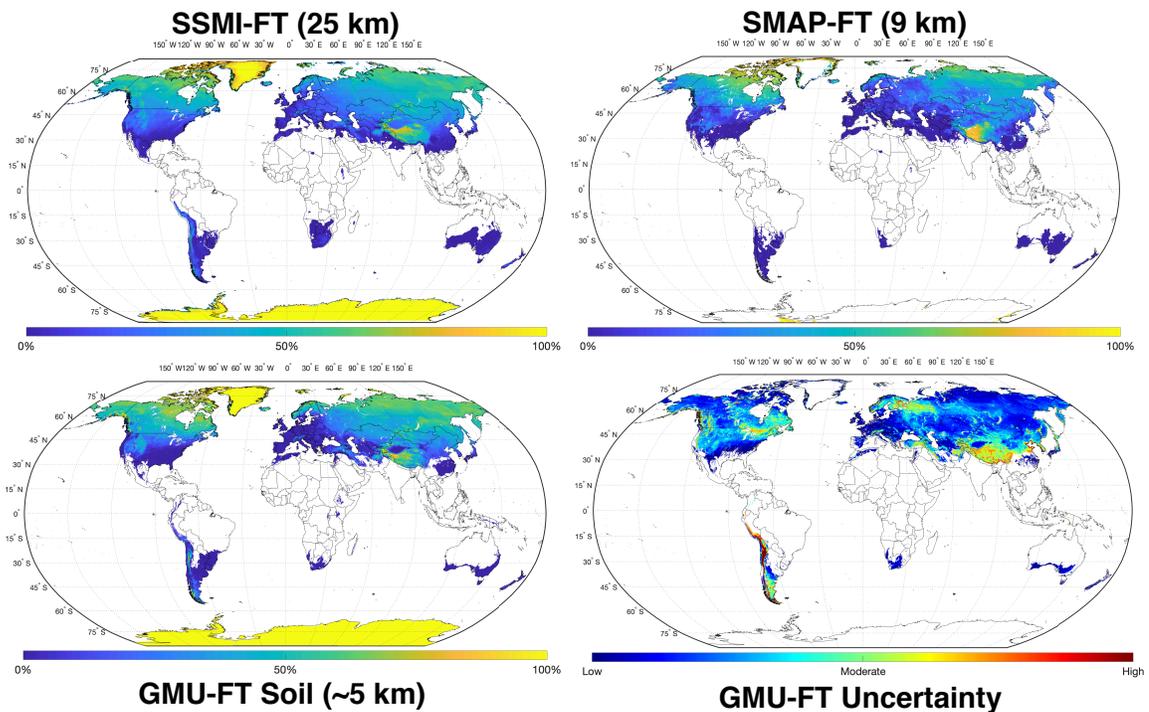
**Figure 55** Example outputs from GMU-FT global model for morning (a.m.) period on January 30, 2020

The inclusion of probability metrics as presented can be used as a proxy for estimate uncertainty. This is an additional benefit of classification through the RF approach as compared to existing methods. As an example of its value, the regions with missing data from a given sensor on a given day can be directly characterized (see swath areas in middle and bottom panels, **Figure 55**). Since full global repeat cycles of SMAP and SSMIS are on the order of 8-days, certain regions in the mid-latitudes are not imaged on a daily basis. No interpolation is used herein, thus the number of assumptions is minimized and the product presents estimates based on the best available data for a given time period.

### ***Global Validation and Product Comparison***

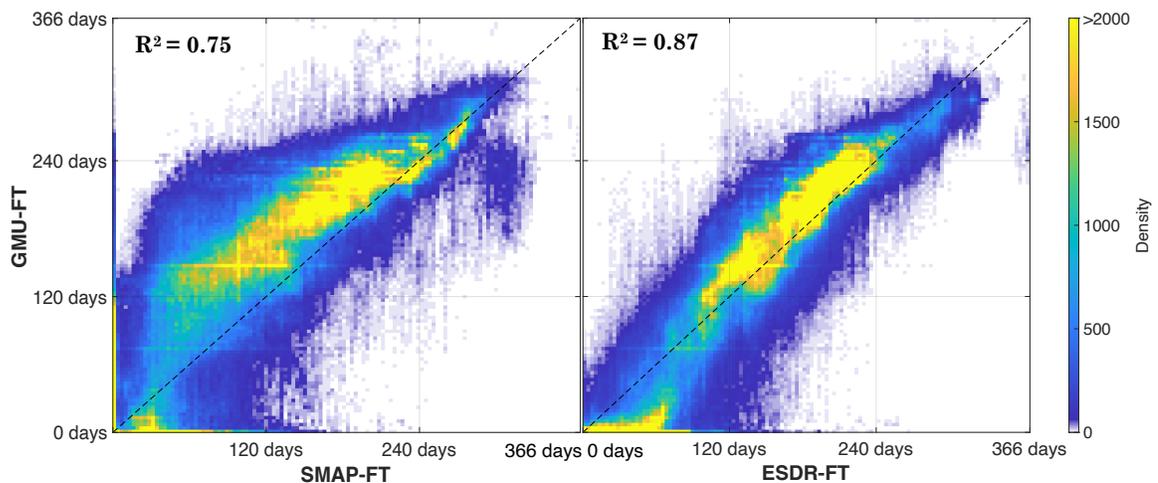
Using the year 2020, the proportion of the year which is classified frozen by various products is presented in **Figure 56**. Results show comparable distributions across each product. In comparisons to SSMI-FT, spatial distributions of annual frozen proportion, while similar, have clear reductions over temperate regions (e.g., CONUS and much of Europe). Due to the introduction of a soil specific classifier, this approach suggests that SSMI-FT indicates freezes in these regions when the soil remains thawed. This is expected based on observational capabilities of Ka-band, which is not sensitive to the 5 cm depth soil layer and remains more representative of air temperatures in these regions. Over such regions, SMAP classifications remain more similar relative to GMU-FT soil classifications. This agrees with previous results and the large body of literature suggesting SMAP L-band observations as a more effective soil state discriminant.

The introduction of global uncertainty metrics is also detailed in **Figure 56**. The metric presents the relative frequency of uncertainty flags (40 – 60% frozen soil probability) globally for the year 2020. Confidence is shown to be reduced in boreal forest regions as well as over HMA and Andes mountains in South America. Based on reductions in data availability closer to the equator, vegetation complexity, topographic variability, and low moisture conditions in the case of HMA, this assessment of uncertainty underlines known observational limitations but is output as a part of the new FT classifier.



**Figure 56** Global validation of GMU-FT against operational FT products

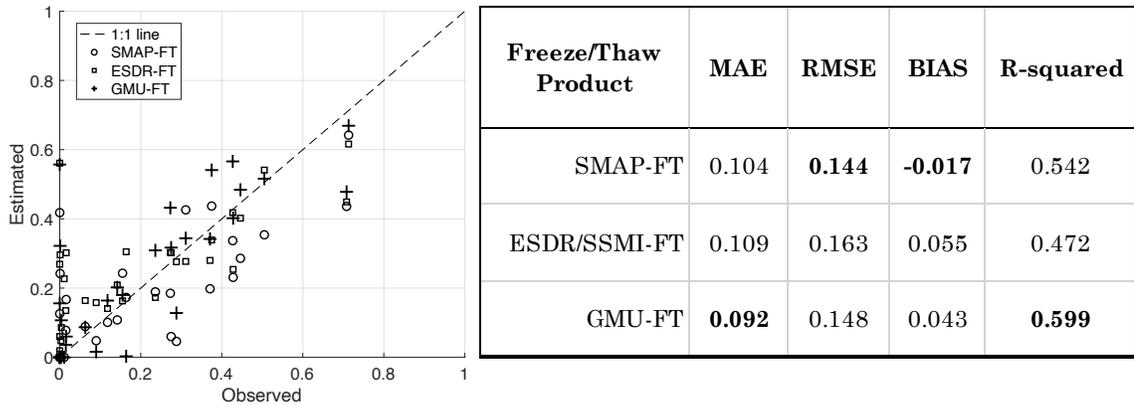
In quantifying the relationship between existing products for 2020, a comparison of the length of the frozen period (days) is performed over the global domain (**Figure 57**). Regions marked as always frozen are excluded from such comparisons. Results indicate the new George Mason product (GMU-FT) to be more closely related to the ESDR-FT record in terms of the duration of the frozen period ( $R^2 = 0.87$ ) compared to SMAP-FT ( $R^2 = 0.75$ ). Specifically, the GMU-FT classifier suggests a considerably longer frozen period relative to SMAP-FT in mid-latitude cold regions (60 – 180 frozen days) as opposed to the most persistent cold regions (i.e., Tundra) in which frozen conditions dominate more than half of the year. In contrast, across both product comparisons, a reduction in the number of freezes is observed by GMU-FT in more temperate regions (<60 frozen days). Based on limitations of current classifiers at detecting freezes in transitional regions when soil remains thawed (**Chapters 2 – 3**), this indicates improved observational capacity of the soil FT state by GMU-FT in such areas.



**Figure 57** Density plot comparison (0.5° grid) total number of frozen days between freeze/thaw classifiers

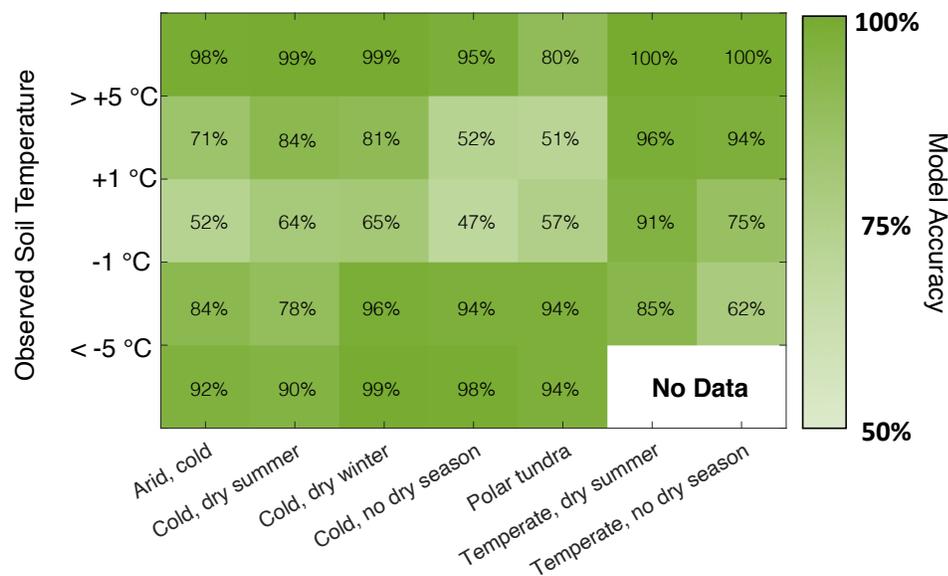
Using a random subset of 30 soil temperature observing sites with continuous data records for the year 2020, the annual frozen period expressed as a fraction of the year is compared to remote sensing observations (presented in **Figure 56**). Against such sites, performance metrics suggest GMU-FT classifications have the highest correlation to the observed frozen period durations across all sites and products (**Table 18**). GMU-FT also presents the lowest mean absolute error (MAE) and is improved relative to soil FT states across all performance metrics compared to SSMI-FT. MAE of GMU-FT (0.09) equates to an approximately 5-day improvement in accurately capturing the duration of frozen soil relative to other products (0.10-0.11).

**Table 18 GMU-FT performance summary of annual frozen proportion compared to FTSPs for the year 2020**



Most error is induced by classifications in far northern regions (>59° N) in which snow cover insulates the surface and soil temperatures remain slightly above freezing. In this case (similar to **Figure 49**), the observed soil freeze proportion remains near 0, even

as products suggest >40% (55% for GMU-FT) of the period is frozen. In such situations, FT classification remains particularly challenging. SMAP-FT is shown to be the most effective at accurately capturing soil freezes in these cases, though uncertainty remains high as soil states are known to vary considerably over small distances depending on snow cover, moisture, and other properties (**Chapter 3**). This remains a limitation of using point observations for validation over large regions. Furthermore, soil temperature observational uncertainties of embedded sensors on the order of 0.5 to 1.0 °C and freezing point depression make these borderline cases particularly difficult to classify. Relative FT uncertainty and the likelihood of mixed FT states increase coincidentally in these cases. **Figure 58** illustrates the reduction in model accuracy with soil temperatures within 1 °C of the freezing point (~64%) as compared to cases in which soil temperatures are observed to be > 5 °C or < -5 °C (>95% classification accuracy).



**Figure 58** GMU-FT classification accuracy of soil freeze/thaw state by temperature category using validation set

Early validation efforts GMU-FT performance is comparable if not improved relative to existing classifiers, especially in capturing soil freeze/thaw state. The effectiveness of spatially varied algorithms and the use of random forests offer effective FT classification with the benefit of improved characterization (i.e., soil specific state and quantifiable uncertainty or model confidence metrics).

### ***Limitations***

Additional validation efforts over a longer duration (2015 – 2021), with additional sites, and with consideration to new metrics such as freeze (and thaw) onset timing should be undertaken. Exploring classification accuracy against a larger number of in-situ sites and spatially distributed temperature data remains a next logical step in further validating these outputs. Density of observing sites is very high over North America and Europe, moderate over High Mountain Asia, and relatively sparse elsewhere. The expansion of this dataset as more observations become available will be important to validating the model and improving classification scores over regions in the southern hemisphere and much of Eurasia. The challenge of comparing point to satellite observations also introduces inherent errors and uncertainties, especially in cases where sub-pixel variability is high. Again, the use of ensemble learners such as random forests is intended to allow individual model errors to negate each other, providing the best possible classifier. Or in cases of mixed conditions, probabilistic freeze estimates around 0.5 can be used to quantify uncertainty and the potential existence of mixed soil states.

Results herein also show that a number of predictors may not provide particular value to models in certain regions, thus creating unnecessarily complex classification

models. This is especially true in the case of observations with high covariance, which includes similar microwave sensing bands. Model simplification in future versions can further improve generalizability. In contrast, the considerable training set size and diversity of predictors may help in presenting improved uncertainty scores and did not present limitations regarding model processing demands and prediction time. The validity of this assumption should be investigated further in the future.

While machine learning comes with a multitude of benefits, achieving model interpretability and creating a connection between predictions and known physical principles can be a challenge. In many cases, ML models are seen as a black box and physically based approaches are preferred. This is especially true for operational modeling, as unseen data may limit detection in anomalous conditions. Still, in the case of decision trees, trained models are representative of optimized multi-dimensional threshold algorithms. Such approaches provide the backbone of existing FT classifiers (SMAP-FT and ESDR-FT). As such, the question of model generalizability in the case of GMU-FT would also draw into question existing detection approaches. Subsequently, the sufficiently large and diverse observational dataset, good validation results, and special attention to the avoidance of overfitting suggest this approach as both scientifically valid and robust.

#### **4.4 Chapter Outcomes**

**Chapter 4** presents the culmination of work included within this dissertation. The training and implementation of an enhanced soil-specific freeze/thaw classification model is detailed. Specifically, decision trees and random forests are applied using satellite

microwave observations, static land surface variables, and snow cover as predictors.

First, at various in-situ stations (i.e., point sites) within uniquely different regions, then at a global scale using unique classifiers by climate region. Validation results presented indicate strong model performance and classifier skill approaching 90% globally. Other major outcomes include:

- *Regionally, freeze/thaw signals and the value of various predictive features is shown to vary. Differences in climatology including temperature regimes, snowfall, and moisture states outline the value of spatially variable classifiers*
- *The accuracy of random forest classifiers for defining soil FT states is demonstrated, along with the value of introducing a prediction confidence (i.e., frozen soil probability) metric*
- *Classifications from this new model, termed GMU-FT, are computed for the year 2020 and validated against existing remote sensing freeze/thaw products from SMAP and the Earth System Data Record (ESDR, SSMIS)*
- *Results suggest comparable and largely improved accuracy relative to these products for soil freeze/thaw state detection*

## **5 RESEARCH CONTRIBUTIONS, FUTURE OPPORTUNITIES, AND DISSEMINATION**

A comprehensive assessment of freeze/thaw classifiers, their limitations, and opportunities for their improvement is completed in this project. The drivers of spatial variability in freeze/thaw cycling are also explored using a combination of field work, airborne sensing with Unpiloted Aerial Vehicles, model outputs, and global remote sensing observations. The value of combining the strengths of several observations and regionally variable classifiers is shown to have exceptional predictive capacity. As a result, a new algorithm for soil specific freeze detection is derived and associated classification uncertainty metrics are also introduced to improve current freeze classifiers, many of which are discovered to have varied physical meanings.

Regarding project contributions and outcomes, several important findings are outlined by task as follows:

### **Task 1. Exploring existing freeze/thaw records; their limitations, sensitivities, and relationships to land surface variables (Chapters 2 - 3)**

- **Existing records have varied relationships to surface temperature states including air, skin, and soil temperatures**
- **Currently, products remain most representative of air temperature**
- **Globally, existing metrics are found to agree 83.5% of the time, with highest uncertainty due to variable moisture, climatology, and land cover**

- **Coastal areas and those with high surface water fraction, snow, and ice also challenge current classifiers**

**Task 2. Determine the driving factors of freeze/thaw transitions at higher resolutions, the potential for resolution enhancement, and utility of new classifiers through focused regional studies and in-situ/ground observations (Chapter 3)**

- **Snow cover is identified as an especially important control on soil freeze/thaw status**
- **Binary classifiers are insufficient to accurately represent surface states due to high fine-scale variability**
- **Topography and land cover type are found to provide important controls on the surface energy balance and are the dominant drivers of sub-pixel variability**

**Task 3. Algorithm and predictor selection, sensitivity analyses, and validation of a new data driven global freeze/thaw classifier (Chapter 4)**

- **Random forest classifiers are presented as an effective, global scale data fusion approach for freeze detection**
- **Regionally varied tuning of hyperparameters is used for model optimization as regional differences in predictor importance are shown**
- **Early validation efforts of GMU-FT suggest global classification accuracy of soil states near 88% with comparable or improved performance relative to existing operational freeze/thaw products**

Fusion of both physical principles and machine learning through random forests is shown to be a promising combination for global scale classification tasks. Combining known sensitivities and band physics, with an exhaustive network of in-situ observations provides a classifier directly optimized to the ground truth. Future efforts to modeling freeze states leveraging these lessons could provide estimates of surface freeze/thaw states on any scale as long as basic surface properties, land cover, snow cover, topography, and air temperature is known. Continued refinement and global validation of this approach is ongoing as well as is the expansion of the data record to include the period of 2015 – present day. This will enable assessments of variability in interannual frozen periods and freeze/thaw cycling. Next, the transition from algorithm development to implementation for forecasting, agricultural decision making, hydrological modeling, and global change studies present potential applications for such a dataset. Continued collaborations with the SoilTemp community to grow the usage of this dataset and support its improvement are planned.

Relevant publications coming from this work have been cited in the text and included in **Chapter 1**, which also details two additional manuscripts to be published. Many of these results have been shared at leading conferences such as for the American Geophysical Union (AGU; 2018, 2019, 2020, 2021) and the American Meteorological Society (AMS; 2019). All relevant codes used to derive the GMU-FT classification are available publicly on GitHub ([https://github.com/jjohns60/GMU\\_FT](https://github.com/jjohns60/GMU_FT)). Any requests for data sets produced herein should be sent to **jjohns60@gmu.edu**. Efforts to publish the extended record with the National Snow and Ice Data Center (NSIDC) are ongoing.

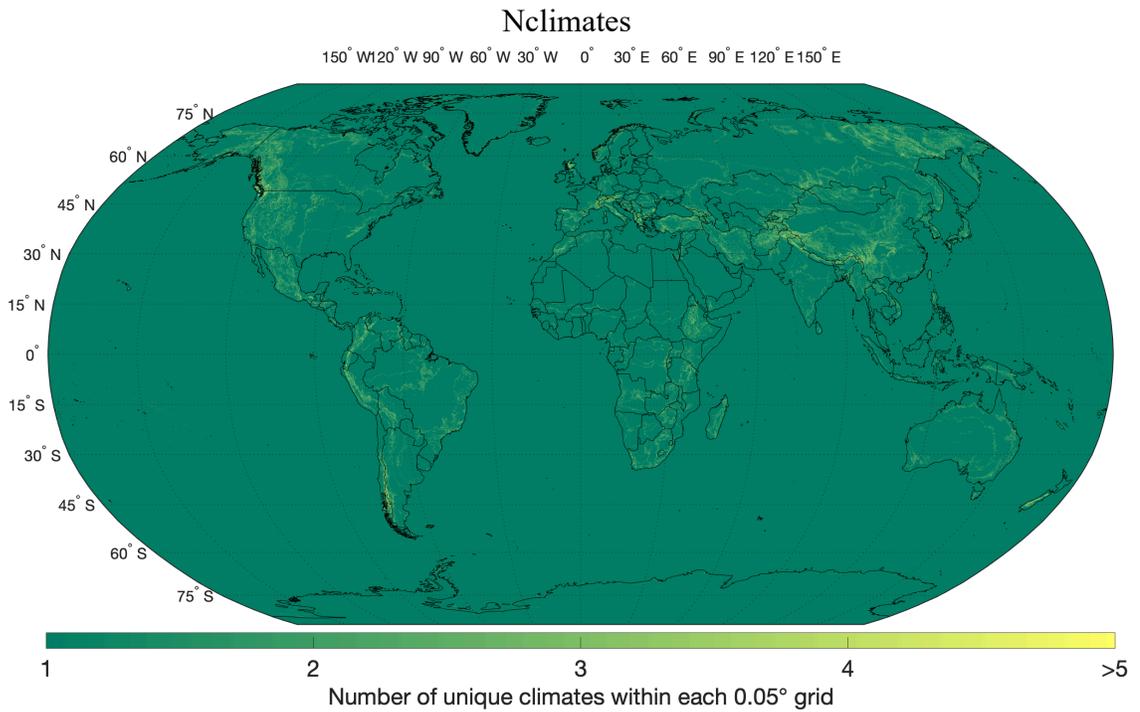
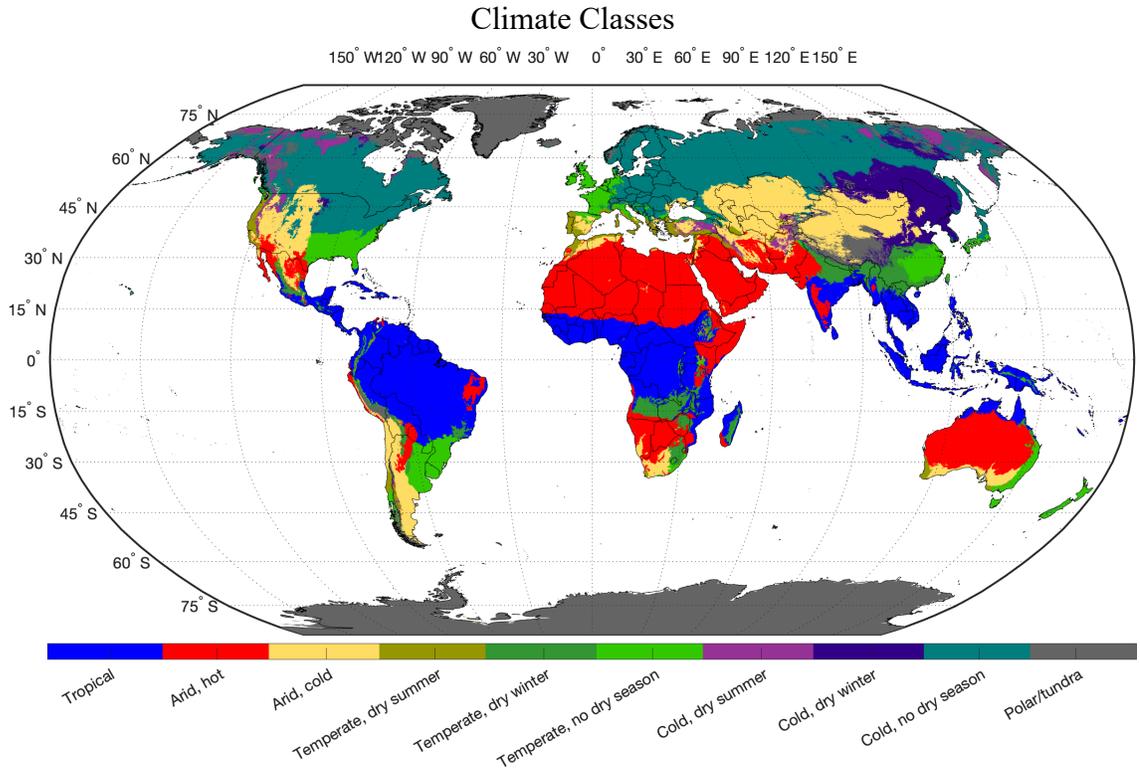
## APPENDIX

### Appendix A: Model Prediction Features Detailed

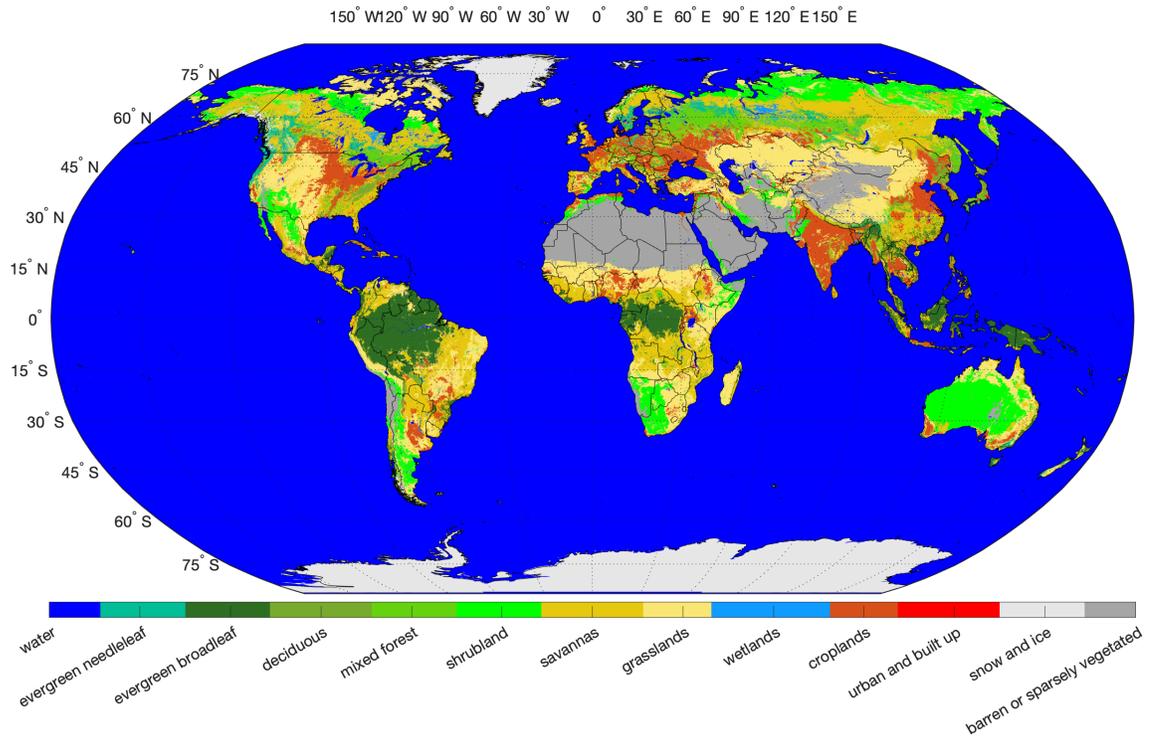
Freeze/Thaw Model Predictor Descriptions		
Feature	Description	References
Climate class	Köppen-Geiger classification of present-day climate class (1980-2016), according to methods and outputs in [1]. Reclassified from 30 classes to 10: A-classes combined to 'Tropical'; Bwh, Bsh to 'Arid, hot'; Bwk, Bsk to 'Arid, cold'; Cs-classes to 'Temperate, dry summer'; Cw-classes to 'Temperate, dry winter'; Cf-classes to 'Temperate, no dry season'; Ds-classes to 'Cold, dry summer'; Dw-classes to 'Cold, dry winter'; Df-classes to 'Cold, no dry season'; ET and EF to 'Polar'	[1] Beck et al., 2018
Nclimates	Number of unique climate classes (Original 30 class scheme, plus open water class) falling within each 0.05° grid cell	
Land cover	Land cover classification according to MODIS MCD12C1 products [2] using the IGBP 17-class classification scheme. Deciduous classes combined, shrubland classes combined, savannas combined, and croplands combined to reduce total classes to 13	
Forest proportion	Indicates an estimate of the average proportion of forest cover contained in each 0.05° grid cell. Uses data from [2] and taken as the average during the 6-year period from 2015-2020. Is a proportion ranging between 0 and 1, computed by the following equation: $forest\ proportion = [0.8(\%LC\ forested\ classes\ 1-5) + 0.45(\%LC\ woody\ savanna\ class\ 8) + 0.2(\%LC\ savanna\ class\ 9) + 0.1(\%LC\ cropland\ class) + 0.25(LC\ cropland/natural\ veg\ class\ 14)]/80$	[2] Friedl and Sulla-Menashe, 2015
Water proportion	Indicates the average proportion of surface water within each 0.05° grid cell. From [2], the included cell water proportion value	
Aspect	Indicates the aspect of a given site depending on its slope, calculated from 1/60° elevations in [3]. Is categorized using a scheme similar to [4], but adjusted hemispherically. Using this method 8 cardinal directions are partitioned as such that odd numbers (1,3,5,7) indicate east facing aspects and even numbers (2,4,6,8) indicate west facing. However, in the northern hemisphere the north facing slopes receive values between 1 and 4 (south facing 5 – 8) and in the southern hemisphere the south facing slopes receive values between 1 and 4 (north facing 5 – 8). Cases in which normal vector is approximately vertical is indicated by 0	[3] ETOPO1; [4] Gómez-Plaza et al., 2001
Topographic Prominence Index (TPI)	Computed for each 0.05° grid cell using 1/60° data from [3]. The metric is computed as the elevation of the center pixel, minus the average of all surrounding pixels. Is normalized (0 - 1) where 0 indicates <= -150 m and >= 150 m. Lower values indicate surrounding topography is more prominent than center of pixel, while higher values indicate the central pixel as higher than the surroundings.	
Elevation St. Dev	Calculated as the standard deviation of all elevations within each 0.05° grid cell using the 1/60° resolution elevation inputs from [3]. Values are normalized (0 - 1) using the range 0 m to 500 m. Values above or below this range are set to 0 and 500 respectively before normalization	
1.41V GHz TB	Vertically and horizontally polarized 1.41 GHz (L-band) microwave brightness temperatures derived from SMAP observations [5],[6] at 3.125 km resolution	[5] Long et al., 2019; [6] Brodzik et al.,
1.41H GHz TB		
19V GHz TB	Assorted vertically and horizontally polarized microwave brightness temperature observations in bands ranging from 19 GHz (K-band) to 37 GHz (Ka-band). Higher frequency	[7] Brodzik and Long, 2016; [8] Meier and Stewart, 2020
19H GHz TB	37 GHz observations are matched to 3.125 km grid, whereas others use 6.25 km grid spacing	
22V GHz TB		
37H GHz TB	[7], [8]	
Land Surface Temperature	Thermal infrared band observations of physical surface skin temperature. From MODIS global 0.05° skin temperature products 11C1 v061 [9] (4x daily, Terra/Aqua)	[9] Wan et al., 2021
Normalized Difference Snow Index (NDSI)	Provides the percentage of snow observed over land within each MODIS 0.05° global grid cell. Can be cloud obscured, see [10]	[10] Hall and Riggs, 2021
GSL Snow Cover Extent	Weekly, binary snow cover presence from the Rutgers University global snow laboratory at 24 km grid spacing [11]	[11] Robinson and Estilov, 2021
1.41 GHz NPR	The normalized polarization ratio (NPR) computed as the difference between 1.41 GHz vertical minus horizontal polarization brightness temperatures divided by their sum	
19 GHz NPR	Same as above, but computed using 19 GHz H- and V-polarized observations	
SG 37H GHz - 19H GHz	Spectral gradient (SG) calculated as the difference between the 37 GHz and 19 GHz horizontally polarized brightness temperatures, divided by the frequency difference (18 GHz). In units of °K/GHz	
SG 37H GHz - 1.41H GHz	Same as above, but computed between the 37 GHz and 1.41 GHz observations	

Note: All features used in training are taken as the nearest value to each in-situ site, both in space and time

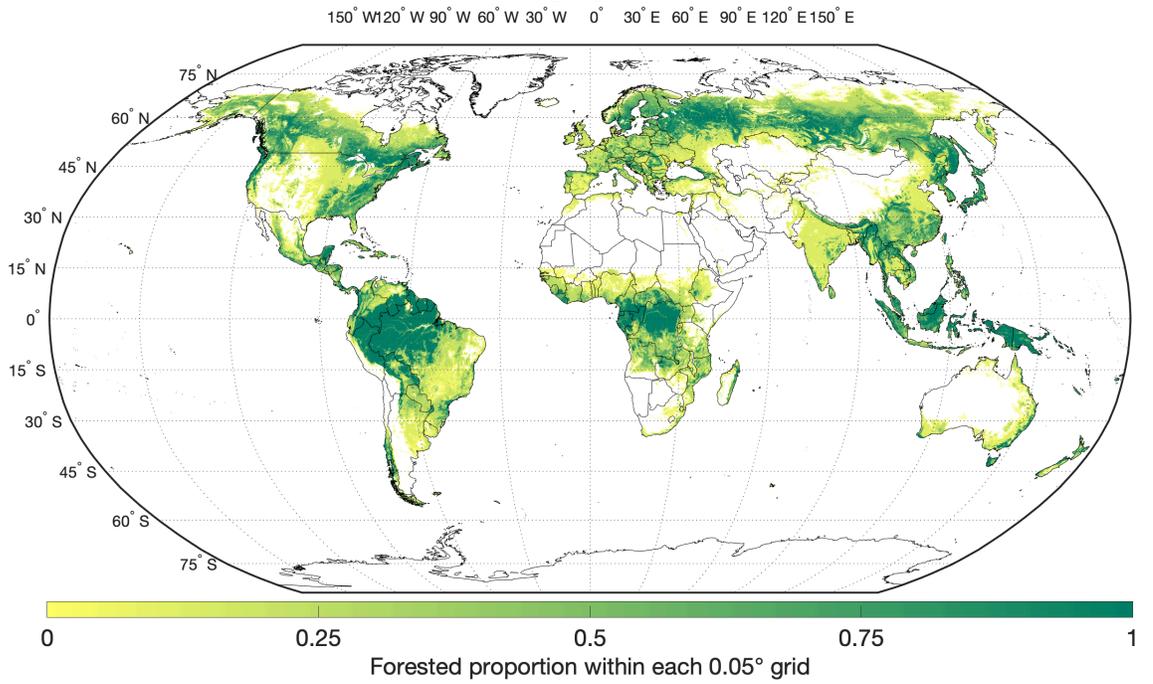
## Appendix B: Global Maps of Static Predictors at 1/20° (0.05°) resolution



## Land cover

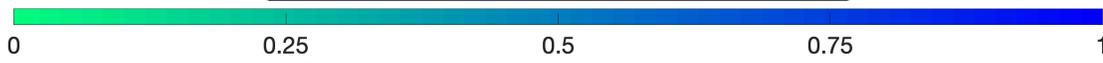
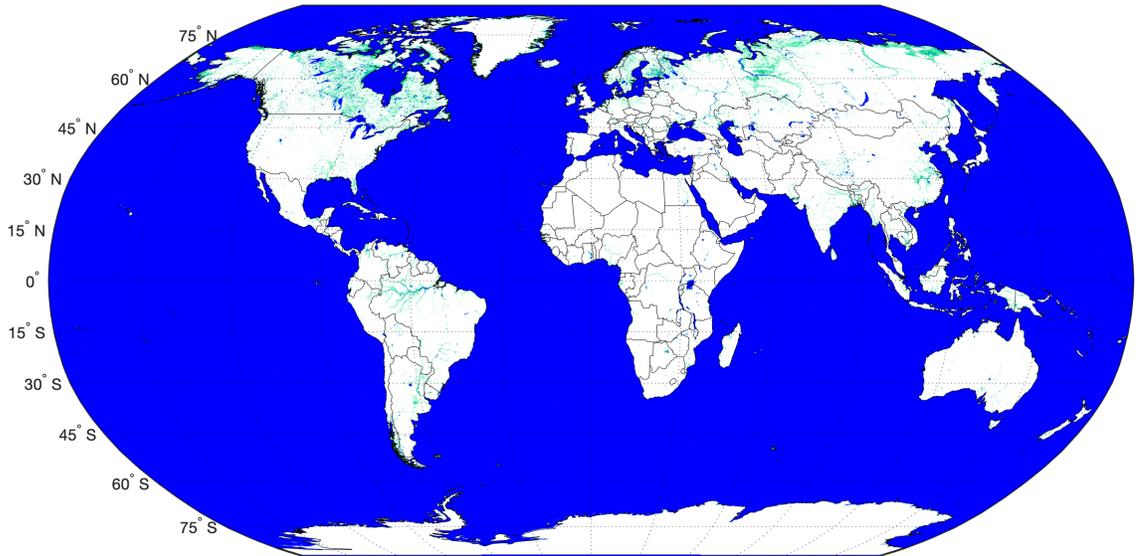


## Forest proportion



# Water proportion

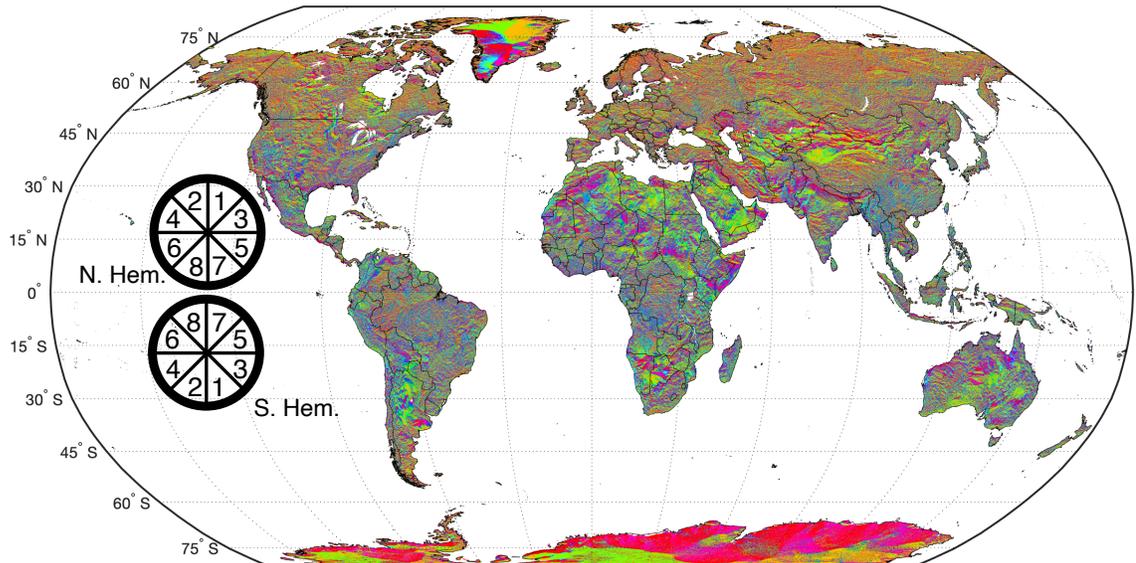
150° W 120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E



Proportion of surface water within each 0.05° grid

# Aspect 8-class

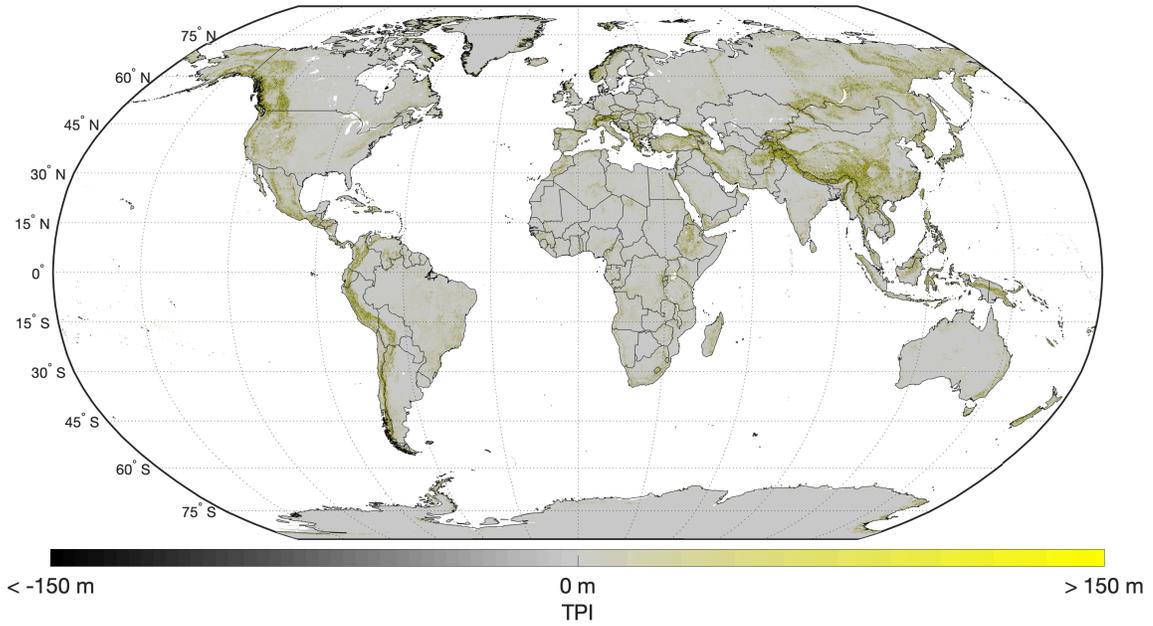
150° W 120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E



no aspect

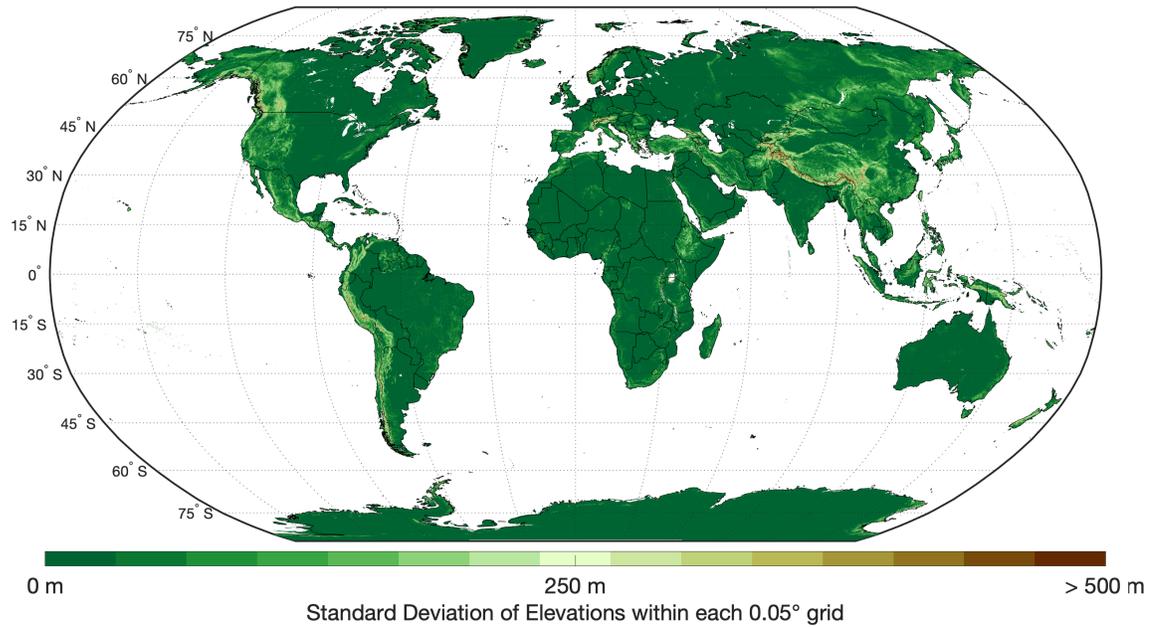
# Topographic Prominence Index (TPI)

150° W 120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E



# Elevation Standard Deviation

150° W 120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E



## REFERENCES

- Abraham, S., C. Huynh, and H. Vu. 2020. "Classification of Soils into Hydrologic Groups Using Machine Learning." *Data*, 5 (1): 2. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/data5010002>.
- Acevo-Herrera, R., A. Aguasca, X. Bosch-Lluis, A. Camps, J. Martínez-Fernández, N. Sánchez-Martín, and C. Pérez-Gutiérrez. 2010. "Design and First Results of an UAV-Borne L-Band Radiometer for Multiple Monitoring Purposes." *Remote Sensing*, 2 (7): 1662–1679. Molecular Diversity Preservation International. <https://doi.org/10.3390/rs2071662>.
- Ala-Aho, P., et al. 2021. "What conditions favor the influence of seasonally frozen ground on hydrological partitioning? A systematic review." *Environ. Res. Lett.*, 16 (4): 043008. IOP Publishing. <https://doi.org/10.1088/1748-9326/abe82c>.
- Amante, C., and B. W. Eakins. 2009. "ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum" NESDIS, NGDC-24, 19.
- Arndt, K. A., W. C. Oechel, J. P. Goodrich, B. A. Bailey, A. Kalthori, J. Hashemi, C. Sweeney, and D. Zona. 2019. "Sensitivity of Methane Emissions to Later Soil Freezing in Arctic Tundra Ecosystems." *Journal of Geophysical Research: Biogeosciences*, 124 (8): 2595–2609. <https://doi.org/10.1029/2019JG005242>.
- NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team. 2019. "ASTER Global Digital Elevation Model V003 [Data set]." NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/ASTER/ASTGTM.003>
- Bahel, V., S. Pillai, and M. Malhotra. 2020. "A Comparative Study on Various Binary Classification Algorithms and their Improved Variant for Optimal Performance." 2020 IEEE Region 10 Symposium (TENSYP), 495–498.
- Barkan, O., J. Weill, and A. Averbuch. 2016. "Gaussian Process Regression for Out-of-Sample Extension." <http://arxiv.org/abs/1603.02194>
- Batani, S., C. Huang, S. A. Margulis, E. Podest, and K. McDonald. 2013. "Feasibility of Characterizing Snowpack and the Freeze–Thaw State of Underlying Soil Using Multifrequency Active/Passive Microwave Data." *Geoscience and Remote Sensing, IEEE Transactions on*, 51: 4085–4102. <https://doi.org/10.1109/TGRS.2012.2229466>.

- Bayard, D., M. Stähli, A. Parriaux, and H. Flüeler. 2005. "The influence of seasonally frozen soil on the snowmelt runoff at two Alpine sites in southern Switzerland." *Journal of Hydrology*, 309 (1): 66–84. <https://doi.org/10.1016/j.jhydrol.2004.11.012>.
- Beck, H. E., N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood. 2018. "Present and future Köppen-Geiger climate classification maps at 1-km resolution." *Scientific Data*, 5 (1): 180214. Nature Publishing Group. <https://doi.org/10.1038/sdata.2018.214>.
- Belgiu, M., and L. Drăguț. 2016. "Random Forest in remote sensing: A review of applications and future directions." *ISPRS Journal of Photogrammetry and Remote Sensing*, 114: 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Bell, J. E., M. A. Palecki, C. B. Baker, W. G. Collins, J. H. Lawrimore, R. D. Leeper, M. E. Hall, J. Kochendorfer, T. P. Meyers, T. Wilson, and H. J. Diamond. 2013. "U.S. Climate Reference Network Soil Moisture and Temperature Observations." *Journal of Hydrometeorology*, 14 (3): 977–988. American Meteorological Society.
- Berni, J. A. J., P. J. Zarco-Tejada, L. Suarez, and E. Fereres. 2009. "Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring from an Unmanned Aerial Vehicle." *IEEE Transactions on Geoscience and Remote Sensing*, 47 (3): 722–738. <https://doi.org/10.1109/TGRS.2008.2010457>.
- Best, M. J., et al. 2011. "The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes." *Geoscientific Model Development*, 4 (3): 677–699. Copernicus GmbH. <https://doi.org/10.5194/gmd-4-677-2011>.
- Betts, A. K., R. Desjardins, D. Worth, S. Wang, and J. Li. 2014. "Coupling of winter climate transitions to snow and clouds over the Prairies: WINTER CLIMATE TRANSITIONS WITH SNOW." *Journal of Geophysical Research: Atmospheres*, 119 (3): 1118–1139. <https://doi.org/10.1002/2013JD021168>.
- Biskaborn, B. K., et al., 2019. "Permafrost is warming at a global scale." *Nature Communications*, 10 (1): 264. Nature Publishing Group. <https://doi.org/10.1038/s41467-018-08240-4>.
- Bohn, T. J., M. Y. Sonessa, and D. P. Lettenmaier. 2010. "Seasonal Hydrologic Forecasting: Do Multimodel Ensemble Averages Always Yield Improvements in Forecast Skill?" *J. Hydrometeorol.*, 11 (6): 1358–1372. American Meteorological Society. <https://doi.org/10.1175/2010JHM1267.1>.

- Breiman, L. 2001. "Random Forests." *Machine Learning*, 45 (1): 5–32.  
<https://doi.org/10.1023/A:1010933404324>.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. New York: Routledge.
- Brodzik, M.J., and Long, D.G. 2016. "MEaSURES Calibrated Enhanced-Resolution Passive Microwave Daily EASE-Grid 2.0 Brightness Temperature ESDR, Version 1." NASA National Snow and Ice Data Center DAAC.
- Brodzik, Mary J., Long, David G., and Hardman, Molly A. 2021. "SMAP Radiometer Twice-Daily rSIR-Enhanced EASE-Grid 2.0 Brightness Temperatures, Version 2." NASA National Snow and Ice Data Center DAAC.
- Brucker, L., E. P. Dinnat, and L. S. Koenig. 2014. "Weekly gridded Aquarius L-band radiometer/scatterometer observations and salinity retrievals over the polar regions – Part 2: Initial product analysis." *The Cryosphere*, 8 (3): 915–930. Copernicus GmbH. <https://doi.org/10.5194/tc-8-915-2014>.
- Campbell, R. B. 1951. "The Freezing Point of Water in Puddled and Unpuddled Soils at Different Soil Moisture Tension Values." 53.
- Castanedo, F. 2013. "A Review of Data Fusion Techniques." *The Scientific World Journal*, 2013: e704504. Hindawi. <https://doi.org/10.1155/2013/704504>.
- Chai, L., L. Zhang, Y. Zhang, Z. Hao, L. Jiang, and S. Zhao. 2014. "Comparison of the classification accuracy of three soil freeze–thaw discrimination algorithms in China using SSMIS and AMSR-E passive microwave imagery." *International Journal of Remote Sensing*, 35 (22): 7631–7649.  
<https://doi.org/10.1080/01431161.2014.975376>.
- Channan, S., K. Collins, and W. R. Emanuel. 2014. "Global mosaics of the standard MODIS land cover type data." University of Maryland and the Pacific Northwest National Laboratory.
- Charrier, G., J. Ngao, M. Saudreau, and T. Améglio. 2015. "Effects of environmental factors and management practices on microclimate, winter physiology, and frost resistance in trees." *Frontiers in Plant Science*, 6.
- Chen, L., K. Li, G. Song, D. Zhang, and C. Liu. 2021. "Effect of freeze–thaw cycle on physical and mechanical properties and damage characteristics of sandstone." *Sci Rep*, 11 (1): 12315. Nature Publishing Group. <https://doi.org/10.1038/s41598-021-91842-8>.

- Chen, X., L. Liu, and A. Bartsch. 2019. “Detecting soil freeze/thaw onsets in Alaska using SMAP and ASCAT data.” *Remote Sensing of Environment*, 220: 59–70. <https://doi.org/10.1016/j.rse.2018.10.010>.
- Chen, Z., Z. Zhu, H. Jiang, and S. Sun. 2020. “Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods.” *Journal of Hydrology*, 591: 125286. <https://doi.org/10.1016/j.jhydrol.2020.125286>.
- Colliander, A., K. McDonald, R. Zimmermann, T. Linke, R. Schroeder, J. Kimball, and E. Njoku. 2010. “Quikscat backscatter sensitivity to landscape freeze/thaw state over ALECTRA sites in Alaska from 2000 to 2007: Application to SMAP validation planning.” 2010 IEEE International Geoscience and Remote Sensing Symposium, 1269–1272.
- Cosgrove, B. et al. 2003. “Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project.” *Journal of Geophysical Research: Atmospheres*, 108 (D22). <https://doi.org/10.1029/2002JD003118>.
- Daniel, J. S., J. M. Jacobs, H. Miller, A. Stoner, J. Crowley, M. Khalkhali, and A. Thomas. 2017. “Climate change: potential impacts on frost–thaw conditions and seasonal load restriction timing for low-volume roadways.” *Road Materials and Pavement Design*, 0 (0): 1–21. <https://doi.org/10.1080/14680629.2017.1302355>.
- Danielson, J. J., and D. B. Gesch. 2011. “Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010).” 34.
- De Lannoy, G. J. M., and R. H. Reichle. 2016. “Global Assimilation of Multiangle and Multipolarization SMOS Brightness Temperature Observations into the GEOS-5 Catchment Land Surface Model for Soil Moisture Estimation.” *J. Hydrometeor.*, 17 (2): 669–691. American Meteorological Society. <https://doi.org/10.1175/JHM-D-15-0037.1>.
- Derber, J. C., D. F. Parrish, and S. J. Lord. 1991. “The New Global Operational Analysis System at the National Meteorological Center.” *Wea. Forecasting*, 6 (4): 538–547. American Meteorological Society. [https://doi.org/10.1175/1520-0434\(1991\)006<0538:TNGOAS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1991)006<0538:TNGOAS>2.0.CO;2).
- Derksen, C., X. Xu, R. Scott Dunbar, A. Colliander, Y. Kim, J. S. Kimball, T. A. Black, E. Euskirchen, A. Langlois, M. M. Lorant, P. Marsh, K. Rautiainen, A. Roy, A. Royer, and J. Stephens. 2017. “Retrieving landscape freeze/thaw state from Soil Moisture Active Passive (SMAP) radar and radiometer measurements.” *Remote Sensing of Environment*, 194 (Supplement C): 48–62. <https://doi.org/10.1016/j.rse.2017.03.007>.

- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki. 2006. "GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface." *Bull. Amer. Meteor. Soc.*, 87 (10): 1381–1398. American Meteorological Society. <https://doi.org/10.1175/BAMS-87-10-1381>.
- Dobson, M., F. Ulaby, M. Hallikainen, and M. El-rayes. 1985. "Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models." *IEEE Transactions on Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.1985.289498>.
- dos Santos, J. A., P.-H. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. X. Falao. 2012. "Multiscale Classification of Remote Sensing Images." *IEEE Transactions on Geoscience and Remote Sensing*, 50 (10): 3764–3775. <https://doi.org/10.1109/TGRS.2012.2186582>.
- Du, J., J. S. Kimball, M. Azarderakhsh, R. S. Dunbar, M. Moghaddam, and K. C. McDonald. 2015. "Classification of Alaska Spring Thaw Characteristics Using Satellite L-Band Radar Remote Sensing." *IEEE Transactions on Geoscience and Remote Sensing*, 53 (1): 542–556. <https://doi.org/10.1109/TGRS.2014.2325409>.
- Du, J., J. D. Watts, L. Jiang, H. Lu, X. Cheng, C. Duguay, M. Farina, Y. Qiu, Y. Kim, J. S. Kimball, and P. Tarolli. 2019. "Remote Sensing of Environmental Changes in Cold Regions: Methods, Achievements and Challenges." *Remote Sensing*, 11 (16): 1952. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/rs11161952>.
- Du, P. 2019. "Ensemble Machine Learning-Based Wind Forecasting to Combine NWP Output with Data from Weather Station." *IEEE Transactions on Sustainable Energy*, 10 (4): 2133–2141. <https://doi.org/10.1109/TSTE.2018.2880615>.
- Duan, S.-B., Z.-L. Li, H. Li, F.-M. Götsche, H. Wu, W. Zhao, P. Leng, X. Zhang, and C. Coll. 2019. "Validation of Collection 6 MODIS land surface temperature product using in situ measurements." *Remote Sensing of Environment*, 225: 16–29. <https://doi.org/10.1016/j.rse.2019.02.020>.
- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar. 2000. "A catchment-based approach to modeling land surface processes in a general circulation model: 2. Parameter estimation and model demonstration." *Journal of Geophysical Research: Atmospheres*, 105 (D20): 24823–24838. <https://doi.org/10.1029/2000JD900328>.
- Dunbar, S., X. Xu, A. Colliander, C. Derksen, J. S. Kimball, and Y. Kim. 2018. "SMAP - Algorithm Theoretical Basis Document (ATBD) - Level 3 Radiometer Freeze/Thaw Data Products." California Institute of Technology.

- Dunbar, S., Xu, X., Colliander, A., Derksen, C., Kimball, J., & Kim, Y. (2014-2020). SMAP Level 3 Radiometer Freeze/Thaw Data Products (L3\_FT\_P and L3\_FT\_P\_E). 38. Revisions A – C.
- Dunne, T., and R. D. Black. 1971. “Runoff Processes during Snowmelt.” *Water Resources Research*, 7 (5): 1160–1172.  
<https://doi.org/10.1029/WR007i005p01160>.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley. 2003. “Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model.” *Journal of Geophysical Research: Atmospheres*, 108 (D22).  
<https://doi.org/10.1029/2002JD003296>.
- England, A. W. 1990. “Radiobrightness of diurnally heated, freezing soil.” *IEEE Transactions on Geoscience and Remote Sensing*, 28 (4): 464–476.  
<https://doi.org/10.1109/tgrs.1990.572923>.
- England, A. W., J. F. Galantowicz, and B. W. Zuerndorfer. 1991. “A volume scattering explanation for the negative spectral gradient of frozen soil.” [Proceedings] *IGARSS’91 Remote Sensing: Global Monitoring for Earth Management*, 1175–1177.
- Espinoza, C. Z., L. R. Khot, S. Sankaran, and P. W. Jacoby. 2017. “High Resolution Multispectral and Thermal Remote Sensing-Based Water Stress Assessment in Subsurface Irrigated Grapevines.” *Remote Sensing*, 9 (9): 961.  
<https://doi.org/10.3390/rs9090961>.
- ETOPO1 “1 Arc-Minute Global Relief Model.” Accessed February 24, 2022.  
<https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.dem:316>.
- Farhadi, L., R. H. Reichle, G. J. M. De Lannoy, and J. S. Kimball. 2014. “Assimilation of Freeze–Thaw Observations into the NASA Catchment Land Surface Model.” *J. Hydrometeor.*, 16 (2): 730–743. <https://doi.org/10.1175/JHM-D-14-0065.1>.
- Freitas, S. C., I. F. Trigo, J. Macedo, C. Barroso, R. Silva, and R. Perdigão. 2013. “Land surface temperature from multiple geostationary satellites.” *International Journal of Remote Sensing*, 34 (9–10): 3051–3068.  
<https://doi.org/10.1080/01431161.2012.716925>.
- Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. 2010. “MODIS Collection 5 global land cover: Algorithm refinements

- and characterization of new datasets.” *Remote Sensing of Environment*, 114 (1): 168–182. <https://doi.org/10.1016/j.rse.2009.08.016>.
- Friedl, Mark, and Sulla-Menashe, Damien. 2015. “MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006.” NASA EOSDIS Land Processes DAAC.
- Gelaro, R., et al. 2017. “The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2).” *J. Climate*, 30 (14): 5419–5454. American Meteorological Society. <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gómez-Plaza, A., M. Martínez-Mena, J. Albaladejo, and V. M. Castillo. 2001. “Factors regulating spatial distribution of soil water content in small semiarid catchments.” *Journal of Hydrology*, 253 (1): 211–226. [https://doi.org/10.1016/S0022-1694\(01\)00483-8](https://doi.org/10.1016/S0022-1694(01)00483-8).
- Goulden, M. L., J. W. Munger, S.-M. Fan, B. C. Daube, and S. C. Wofsy. 1996. “Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy.” *Global Change Biology*, 2 (3): 169–182. <https://doi.org/10.1111/j.1365-2486.1996.tb00070.x>.
- Grody, N. C. 1991. “Classification of snow cover and precipitation using the special sensor microwave imager.” *Journal of Geophysical Research: Atmospheres*, 96 (D4): 7423–7435. <https://doi.org/10.1029/91JD00045>.
- Guo, Z., P. A. Dirmeyer, X. Gao, and M. Zhao. 2007. “Improving the quality of simulated soil moisture with a multi-model ensemble approach.” *Quarterly Journal of the Royal Meteorological Society*, 133 (624): 731–747. <https://doi.org/10.1002/qj.48>.
- Hall, D. K., and Riggs, G. A. 2021. “MODIS/Terra Snow Cover Daily L3 Global 0.05Deg CMG, Version 61.” NASA National Snow and Ice Data Center DAAC.
- Han, W., C. Huang, H. Duan, J. Gu, and J. Hou. 2020. “Lake Phenology of Freeze-Thaw Cycles Using Random Forest: A Case Study of Qinghai Lake.” *Remote Sensing*, 12 (24): 4098. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/rs12244098>.
- Harvey, M. C., J. V. Rowland, and K. M. Luketina. 2016. “Drone with thermal infrared camera provides high resolution georeferenced imagery of the Waikite geothermal area, New Zealand.” *Journal of Volcanology and Geothermal Research*, 325: 61–69. <https://doi.org/10.1016/j.jvolgeores.2016.06.014>.

- Henry, H. A. L. 2008. "Climate change and soil freezing dynamics: historical trends and projected changes." *Climatic Change*, 87 (3–4): 421–434. <https://doi.org/10.1007/s10584-007-9322-8>.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. "Very high-resolution interpolated climate surfaces for global land areas." *International Journal of Climatology*, 25 (15): 1965–1978. <https://doi.org/10.1002/joc.1276>.
- Hjort, J., O. Karjalainen, J. Aalto, S. Westermann, V. E. Romanovsky, F. E. Nelson, B. Etzelmüller, and M. Luoto. 2018. "Degrading permafrost puts Arctic infrastructure at risk by mid-century." *Nature Communications*, 9 (1): 5147. Nature Publishing Group. <https://doi.org/10.1038/s41467-018-07557-4>.
- Houser, P., J., Johnston. 2020. SnowEx In-situ Observations [Data set]. To be published
- Houtz, D., R. Naderpour, and M. Schwank. 2020. "Portable L-Band Radiometer (PoLRa): Design and Characterization." *Remote Sensing*, 12 (17): 2780. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/rs12172780>.
- Hu, T., T. Zhao, J. Shi, S. Wu, D. Liu, H. Qin, and K. Zhao. 2017. "High-Resolution Mapping of Freeze/Thaw Status in China via Fusion of MODIS and AMSR2 Data." *Remote Sensing*, 9 (12): 1339. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/rs9121339>.
- Hugelius, G., T. Virtanen, D. Kaverin, A. Pastukhov, F. Rivkin, S. Marchenko, V. Romanovsky, and P. Kuhry. 2011. "High-resolution mapping of ecosystem carbon storage and potential effects of permafrost thaw in periglacial terrain, European Russian Arctic." *Journal of Geophysical Research: Biogeosciences*, 116 (G3). <https://doi.org/10.1029/2010JG001606>.
- Iwata, Y., M. Nemoto, S. Hasegawa, Y. Yanai, K. Kuwao, and T. Hirota. 2011. "Influence of rain, air temperature, and snow cover on subsequent spring-snowmelt infiltration into thin frozen soil layer in northern Japan." *Journal of Hydrology*, 401 (3): 165–176. <https://doi.org/10.1016/j.jhydrol.2011.02.019>.
- Ji, L., G. B. Senay, and J. P. Verdin. 2015. "Evaluation of the Global Land Data Assimilation System (GLDAS) Air Temperature Data Products." *Journal of Hydrometeorology*, 16 (6): 2463–2480. <https://doi.org/10.1175/JHM-D-14-0230.1>.
- Jia, X., and J. A. Richards. 1999. "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification." *IEEE Transactions on Geoscience and Remote Sensing*, 37 (1): 538–542. <https://doi.org/10.1109/36.739109>.

- Jin, R., X. Li, and T. Che. 2009. "A decision tree algorithm for surface soil freeze/thaw classification over China using SSM/I brightness temperature." *Remote Sensing of Environment*, 113 (12): 2651–2660. <https://doi.org/10.1016/j.rse.2009.08.003>.
- Johnston, J., V. Maggioni, and P. Houser. 2019. "Investigating the Relationship Between Satellite-Based Freeze/Thaw Products and Land Surface Temperature." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1–25. <https://doi.org/10.1109/JSTARS.2019.2926942>.
- Johnston, J. M., P. R. Houser, V. Maggioni, R. S. Kim, and C. Vuyovich. 2021. "Informing Improvements in Freeze/Thaw State Classification Using Subpixel Temperature." *IEEE Transactions on Geoscience and Remote Sensing*, 1–19. <https://doi.org/10.1109/TGRS.2021.3099292>.
- Johnston, J., V. Maggioni, and P. Houser. 2020. "Comparing global passive microwave freeze/thaw records: Investigating differences between Ka- and L-band products." *Remote Sensing of Environment*, 247: 111936. <https://doi.org/10.1016/j.rse.2020.111936>.
- Judge, J., J. F. Galantowicz, A. W. England, and P. Dahl. 1997. "Freeze/thaw classification for prairie soils using SSM/I radiobrightnesses." *IEEE Transactions on Geoscience and Remote Sensing*, 35 (4): 827–832. <https://doi.org/10.1109/36.602525>.
- Kalra, A., T. Piechota, R. Davies, and G. Tootle. 2008. "Changes in U.S. Streamflow and Western U.S. Snowpack." *Journal of Hydrologic Engineering - J HYDROL ENG*, 13. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2008\)13:3\(156\)](https://doi.org/10.1061/(ASCE)1084-0699(2008)13:3(156)).
- Karunaratne, K. C., and C. R. Burn. 2003. "Freezing n-factors in discontinuous permafrost terrain, Takhini River, Yukon Territory, Canada." *Proceedings of the 8th International Conference on Permafrost*, 519–524. University of Zurich, Switzerland.
- Kawanishi, T., T. Sezai, Y. Ito, K. Imaoka, T. Takeshima, Y. Ishido, A. Shibata, M. Miura, H. Inahata, and R. W. Spencer. 2003. "The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies." *IEEE Transactions on Geoscience and Remote Sensing*, 41 (2): 184–194. <https://doi.org/10.1109/TGRS.2002.808331>.
- Kelly, J., N. Kljun, P.-O. Olsson, L. Mihai, B. Liljeblad, P. Weslien, L. Klemedtsson, and L. Eklundh. 2019. "Challenges and Best Practices for Deriving Temperature Data

- from an Uncalibrated UAV Thermal Infrared Camera.” *Remote Sensing*, 11 (5): 567. <https://doi.org/10.3390/rs11050567>.
- Kerr, Y. et al. 2012. “The SMOS Soil Moisture Retrieval Algorithm.” *IEEE Transactions on Geoscience and Remote Sensing*, 50 (5): 1384–1403. <https://doi.org/10.1109/TGRS.2012.2184548>.
- Kerr, Y. H., and E. G. Njoku. 1990. “A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space.” *IEEE Transactions on Geoscience and Remote Sensing*, 28 (3): 384–393. <https://doi.org/10.1109/36.54364>.
- Kim, E., et al. 2016. “SLAPex Freeze/Thaw 2015: The First Dedicated Soil Freeze/Thaw Airborne Campaign.” Beijing.
- Kim, R. S., S. Kumar, C. Vuyovich, P. Houser, J. Lundquist, L. Mudryk, M. Durand, A. Barros, E. J. Kim, B. A. Forman, E. D. Gutmann, M. L. Wrzesien, C. Garnaud, M. Sandells, H.-P. Marshall, N. Cristea, J. M. Pflug, J. Johnston, Y. Cao, D. Mocko, and S. Wang. 2020. “Snow Ensemble Uncertainty Project (SEUP): Quantification of snow water equivalent uncertainty across North America via ensemble land surface modeling.” *The Cryosphere Discussions*, 1–32. Copernicus GmbH. <https://doi.org/10.5194/tc-2020-248>.
- Kim, Y., J. S. Kimball, K. C. McDonald, and J. Glassy. 2011. “Developing a Global Data Record of Daily Landscape Freeze/Thaw Status Using Satellite Passive Microwave Remote Sensing.” *IEEE Transactions on Geoscience and Remote Sensing*, 49 (3): 949–960. <https://doi.org/10.1109/TGRS.2010.2070515>.
- Kim, Y., J. S. Kimball, K. Zhang, and K. C. McDonald. 2012. “Satellite detection of increasing Northern Hemisphere non-frozen seasons from 1979 to 2008: Implications for regional vegetation growth.” *Remote Sensing of Environment*, 121 (Supplement C): 472–487. <https://doi.org/10.1016/j.rse.2012.02.014>.
- Kim, Y., J. S. Kimball, K. Zhang, K. Didan, I. Velicogna, and K. C. McDonald. 2014. “Attribution of divergent northern vegetation growth responses to lengthening non-frozen seasons using satellite optical-NIR and microwave remote sensing.” *International Journal of Remote Sensing*, 35 (10): 3700–3721. Taylor & Francis. <https://doi.org/10.1080/01431161.2014.915595>.
- Kim, Y., J. S. Kimball, D. A. Robinson, and C. Derksen. 2015. “New satellite climate data records indicate strong coupling between recent frozen season changes and snow cover over high northern latitudes.” *Environ. Res. Lett.*, 10 (8): 084004. <https://doi.org/10.1088/1748-9326/10/8/084004>.

- Kim, Y., J. S. Kimball, J. Glassy, and K. C. McDonald. 2018. "MEaSURES Global Record of Daily Landscape Freeze/Thaw Status, Version 4." Boulder, Colorado: NASA National Snow and Ice Data Center Distributed Active Archive Center, 2017. <https://doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0477.004>.
- Kim, Y., J. S. Kimball, J. Du, and J. Glassy. 2021a. "MEaSURES Global Record of Daily Landscape Freeze/Thaw Status, Version 5." Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/10.5067/LJ6SLXNJB2CQ>.
- Kim, Y., J. S. Kimball, J. Du, T. Kundig, J. S. Kimball, J. Du, and J. Glassy. 2021b. "A Global Data Record of Daily Landscape Freeze/Thaw status Version 5.0." NSIDC.
- Kim, Y., J. S. Kimball, J. Glassy, and J. Du. 2017. "An extended global Earth system data record on daily landscape freeze–thaw status determined from satellite passive microwave remote sensing." *Earth System Science Data*, 9 (1): 133–147. <https://doi.org/10.5194/essd-9-133-2017>.
- Kim, Y., J. S. Kimball, X. Xu, R. S. Dunbar, A. Colliander, and C. Derksen. 2019. "Global Assessment of the SMAP Freeze/Thaw Data Record and Regional Applications for Detecting Spring Onset and Frost Events." *Remote Sensing*, 11 (11): 1317. <https://doi.org/10.3390/rs11111317>.
- Kimball, J. S., K. C. McDonald, Y. Kim, and J. Glassy. 2009. "An Earth System Data Record for Land Surface Freeze/Thaw State. Algorithm Theoretical Basis Document (ATBD), Version 1."
- Kimball, J. S., K. C. McDonald, S. W. Running, and S. E. Frolking. 2004. "Satellite radar remote sensing of seasonal growing seasons for boreal and subalpine evergreen forests." *Remote Sensing of Environment*, 90 (2): 243–258. <https://doi.org/10.1016/j.rse.2004.01.002>.
- Koren, V., J. Schaake, K. Mitchell, Q.-Y. Duan, F. Chen, and J. M. Baker. 1999. "A parameterization of snowpack and frozen ground intended for NCEP weather and climate models." *J. Geophys. Res.*, 104 (D16): 19569–19585. <https://doi.org/10.1029/1999JD900232>.
- Koren, V., M. Smith, and Z. Cui. 2014. "Physically-based modifications to the Sacramento Soil Moisture Accounting model. Part A: Modeling the effects of frozen ground on the runoff generation process." *Journal of Hydrology*, 519: 3475–3491. <https://doi.org/10.1016/j.jhydrol.2014.03.004>.

- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar. 2000. "A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure." *Journal of Geophysical Research: Atmospheres*, 105 (D20): 24809–24822. <https://doi.org/10.1029/2000JD900327>.
- Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel. 2006. "World Map of the Köppen-Geiger climate classification updated." *Meteorologische Zeitschrift*, 15 (3): 259–263. <https://doi.org/10.1127/0941-2948/2006/0130>.
- Kraatz, S., J. M. Jacobs, R. Schröder, E. Cho, M. Cosh, M. Seyfried, J. Prueger, and S. Livingston. 2018. "Evaluation of SMAP Freeze/Thaw Retrieval Accuracy at Core Validation Sites in the Contiguous United States." *Remote Sensing*, 10 (9): 1483. <https://doi.org/10.3390/rs10091483>.
- Kreyling, J., C. Beierkuhnlein, K. Pritsch, M. Schloter, and A. Jentsch. 2008. "Recurrent soil freeze-thaw cycles enhance grassland productivity." *New Phytol.*, 177 (4): 938–945. <https://doi.org/10.1111/j.1469-8137.2007.02309.x>.
- Kulanuwat, L., C. Chantrapornchai, M. Maleewong, P. Wongchaisuwat, S. Wimala, K. Sarinnapakorn, and S. Boonya-aroonnet. 2021. "Anomaly Detection Using a Sliding Window Technique and Data Imputation with Machine Learning for Hydrological Time Series." *Water*, 13 (13): 1862. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/w13131862>.
- La Cava, W., C. Bauer, J. H. Moore, and S. A. Pendergrass. 2020. "Interpretation of machine learning predictions for patient outcomes in electronic health records." *AMIA Annu Symp Proc*, 2019: 572–581.
- Lembrechts, J. J., et al. 2020. "SoilTemp: A global database of near-surface temperature." *Global Change Biology*, 26 (11): 6616–6629. <https://doi.org/10.1111/gcb.15123>.
- Li, C. 2013. "Probability Estimation in Random Forests." 35.
- Lin, Y.-C., and A. Habib. 2021. "Quality control and crop characterization framework for multi-temporal UAV LiDAR data over mechanized agricultural fields." *Remote Sensing of Environment*, 256: 112299. <https://doi.org/10.1016/j.rse.2021.112299>.
- Long, D. G., M. J. Brodzik, and M. A. Hardman. 2019. "Enhanced-Resolution SMAP Brightness Temperature Image Products." *IEEE Trans. Geosci. Remote Sensing*, 57 (7): 4151–4163. <https://doi.org/10.1109/TGRS.2018.2889427>.
- Luo, L., et al. 2018. "Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains." *Journal of*

Geophysical Research: Atmospheres.

[https://doi.org/10.1029/2002JD003246@10.1002/\(ISSN\)2169-8996.GCIP3](https://doi.org/10.1029/2002JD003246@10.1002/(ISSN)2169-8996.GCIP3).

- Lyu, H., K. A. McColl, X. Li, C. Derksen, A. Berg, T. A. Black, E. Euskirchen, M. Loranty, J. Pulliainen, K. Rautiainen, T. Rowlandson, A. Roy, A. Royer, A. Langlois, J. Stephens, H. Lu, and D. Entekhabi. 2018. “Validation of the SMAP freeze/thaw product using categorical triple collocation.” *Remote Sensing of Environment*, 205: 329–337. <https://doi.org/10.1016/j.rse.2017.12.007>.
- Manfreda, S., et al. 2018. “On the Use of Unmanned Aerial Systems for Environmental Monitoring.” *Remote Sensing*, 10 (4): 641. <https://doi.org/10.3390/rs10040641>.
- Martins, J. P. 2019. “The Hourly Land Surface Temperature from the Copernicus Global Land Service – Part 1: the updated algorithm with inclusion of vegetation dynamics and of Indian Ocean Data Coverage mission.”
- Martins, L., G. Vasconcelos, P. B. Lourenço, and C. Palha. 2016. “Influence of the Freeze-Thaw Cycles on the Physical and Mechanical Properties of Granites.” *Journal of Materials in Civil Engineering*, 28 (5): 04015201. American Society of Civil Engineers. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0001488](https://doi.org/10.1061/(ASCE)MT.1943-5533.0001488).
- Massey, F. J. 1951. “The Kolmogorov-Smirnov Test for Goodness of Fit.” *Journal of the American Statistical Association*, 46 (253): 68. <https://doi.org/10.2307/2280095>.
- Mavrovic, A., R. Pardo Lara, A. Berg, F. Demontoux, A. Royer, and A. Roy. 2021. “Soil dielectric characterization during freeze–thaw transitions using L-band coaxial and soil moisture probes.” *Hydrology and Earth System Sciences*, 25 (3): 1117–1131. Copernicus GmbH. <https://doi.org/10.5194/hess-25-1117-2021>.
- McNairn, H., et al. 2015. “The Soil Moisture Active Passive Validation Experiment 2012 (SMAPVEX12): Prelaunch Calibration and Validation of the SMAP Soil Moisture Algorithms.” *IEEE Transactions on Geoscience and Remote Sensing*, 53 (5): 2784–2801. <https://doi.org/10.1109/TGRS.2014.2364913>.
- Mei, Y., V. Maggioni, P. Houser, Y. Xue, and T. Rouf. 2020. “A Nonparametric Statistical Technique for Spatial Downscaling of Precipitation Over High Mountain Asia.” *Water Resources Research*, 56 (11): e2020WR027472. <https://doi.org/10.1029/2020WR027472>.
- Meier, W. N., and J. S. Stewart. 2020. “Assessing the Potential of Enhanced Resolution Gridded Passive Microwave Brightness Temperatures for Retrieval of Sea Ice Parameters.” *Remote Sensing*, 12 (16): 2552. <https://doi.org/10.3390/rs12162552>.

- Mesas-Carrascosa, F.-J., F. Pérez-Porras, J. E. Meroño de Larriva, C. Mena Frau, F. Agüera-Vega, F. Carvajal-Ramírez, P. Martínez-Carricondo, and A. García-Ferrer. 2018. “Drift Correction of Lightweight Microbolometer Thermal Sensors On-Board Unmanned Aerial Vehicles.” *Remote Sensing*, 10 (4): 615. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/rs10040615>.
- Mesinger, F., et al. 2006. “North American Regional Reanalysis.” *Bull. Amer. Meteor. Soc.*, 87 (3): 343–360. <https://doi.org/10.1175/BAMS-87-3-343>.
- Miller, L. H. 1956. “Table of Percentage Points of Kolmogorov Statistics.” *Journal of the American Statistical Association*, 51 (273): 111–121. <https://doi.org/10.1080/01621459.1956.10501314>.
- Molod, A., L. Takacs, M. Suarez, and J. Bacmeister. 2015. “Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2.” *Geoscientific Model Development*, 8 (5): 1339–1356. Copernicus GmbH. <https://doi.org/10.5194/gmd-8-1339-2015>.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaigis. 1996. “The ECMWF Ensemble Prediction System: Methodology and validation.” *Quarterly Journal of the Royal Meteorological Society*, 122 (529): 73–119. <https://doi.org/10.1002/qj.49712252905>.
- Monsell, B. C. 2007. “The X-13A-S Seasonal Adjustment Program.” U.S Census Bureau.
- Mortin, J., T. M. Schröder, A. W. Hansen, B. Holt, and K. C. McDonald. 2012. “Mapping of seasonal freeze-thaw transitions across the pan-Arctic land and sea ice domains with satellite radar.” *Journal of Geophysical Research: Oceans*, 117 (C8). <https://doi.org/10.1029/2012JC008001>.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth. 2004. “Quantification of modelling uncertainties in a large ensemble of climate change simulations.” *Nature*, 430 (7001): 768–772. Nature Publishing Group. <https://doi.org/10.1038/nature02771>.
- Muzalevskiy, K., and Z. Ruzicka. 2020. “Detection of soil freeze/thaw states in the Arctic region based on combined SMAP and AMSR-2 radio brightness observations.” *International Journal of Remote Sensing*, 41 (14): 5046–5061. Taylor & Francis. <https://doi.org/10.1080/01431161.2020.1724348>.
- National Centers for Environmental Information (NCEI). 2019. Climate at a Glance. <https://www.ncdc.noaa.gov/cag/divisional/mapping/110/tavg/201512/1/anomaly>

- Nelson, F. E., and S. I. Outcalt. 1987. "A Computational Method for Prediction and Regionalization of Permafrost." *Arctic and Alpine Research*, 19 (3): 279–288. <https://doi.org/10.1080/00040851.1987.12002602>.
- Ni, J., Y. Cheng, Q. Wang, C. W. W. Ng, and A. Garg. 2019. "Effects of vegetation on soil temperature and water content: Field monitoring and numerical modelling." *Journal of Hydrology*, 571: 494–502. <https://doi.org/10.1016/j.jhydrol.2019.02.009>.
- Niu, G.-Y., and Z.-L. Yang. 2006. "Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale." *J. Hydrometeorol.*, 7 (5): 937–952. American Meteorological Society. <https://doi.org/10.1175/JHM538.1>.
- Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia. 2011. "The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements." *Journal of Geophysical Research: Atmospheres*, 116 (D12). <https://doi.org/10.1029/2010JD015139>.
- Owe, M., and A. A. Van de Griend. 1998. "Comparison of soil moisture penetration depths for several bare soils at two microwave frequencies and implications for remote sensing." *Water Resources Research*, 34 (9): 2319–2327. <https://doi.org/10.1029/98WR01469>.
- Pacheco et al., 2019. Real-Time In-Situ Soil Monitoring for Agriculture (RISMA) Network Metadata [Data set]. Historical Data, <https://agriculture.canada.ca/SoilMonitoringStations/historical-data-en.html>
- Painter, T. H., et al. 2016. "The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo." *Remote Sensing of Environment*, 184: 139–152. <https://doi.org/10.1016/j.rse.2016.06.018>.
- Pal, M. 2005. "Random forest classifier for remote sensing classification." *International Journal of Remote Sensing*, 26 (1): 217–222. Taylor & Francis. <https://doi.org/10.1080/01431160412331269698>.
- Pappu, V., and P. M. Pardalos. 2014. "High-Dimensional Data Classification." *Clusters, Orders, and Trees: Methods and Applications*, Springer Optimization and Its Applications, F. Aleskerov, B. Goldengorin, and P. M. Pardalos, eds., 119–150. New York, NY: Springer New York.
- Parkin, T. B. 1993. "Spatial Variability of Microbial Processes in Soil—A Review."

- Podest, E., K. C. McDonald, and J. S. Kimball. 2014. "Multisensor Microwave Sensitivity to Freeze/Thaw Dynamics Across a Complex Boreal Landscape." *IEEE Transactions on Geoscience and Remote Sensing*, 52 (11): 6818–6828. <https://doi.org/10.1109/TGRS.2014.2303635>.
- Prince, M., A. Roy, A. Royer, and A. Langlois. 2019. "Timing and spatial variability of fall soil freezing in boreal forest and its effect on SMAP L-band radiometer measurements." *Remote Sensing of Environment*, 231: 111230. <https://doi.org/10.1016/j.rse.2019.111230>.
- Rasmussen, C. E. 2004. "Gaussian Processes in Machine Learning." *Advanced Lectures on Machine Learning, Lecture Notes in Computer Science*, O. Bousquet, U. von Luxburg, and G. Rätsch, eds., 63–71. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Rautiainen, K., J. Lemmetyinen, J. Pulliainen, J. Vehvilainen, M. Drusch, A. Kontu, J. Kainulainen, and J. Seppanen. 2012. "L-Band Radiometer Observations of Soil Processes in Boreal and Subarctic Environments." *IEEE Transactions on Geoscience and Remote Sensing*, 50 (5): 1483–1497. <https://doi.org/10.1109/TGRS.2011.2167755>.
- Rautiainen, K., J. Lemmetyinen, M. Schwank, A. Kontu, C. B. Ménard, C. Mätzler, M. Drusch, A. Wiesmann, J. Ikonen, and J. Pulliainen. 2014. "Detection of soil freezing from L-band passive microwave observations." *Remote Sensing of Environment*, 147: 206–218. <https://doi.org/10.1016/j.rse.2014.03.007>.
- Rautiainen, K., T. Parkkinen, J. Lemmetyinen, M. Schwank, A. Wiesmann, J. Ikonen, C. Derksen, S. Davydov, A. Davydova, J. Boike, M. Langer, M. Drusch, and J. Pulliainen. 2016. "SMOS prototype algorithm for detecting autumn soil freezing." *Remote Sensing of Environment, Special Issue: ESA's Soil Moisture and Ocean Salinity Mission - Achievements and Applications*, 180: 346–360. <https://doi.org/10.1016/j.rse.2016.01.012>.
- Reda, I., and A. Andreas. 2008. *Solar Position Algorithm for Solar Radiation Applications (Revised)*. NREL/TP-560-34302, 15003974.
- Reichle, R. H., S. V. Kumar, S. P. P. Mahanama, R. D. Koster, and Q. Liu. 2010. "Assimilation of Satellite-Derived Skin Temperature Observations into Land Surface Models." *J. Hydrometeor.*, 11 (5): 1103–1122. <https://doi.org/10.1175/2010JHM1262.1>.
- Resop, J. P., L. Lehmann, and W. C. Hession. 2019. "Drone Laser Scanning for Modeling Riverscape Topography and Vegetation: Comparison with Traditional

- Aerial Lidar.” *Drones*, 3 (2): 35. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/drones3020035>.
- Rignot, E., and J. B. Way. 1994. “Monitoring freeze—thaw cycles along North—South Alaskan transects using ERS-1 SAR.” *Remote Sensing of Environment*, 49 (2): 131–137. [https://doi.org/10.1016/0034-4257\(94\)90049-3](https://doi.org/10.1016/0034-4257(94)90049-3).
- Robinson, D. A., and T. W. Estilow. 2021. “Rutgers Northern Hemisphere 24 km Weekly Snow Cover Extent, September 1980 Onward, Version 1.” NSIDC.
- Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll. 2004. “The Global Land Data Assimilation System.” *Bull. Amer. Meteor. Soc.*, 85 (3): 381–394. <https://doi.org/10.1175/BAMS-85-3-381>.
- Rowlandson, T. L., A. A. Berg, A. Roy, E. Kim, R. Pardo Lara, J. Powers, K. Lewis, P. Houser, K. McDonald, P. Toose, A. Wu, E. De Marco, C. Derksen, J. Entin, A. Colliander, X. Xu, and A. Mavrovic. 2018. “Capturing agricultural soil freeze/thaw state through remote sensing and ground observations: A soil freeze/thaw validation campaign.” *Remote Sensing of Environment*, 211: 59–70. <https://doi.org/10.1016/j.rse.2018.04.003>.
- Roy, A., A. Royer, C. Derksen, L. Brucker, A. Langlois, A. Mialon, and Y. H. Kerr. 2015. “Evaluation of Spaceborne L-Band Radiometer Measurements for Terrestrial Freeze/Thaw Retrievals in Canada.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (9): 4442–4459. <https://doi.org/10.1109/JSTARS.2015.2476358>.
- Roy, A., P. Toose, M. Williamson, T. Rowlandson, C. Derksen, A. Royer, A. A. Berg, J. Lemmetyinen, and L. Arnold. 2017a. “Response of L-Band brightness temperatures to freeze/thaw and snow dynamics in a prairie environment from ground-based radiometer measurements.” *Remote Sensing of Environment*, 191 (Supplement C): 67–80. <https://doi.org/10.1016/j.rse.2017.01.017>.
- Roy, A., et al. 2017b. “Spatial Variability of L-Band Brightness Temperature during Freeze/Thaw Events over a Prairie Environment.” *Remote Sensing*, 9 (9): 894. <https://doi.org/10.3390/rs9090894>.
- Roy, A., P. Toose, A. Mavrovic, C. Pappas, A. Royer, C. Derksen, A. Berg, T. Rowlandson, M. El-Amine, A. Barr, A. Black, A. Langlois, and O. Sonnentag. 2020. “L-Band response to freeze/thaw in a boreal forest stand from ground- and tower-based radiometer observations.” *Remote Sensing of Environment*, 237: 111542. <https://doi.org/10.1016/j.rse.2019.111542>.

- Royer, A., K. Goïta, J. Kohn, and D. D. Sève. 2010. "Monitoring Dry, Wet, and No-Snow Conditions from Microwave Satellite Observations." *IEEE Geoscience and Remote Sensing Letters*, 7 (4): 670–674.  
<https://doi.org/10.1109/LGRS.2010.2045733>.
- Sagan, V., et al. 2019. "UAV-Based High Resolution Thermal Imaging for Vegetation Monitoring, and Plant Phenotyping Using ICI 8640 P, FLIR Vue Pro R 640, and thermoMap Cameras." *Remote Sensing*, 11 (3): 330.  
<https://doi.org/10.3390/rs11030330>.
- Sasaki, Y. 2007. "The truth of the F-measure." 5.
- Schwanghart, W. 2021. TopoToolbox - a set of Matlab functions for topographic analysis. <https://github.com/wschwanghart/topotoolbox>
- Sen, P. K. 2005. "Gini diversity index, hamming distance and curse of dimensionality." *Metron - International Journal of Statistics*.
- Shanley, J. B., and A. Chalmers. 1999. "The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont." *Hydrological Processes*, 13 (12-13): 1843–1857.  
[https://doi.org/10.1002/\(SICI\)1099-1085\(199909\)13:12/13<1843::AID-HYP879>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1843::AID-HYP879>3.0.CO;2-G).
- Sheng, H., H. Chao, C. Coopmans, J. Han, M. McKee, and Y. Chen. 2010. "Low-cost UAV-based thermal infrared remote sensing: Platform, calibration and applications." *Proceedings of 2010 IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications*, 38–43.
- Simpson, E. H. 1949. "Measurement of Diversity." *Nature*, 163 (4148): 688–688. Nature Publishing Group. <https://doi.org/10.1038/163688a0>.
- Song, Y., and Y. Lu. 2015. "Decision tree methods: applications for classification and prediction." *Shanghai Arch Psychiatry*, 27 (2): 130–135.  
<https://doi.org/10.11919/j.issn.1002-0829.215044>.
- Stadler, D., H. Flühler, and P.-E. Jansson. 1997. "Modelling vertical and lateral water flow in frozen and sloped forest soil plots." *Cold Regions Science and Technology*, 26 (3): 181–194. [https://doi.org/10.1016/S0165-232X\(97\)00017-7](https://doi.org/10.1016/S0165-232X(97)00017-7).
- Suarez, M. J., A. daSilva, D. Dee, S. Bloom, M. Bosilovich, S. Pawson, S. Schubert, M.-L. Wu, M. Sienkiewicz, and I. Stajner. 2005. Documentation and Validation of the Goddard Earth Observing System (GEOS) Data Assimilation System, Version 4.

- Sugiura, R., N. Noguchi, and K. Ishii. 2007. "Correction of Low-altitude Thermal Images applied to estimating Soil Water Status." *Biosystems Engineering*, 96 (3): 301–313. <https://doi.org/10.1016/j.biosystemseng.2006.11.006>.
- Sulla-Menashe, D., and M. A. Friedl. 2018. "User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product."
- Suzuki, S. 2004. "Verification of freezing point depression method for measuring matric potential of soil water." *Soil Science and Plant Nutrition*, 50 (8): 1277–1280. <https://doi.org/10.1080/00380768.2004.10408604>.
- Teufel, B., G. T. Diro, K. Whan, S. M. Milrad, D. I. Jeong, A. Ganji, O. Huziy, K. Winger, J. R. Gyakum, R. de Elia, F. W. Zwiers, and L. Sushama. 2017. "Investigation of the 2013 Alberta flood from weather and climate perspectives." *Clim Dyn*, 48 (9–10): 2881–2899. <https://doi.org/10.1007/s00382-016-3239-8>.
- Turner, D., A. Lucieer, Z. Malenovský, D. H. King, and S. A. Robinson. 2014. "Spatial Co-Registration of Ultra-High Resolution Visible, Multispectral and Thermal Images Acquired with a Micro-UAV over Antarctic Moss Beds." *Remote Sensing*, 6 (5): 4003–4024. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/rs6054003>.
- Ulaby, F. T., R. K. Moore, and A. K. Fung. 1986. *Microwave remote sensing: Active and passive. Volume 1-3 - Microwave remote sensing fundamentals and radiometry.*
- Ulaby, F. T. 1981. *Microwave remote sensing: active and passive. Remote sensing: no. 2-[4].* Reading, Mass: Addison-Wesley Pub. Company, Advanced Book Program/World Science Division.
- USDA Natural Resources Conservation Service. 2020a. *Soil Climate Analysis Network (SCAN) [Data set].* Ag Data Commons. Accessed 2021-03-16, <https://data.nal.usda.gov/dataset/soil-climate-analysis-network-scan>
- USDA Natural Resources Conservation Service. 2020b. *SNOWpack TELEmetry Network (SNOTEL) [Data set].* Ag Data Commons. Accessed 2021-03-16, <https://data.nal.usda.gov/dataset/snowpack-telemetry-network-snotel>
- USGS. 2017. "Grand Mesa, Uncompahgre, and Gunnison National Forests, Draft Forest Assessments: Terrestrial Ecosystems: Integrity and System Drivers and Stressors." 142.

- Wan, Zhengming, Hook, Simon, and Hulley, Glynn. 2015. "MOD11C1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 0.05Deg CMG V006." NASA EOSDIS Land Processes DAAC.
- Wan, Zhengming, Hook, Simon, and Hulley, Glynn. 2021. "MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 0.05Deg CMG V061." NASA EOSDIS Land Processes DAAC.
- Wan, Z. 1999. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD). 77. Institute for Computational Earth System Science.
- Wang, L., P. Toose, R. Brown, and C. Derksen. 2016. "Frequency and distribution of winter melt events from passive microwave satellite data in the pan-Arctic, 1988–2013." *The Cryosphere*, 14.
- Wang, J., L. Jiang, H. Cui, G. Wang, J. Yang, X. Liu, and X. Su. 2020a. "Evaluation and analysis of SMAP, AMSR2 and MEaSURES freeze/thaw products in China." *Remote Sensing of Environment*, 242: 111734. <https://doi.org/10.1016/j.rse.2020.111734>.
- Wang, S., X. Luo, and Y. Peng. 2020b. "Spatial Downscaling of MODIS Land Surface Temperature Based on Geographically Weighted Autoregressive Model." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 2532–2546. <https://doi.org/10.1109/JSTARS.2020.2968809>.
- Whitehead, K., C. H. Hugenholtz, S. Myshak, O. Brown, A. LeClair, A. Tamminga, T. E. Barchyn, B. Moorman, and B. Eaton. 2014. "Remote sensing of the environment with small unmanned aircraft systems (UASs), part 2: scientific and commercial applications." *J. Unmanned Vehicle Sys.*, 02 (03): 86–102. <https://doi.org/10.1139/juvs-2014-0007>.
- Xia, Y., et al., NCEP/EMC 2009. NLDAS Primary Forcing Data L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2018, 10.5067/6J5LHHOHZHN4
- Xia, Y., et al., NCEP/EMC 2012a. NLDAS Noah Land Surface Model L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2018, 10.5067/47Z13FNQODKV
- Xia, Y., M. Ek, J. Sheffield, B. Livneh, M. Huang, H. Wei, S. Feng, L. Luo, J. Meng, and E. Wood. 2012b. "Validation of Noah-Simulated Soil Temperature in the North

- American Land Data Assimilation System Phase 2.” *J. Appl. Meteor. Climatol.*, 52 (2): 455–471. <https://doi.org/10.1175/JAMC-D-12-033.1>.
- Xia, Y., et al. 2012. “Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products.” *Journal of Geophysical Research: Atmospheres*, 117 (D3). <https://doi.org/10.1029/2011JD016048>.
- Xu, S., Q. Zhao, K. Yin, G. He, Z. Zhang, G. Wang, M. Wen, and N. Zhang. 2021. “Spatial Downscaling of Land Surface Temperature Based on a Multi-Factor Geographically Weighted Machine Learning Model.” *Remote Sensing*, 13 (6): 1186. Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/rs13061186>.
- Xu, X., R. S. Dunbar, C. Derksen, A. Colliander, J. Kimball, and Y. Kim. 2016. “Landscape freeze/thaw products from Soil Moisture Active/Passive (SMAP) radar and radiometer data.” 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 132–135.
- Xu, X., C. Derksen, R. S. Dunbar, A. Colliander, J. Kimball, and Y. Kim. 2017. “Landscape freeze/thaw standard and enhanced products from soil moisture active/passive (SMAP) radiometer data.” 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2550–2553.
- Xu, X., R. S. Dunbar, A. Colliander, Y. Kim, and J. Kimball. 2018. “Soil Moisture Active Passive (SMAP) Project Calibration and Validation for the L3\_FT\_P and L3\_FT\_P\_E Data Product (Version 2).” 47.
- Xue, Y., P. R. Houser, V. Maggioni, Y. Mei, S. V. Kumar, and Y. Yoon. 2019. “Assimilation of Satellite-Based Snow Cover and Freeze/Thaw Observations Over High Mountain Asia.” *Front. Earth Sci.*, 7. <https://doi.org/10.3389/feart.2019.00115>.
- Yan, G., Z.-H. Jiao, T. Wang, and X. Mu. 2020. “Modeling surface longwave radiation over high-relief terrain.” *Remote Sensing of Environment*, 237: 111556. <https://doi.org/10.1016/j.rse.2019.111556>.
- Yang, K. 2021. The multiscale observation network of soil temperature and moisture on the central Tibetan Plateau (2010-2021). National Tibetan Plateau Data Center, DOI: 10.11888/Terre.tpdc.271918.
- Yang, Z.-L., G.-Y. Niu, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, L. Longuevergne, K. Manning, D. Niyogi, M. Tewari, and Y. Xia. 2011. “The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation

- over global river basins.” *Journal of Geophysical Research: Atmospheres*, 116 (D12). <https://doi.org/10.1029/2010JD015140>.
- Zhang, K., J. S. Kimball, Y. Kim, and K. C. McDonald. 2011. “Changing freeze-thaw seasons in northern high latitudes and associated influences on evapotranspiration.” *Hydrological Processes*, 25 (26): 4142–4151. <https://doi.org/10.1002/hyp.8350>.
- Zhang, L., T. Zhao, L. Jiang, and S. Zhao. 2010. “Estimate of Phase Transition Water Content in Freeze–Thaw Process Using Microwave Radiometer.” *IEEE Transactions on Geoscience and Remote Sensing*, 48 (12): 4248–4255. <https://doi.org/10.1109/TGRS.2010.2051158>.
- Zhang, T., and R. L. Armstrong. 2001. “Soil freeze/thaw cycles over snow-free land detected by passive microwave remote sensing.” *Geophysical Research Letters*, 28. <https://doi.org/10.1029/2000GL011952>.
- Zhang, T., R. L. Armstrong, and J. Smith. 2003. “Investigation of the near-surface soil freeze-thaw cycle in the contiguous United States: Algorithm development and validation.” *Journal of Geophysical Research: Atmospheres*, 108 (D22). <https://doi.org/10.1029/2003JD003530>.
- Zhang, X., and S. Sun. 2011. “The impact of soil freezing/thawing processes on water and energy balances.” *Adv. Atmos. Sci.*, 28 (1): 169–177. <https://doi.org/10.1007/s00376-010-9206-0>.
- Zhao, T., J. Shi, T. Hu, L. Zhao, D. Zou, T. Wang, D. Ji, R. Li, and P. Wang. 2017. “Estimation of high-resolution near-surface freeze/thaw state by the integration of microwave and thermal infrared remote sensing data on the Tibetan Plateau.” *Earth and Space Science*, 4 (8): 472–484. <https://doi.org/10.1002/2017EA000277>.
- Zhao, T., L. Zhang, L. Jiang, S. Zhao, L. Chai, and R. Jin. 2011. “A new soil freeze/thaw discriminant algorithm using AMSR-E passive microwave imagery.” *Hydrol. Process.*, 25 (11): 1704–1716. <https://doi.org/10.1002/hyp.7930>.
- Zheng, D., X. Li, T. Zhao, J. Wen, R. Van Der Velde, M. Schwank, X. Wang, Z. Wang, and Z. Su. 2020. “Impact of Soil Permittivity and Temperature Profile on L-Band Microwave Emission of Frozen Soil.” *IEEE Transactions on Geoscience and Remote Sensing*, 1–14. <https://doi.org/10.1109/TGRS.2020.3024971>.
- Zheng, G., Y. Yang, D. Yang, B. Dafflon, Y. Yi, S. Zhang, D. Chen, B. Gao, T. Wang, R. Shi, and Q. Wu. 2020. “Remote sensing spatiotemporal patterns of frozen soil and the environmental controls over the Tibetan Plateau during 2002–2016.” *Remote Sensing of Environment*, 247: 111927. <https://doi.org/10.1016/j.rse.2020.111927>.

Zhong, W., Q. Yuan, T. Liu, and L. Yue. 2022. "Freeze/thaw onset detection combining SMAP and ASCAT data over Alaska: A machine learning approach." *Journal of Hydrology*, 605: 127354. <https://doi.org/10.1016/j.jhydrol.2021.127354>.

Zuerndorfer, B. W., A. W. England, M. C. Dobson, and F. T. Ulaby. 1990. "Mapping freeze/thaw boundaries with SMMR data." *Agric. For. Meteorology.*, 52: 199–225.

Zwieback, S., A. Bartsch, T. Melzer, and W. Wagner. 2012. "Probabilistic Fusion of Ku - and C-band Scatterometer Data for Determining the Freeze/Thaw State." *IEEE Transactions on Geoscience and Remote Sensing*, 50 (7): 2583–2594. <https://doi.org/10.1109/TGRS.2011.2169076>.

## BIOGRAPHY

Jeremy M. Johnston graduated from Albemarle High School in Charlottesville, Virginia, in 2013. He then went on to receive both his Bachelor of Science and Master of Science Degrees from George Mason University in 2017 and 2019, respectively. His research interests include global hydrology, weather and climate, satellite remote sensing, ground-based field data collection and sensors, as well as the applications of UAV technology for remote sensing. Jeremy's goal is to help extend human understanding through observation of macro- and micro-scale processes and subsequently, the application of this knowledge to developing solutions to global environmental challenges. He enjoys spending time outdoors, playing sports, watching movies, and learning just about anything.

He has participated in extensive field activities (2017 – 2020) in the mountains of Colorado to support NASA SnowEx efforts and provide maintenance to ground-based sensing networks. Additionally, he attended a CUAHSI snow measurement field school in Bozeman Montana in January 2019 to support this work. Also in 2019, he traveled to Australia to support a soil moisture field measurement campaign in coordination with Monash University. At George Mason University he has received awards from the Graduate Student Travel Fund to support travel to conferences, as well as from the George Mason Open Access Publishing Fund to publish his scientific research. He has also served as a reviewer for multiple leading journals such as Water Resources Research, the Journal of Hydrology, and GIScience & Remote Sensing.