Full Polarimetric GNSS-R Assessment of the Freeze and Thaw States of the Terrestrial Cryosphere

Nereida Rodriguez-Alvarez, *Senior Member, IEEE*, Joan Francesc Munoz-Martin, *Member, IEEE*, Xavier Bosch-Lluis, *Senior Member, IEEE*, and Kamal Oudrhiri

Abstract— The shift between frozen and thawed conditions on the Earth's surface influences climate, hydrology, and ecology. A primary objective of the Soil Moisture Active Passive (SMAP) mission is to estimate the surface binary freeze/thaw (F/T) state for the northern hemisphere above 45° N latitude with a classification accuracy of 80% at 3-km spatial resolution and 2-day intervals. The objective was to be accomplished by the SMAP L-band radar measurements. In July 2015, the radar transmitter suffered an anomaly that prevented it from nominal operation. The mission partially met the goal exceeding the targeted 80% accuracy but at 36 km resolution. After the radar anomaly, the SMAP mission switched the radar receiver bandpass filter frequency to Global Positioning System (GPS) L2c, acting as a full-polarimetric Global Navigation Satellite System–Reflectometry (GNSS-R) receiver, known as SMAP-Reflectometer (SMAP-R). This work focuses on using SMAP-R signals to classify the F/T state over the Northern latitudes, within the limitations of the dataset. An algorithm based on the seasonal threshold approach that was originally envisioned for the SMAP radar is applied to the SMAP-R data (i.e., bistatic radar). Then the algorithm is evolved using a Random Forest algorithm to aid threshold selection from the discriminator built in the seasonal threshold approach. This algorithm is applied for years 2016 to 2022 over the Northern Hemisphere terrestrial cryosphere and shows an F/T classification accuracy agreement better than 97% with respect to the classification of the official SMAP F/T Radiometry product, proving the potential of polarimetric GNSS-R to derive F/T.

Index Terms— Cryosphere, freeze/thaw (F/T), Global Navigation Satellite System–Reflectometry (GNSS-R), polarimetry, Stokes parameters.

I. INTRODUCTION

THE terrestrial cryosphere refers to areas of the Earth's surface land that are permanently or seasonally frozen.

The observation and the analysis of the Earth's surface freeze/thaw (F/T) state, as well as the timing of the transition between those states, are key to understanding the impact of these processes on climate, hydrological, ecological, and biogeochemical processes. The Soil Moisture Active Passive (SMAP) mission's radar was designed, based on an extensive heritage from the ground, airborne, and satellite-based SAR missions [1], [2], [3], [4] to provide the most accurate remote sensing-based characterization of the Earth's surface F/T state at a global scale. The SMAP mission requirement was set to provide "estimates of surface binary F/T state for the region north of 45° N latitude, which includes the boreal forest zone, with a mean spatial classification accuracy of 80% at 3 km spatial resolution and 2-day average intervals" [5]. When the SMAP radar failed, the SMAP mission implemented a F/T product based on the measurements of the L-band radiometer, which provided a reduced spatial resolution with respect to the original plan of 36 and 9 km [6], and a different algorithm performance, as explained in [5]. Derksen et al. [7] explore the use of SMAP mission data to develop a high-resolution product for identifying the F/T state of landscapes. The study in [7] utilizes in situ observations to examine the sensitivity differences between 3-km radar and 36-km radiometer measurements to landscape F/T conditions. It demonstrates excellent agreement with in situ measurements, fulfilling the SMAP mission's accuracy requirements. The research in [7] highlights the challenges and nuances of detecting F/T transitions, particularly in the context of wet snow conditions, and discusses the potential for optimizing detection methodologies and validating results against other satellite data. The study in [7] provides valuable insight into the capabilities and limitations of current satellite-based methods for monitoring F/T states, which is critical for understanding hydrological and biospheric processes in northern landscapes and high elevations.

In addition to the advancements provided by the SMAP mission, the Soil Moisture and Ocean Salinity (SMOS) mission has also played a critical role in enhancing our understanding of the Earth's F/T cycles, particularly through the development of F/T products. The SMOS mission, utilizing its L-band passive microwave radiometry, has enabled the detection of soil freezing and thawing processes with unprecedented accuracy. Among publications, [8] have pioneered the use of SMOS data to develop prototype algorithms for detecting autumn soil freezing, showcasing the potential of SMOS in contributing to the global monitoring of seasonal F/T cycles.

These developments are crucial for improving the accuracy of hydrological models and enhancing predictions of seasonal and annual changes in soil moisture content, which are vital for agricultural planning, climate modeling, and understanding the dynamics of carbon exchange between the Earth's surface and the atmosphere. Other advancements in the monitoring of landscape F/T status through various remote sensing techniques contribute significantly to our understanding of climate change impacts, hydrological cycles, and environmental processes in both high-latitude and high-altitude regions. For instance, [9] and [10] have developed and extended global Earth system data records on the daily landscape freeze-thaw status, determined from satellite passive microwave remote sensing, emphasizing the methodology and potential applications in climate change effects and hydrological cycle analysis. Similarly, [11] introduced algorithms for detecting soil freezing and freeze-thaw states over high-latitude regions, utilizing Global Navigation Satellite System–Reflectometry (GNSS-R) data, highlighting the significance of these technologies in enhancing our understanding of seasonal freeze-thaw cycles. Moreover, [12] and [13] have focused on the discrimination and estimation of soil F/T states using AMSR-E passive microwave imagery and the integration of microwave and thermal infrared remote sensing data, particularly on the Tibetan Plateau, demonstrating improved monitoring capabilities in complex terrains.

The repurposing of the SMAP radar into a GNSS-R receiver enabled the first full polarimetric GNSS-R instrument in space, with more than eight years of data, starting in September 2015, a few months after the radar transmitter failed to conduct normal operations. While few investigations were conducted two years after the collection started [14], [15], it was not until seven years later when the dataset was mathematically formulated through the Stokes parameter full polarimetric scattering representation in [16] and then properly calibrated in [17]. The Stokes parameter formulation and the proper calibration enabled sensitivity analysis for both the land [18] and the cryosphere [19]. The study in [19] showed the sensitivity of the calibrated SMAP-R Stokes parameters over the cryosphere for applications such as sea ice detection and monitoring, Greenland ice sheet characterization, and F/T state detection and state transition monitoring. In this manuscript, we investigate the feasibility of an algorithm to produce F/T maps from SMAP-R data. Rather than just looking at signal sensitivities as it was done in [19], we now build a SMAP-R methodology based on the algorithm that was originally envisioned for the SMAP radar in [5], but using SMAP-R data alone. With this work, we aim to provide the first polarimetric forward-scattering characterization of the F/T state of surfaces above 45° N latitude at 36 km and 1-month temporal scale as SMAP-R is the first full polarimetric GNSS-R instrument in space and its measurements offer insight on the capability of such instruments. While the results are not meeting original requirements for the SMAP mission it does provide an independent assessment of how well SMAP-R data are sensitive to F/T states by comparing the retrieved values with respect to the validated SMAP F/T official product. Alternatively,

if properly calibrated toward in situ data rather than the SMAP F/T official product itself, SMAP-R could be used in conjunction with current SMAP F/T Radiometer products. This study also serves as justification of the capabilities of future polarimetric GNSS-R missions in development that will also tackle cryosphere science investigations with a better revisit and reduced sparsity, enabling products at 1-week scales. These upcoming polarimetric GNSS-R missions carry dual-polarization GNSS-R instruments and include the upcoming European Space Agency (ESA) HydroGNSS mission [20], constellations of satellites developed by private companies, such as Muon Space or Spire, as well as the China's first commercial GNSS-R instrument Tianjin Yunyao Aerospace Technology Company Ltd. (YUNYAO) [21].

As shown in [22] and [23], both the selection of the receiving antenna polarization and its gain are key for the sensitivity of the measurements to the Earth's geophysical parameters. In Munoz-Martin et al. [22] it was shown how the detection probability varies over the terrestrial cryosphere for a receiving antenna of 10 dB gain and a receiving antenna of 14 dB gain. For more standard GNSS-R instrument designs, antennas are selected to be RHCP/LHCP. If that is the case, the RHCP channel would benefit in general from a larger coherent integration time (∼4 ms). As shown in [22], even using higher coherent integration times and a 14 dB antenna, the measurements over the terrestrial cryosphere, especially over Alaska, Russia, and China, would experience challenges on the RHCP channel.

The manuscript is structured as Section II goes through the theoretical background for the full Stokes parameter GNSS-R formulation and the derivation of the reflectivity; Section III delves into the description of the algorithm for F/T classification based on a seasonal discriminator, starting with the methodology planned for the radar and evolving into the use of random forest to set the seasonal thresholds of different areas with similar electromagnetic features; Section IV presents F/T classification maps obtained from full polarimetric GNSS-R measurements as well as the performance respect to the SMAP F/T official product. Section V provides conclusions.

II. THEORETICAL BACKGROUND

The SMAP radar receiver, acting as a full polarimetric GNSS reflectometer, collects forward scattered signals reflected in the Earth's surface and transmitted by the Global Positioning System (GPS) constellation, using two receiving chains, one at vertical (V) and the other at horizontal (H) polarization. In the case of GPS, the signal is transmitted in right-hand circular polarization (RHCP) that, when observed by a pair of orthogonally polarized antennas, sets up what is known in the radar field as Hybrid Compact Polarimetry (HCP). Raney et al. [24] discuss the use of compact and quad-pol terminologies in the right-hand circular polarization (SAR) field as compared to the original definition of the full polarimetric system from radiometry. Furthermore, the author concludes that in response to a balanced illumination, such as circularly polarized transmissions, the resulting polarimetric

portraits embrace all of the polarimetric scattering information unique to the observed scene, and the system satisfies the original sense of full polarimetric. To measure such polarimetric portraits, the Stokes parameters of the signal are computed. For the specific case of GNSS-R, this can be achieved by receiving the H- and V-pol components of the RHCP transmitted signal and processing them following [16]. Thanks to the opportunistic nature of SMAP-R, born from a radar design, its receiver measures the GPS RHCP signals from a pair of linear orthogonal antennas, enabling full polarimetric GNSS-R.

A. SMAP-R Stokes Parameters

For the sake of providing a complete mathematical background, we summarize SMAP-R's main equations formulated in previous publications, such as [16]. The SMAP-R Stokes parameters of the reflected signal are defined in the following equation:

$$
\begin{aligned}
S_0 &= <|E_{RH}|^2> + <|E_{RV}|^2> = <|E_{RR}|^2> + <|E_{RL}|^2> \\
S_1 &= <|E_{RH}|^2> - <|E_{RV}|^2> \\
S_2 &= 2 <\mathrm{Re}\{E_{RH}E_{RV}^*\}> = <|E_{R+45°}|^2> - <|E_{R-45°}|^2> \\
S_3 &= 2 <\mathrm{Im}\{E_{RH}E_{RV}^*\}> = <|E_{RR}|^2> - <|E_{RL}|^2>
\end{aligned}
\tag{1}
$$

where E_{RH} is the electromagnetic field transmitted by the GPS satellite at RHCP and received at the SMAP H polarization antenna, and E_{RV} is the electromagnetic field transmitted by the GPS satellite at RHCP and received at the SMAP V polarization antenna. Consequently, and for the sake of illustration, $E_{R+45°/R-45°}$ refers to the same transmitted signal received by a $\pm45°$-polarized antenna, and $E_{RR/L}$ to a signal received by two circularly polarized antennas. Usually, these values are provided as normalized by the total signal intensity, S_0, i.e.: $\bar{S}_1 = S_1/S_0$, $\bar{S}_2 = S_2/S_0$, and $\bar{S}_3 = S_3/S_0$. Images of the normalized Stokes parameters over the cryosphere were extensively shown in [19]. In Fig. 1 we provide the normalized Stokes parameters mapped over the terrestrial cryosphere for 2019, as an illustrative example.

Areas such as Greenland or the ocean are filtered with the International Geosphere Biosphere Program (IGBP) dataset, [25].

B. SMAP-R Reflectivity

The reflectivity over land is computed from the calibrated bistatic radar cross section (BRCS, σ), obtained through means of the Stokes parameters as

$$
P_g = \frac{S_0 - S_{0noise}}{G}
\tag{2}
$$

being P_g the received power of the first Stokes parameter, S_{0noise} the noise level of S_0 at 30 delays away from the specular point position, and G the system gain. This absolute power calibration is covered in detail in [16, Sec. IV A]. This is converted to σ applying

$$
\sigma = \frac{P_g(4\pi)^3 R_T^2 R_R^2 L}{P_T \lambda^2 G_T G_R}
\tag{3}
$$

where R_T is the transmitter to specular point distance, R_R is the specular point to receiver distance, L are additional system losses, including atmospheric loss, P_T is the GPS transmit power, G_T is the transmitter antenna gain, and G_R is the SMAP receiving gain. Finally, the reflectivity over land can be computed as in 3 from [18]

$$
\Gamma_0 = \frac{1}{4\pi} \frac{(R_T^2 + R_R^2)}{R_T^2 R_R^2} \sigma.
\tag{4}
$$

Images of the reflectivity Γ_0 over the cryosphere were extensively shown in [19]. In Fig. 2, we provide Γ_0 mapped over the terrestrial cryosphere for 2019, as an illustrative example.

III. F/T CLASSIFICATION ALGORITHM AND PERFORMANCE ANALYSIS

Following what was envisioned for the SMAP radar data, we have developed an equivalent algorithm using SMAP-R data alone. From the different approaches described in [5], the seasonal threshold approach, the baseline one, is selected. This algorithm considers the remote sensing signatures acquired during seasonal reference frozen and thawed states, in this case, the SMAP-R Γ_0 signatures. Analogous to what is described in [5], a seasonal discriminator Δ is built to identify the F/T state at a particular given time, in the following equation:

$$
\Delta(t) = \frac{\Gamma_0(t) - \Gamma_{0fr}}{\Gamma_{0th} - \Gamma_{0fr}}
\tag{5}
$$

where $\Gamma_0(t)$ is the reflectivity measured at time t, for which the F/T state is the classification to be assessed, Γ_{0th} is the reflectivity corresponding to the thawed state, obtained as the Γ_0 mean for the periods when the landscapes are thawed, and Γ_{0fr} is the reflectivity corresponding to the frozen state, obtained as the Γ_0 mean for the periods when the landscapes are frozen.

A. Manual Selection of the F/T Discriminator Thresholds

For SMAP-R Γ_{0th} and Γ_{0fr} states are generated from averaging the Γ_0 measured during July to September and January to March for years 2016 to 2022, i.e., 7 years of data. The periods ensure pixels of constant state (frozen or thawed). Fig. 3 shows the F/T states used as reference for each grid cell.

A threshold T is then defined such as if $\Delta(t) > T$ the surface is thawed and if $\Delta(t) \leq T$ the surface is frozen. This T is not necessarily a fixed number and can be selected based on several conditions. One way to do this is to select a T value depending on the landscape type. Another way is to analyze Γ_0 over the terrestrial cryosphere, understand the levels of those signatures and classify per differences in Γ_0 at the F/T states. Indeed, the Γ_0 levels and the change observed between F/T states is a descriptor of the landscape itself. For example, wetland areas will observe similar Γ_0 levels and F/T state-related changes on those levels across the globe. Forested areas are more complex due to different tree structures and soil composition, which play a significant role in the observed Γ_0 levels and F/T state-related changes on those levels. On this line, we have developed a land classification that is based on
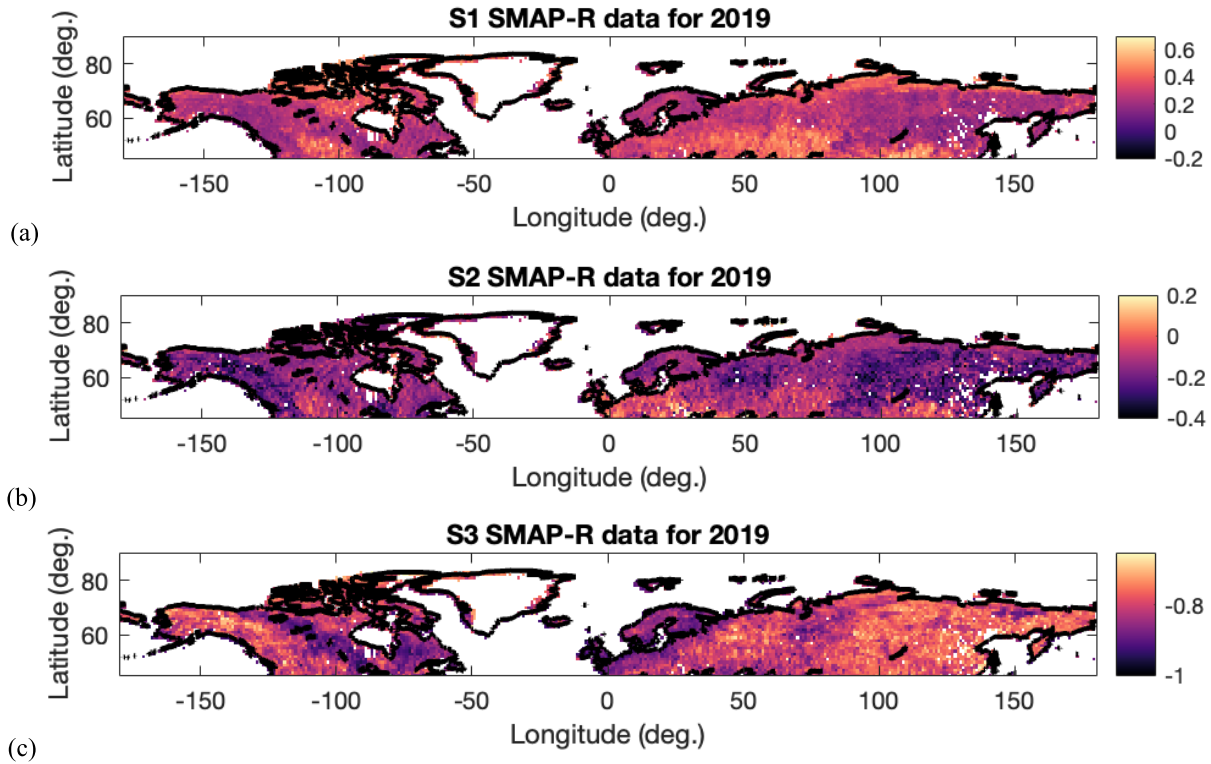
(a)

(b)

(c)

Fig. 1.   SMAP-R normalized Stokes parameters for the terrestrial cryosphere at latitudes over 45° N latitude: (a) normalized second Stokes ($\bar{S}_1$); (b) normalized third Stokes ($\bar{S}_2$); and (c) normalized fourth Stokes ($\bar{S}_3$). Data averaged for the entire year 2019, every cell contains an average of five measurements.
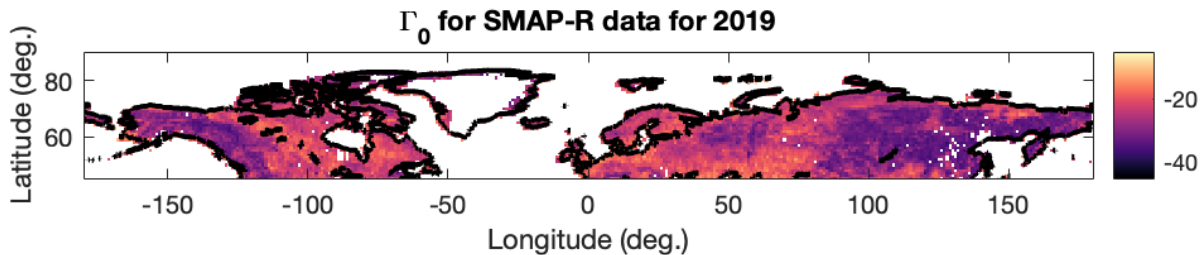

Fig. 2.   View of the SMAP-R reflectivity Γ_0 units in [dB] for the terrestrial cryosphere at latitudes over 45° N latitude. Data averaged for the entire year 2019, every cell contains an average of five measurements.

the Γ_0 levels and the change observed between F/T states. The classification is obtained using a k-means clustering [26] algorithm based on grouping samples by their mean as a representation of the properties of the pixels within the scene. K-means clustering, also known as Lloyd's algorithm, is an iterative, data-partitioning algorithm that assigns the vector of observations to the number of indicated clusters, with each cluster defined by a centroid. The result is shown in Fig. 4.

Following the land classification in Fig. 4, we analyze the time series of the mean value of the discriminator Δ (5) over those seven depicted clusters, which will ultimately allow us to set threshold T values for each cluster. Fig. 5 shows the discriminator in 5, averaged over each cluster, as a function of time. Note that to achieve the final number of clusters, we progressively increased the number of clusters until we found a limit in clearly showing a distinction between frozen and thawed periods for the clusters. e.g., we took pixels in cluster #2 and generated an extra cluster for this cluster. If it

provided a good distinction in any of the two, we added the cluster, if not, we discarded the ability for those pixels to be classified.

Seasonality can be seen in most of the grouped clusters, allowing to select a proper T value to describe the F/T states through the discriminator $\Delta(t)$ in (5). We have selected the values that mark the two states, i.e., T values are handpicked by selecting a value that ensures most January, February, and March monthly means for all years stay below the line, i.e., frozen. T values are represented in Fig. 5 through a dashed gray line and summarized in Table I, saved as a threshold Look-Up-Table (LUT) for the algorithm.

The seasonality in cluster #2 and cluster #6 is not as intense as in the rest of the clusters. Fig. 6 shows the map of pixels in clusters 2 and 6, which do not have a clear seasonality that allows for a discrimination value.

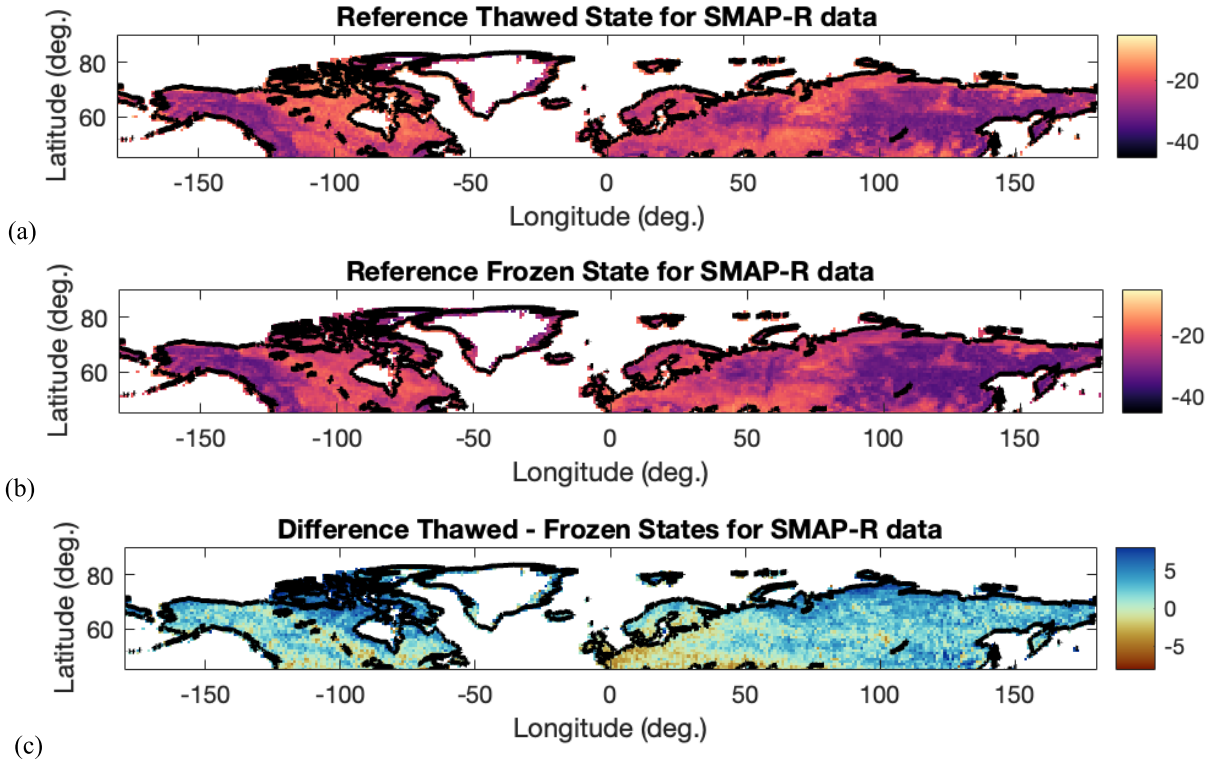As a summary, the diagram in Fig. 7 describes the implemented algorithm.

(a)

(b)

(c)

Fig. 3.   Reference Γ_0 over 45° N latitude for (a) thawed state, Γ_{0th}; (b) frozen state, Γ_{0fr}, (c) difference between thawed and frozen states. Data were averaged over periods of three months (July to September and January to March, respectively) for seven years of data, which provides a mean of eight measurements for each grid cell, providing values of frozen and thawed states for each grid cell.
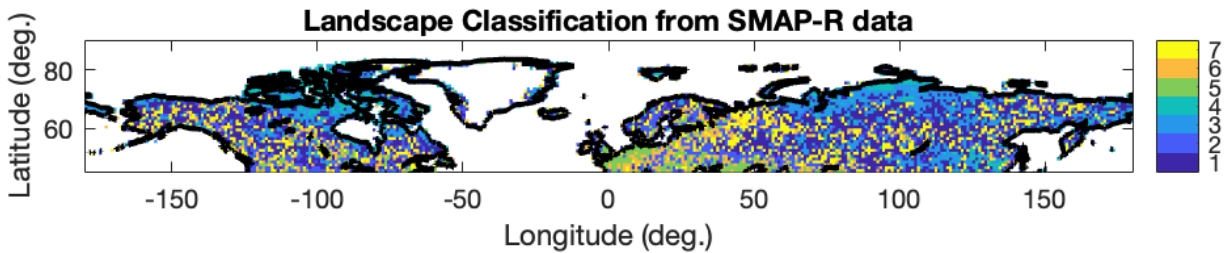

Fig. 4.   Land classification obtained for the terrestrial cryosphere at latitudes over 45° N latitude. This is computed by applying a k-means clustering algorithm to the discriminator in (5) using SMAP-R data.

TABLE I
THRESHOLD LUT: T VALUES FOUND TO DESCRIBE THE SEASONALITY OF DIFFERENT CLUSTERS

Cluster	T values
1	0.4
2	None
3	0.35
4	0.2
5	0.3
6	None
7	0.5

Next, we compute the performance of the algorithm presented in Fig. 7, through confusion matrix analysis.

B. Performance of the Algorithm Based on Manual Selection of the F/T Discriminator Thresholds

The performance is assessed for each year computing the corresponding confusion matrices. A confusion matrix is a tabular representation used in the field of machine learning and statistics to evaluate the performance of a classification algorithm. It is a tool that helps understand the accuracy and effectiveness of a model's predictions by comparing its predictions to the actual or target values. For each confusion matrix, we show:

1) True Thawed (TT): This represents the number of cases where the model correctly predicted the thawed class, target = 0 and predicted = 0.
2) True Frozen (TF): This represents the number of cases where the model correctly predicted the frozen class, target = 1 and predicted = 1.
3) False Thawed (FT): This represents the number of cases where the model incorrectly predicted the thawed class, target = 0 and predicted = 1.
4) False Frozen (FF): This represents the number of cases where the model incorrectly predicted the frozen class, target = 1 and predicted = 0.

where in our analysis, Target = F/T SMAP official product classification at 36 km [27], Predicted = F/T
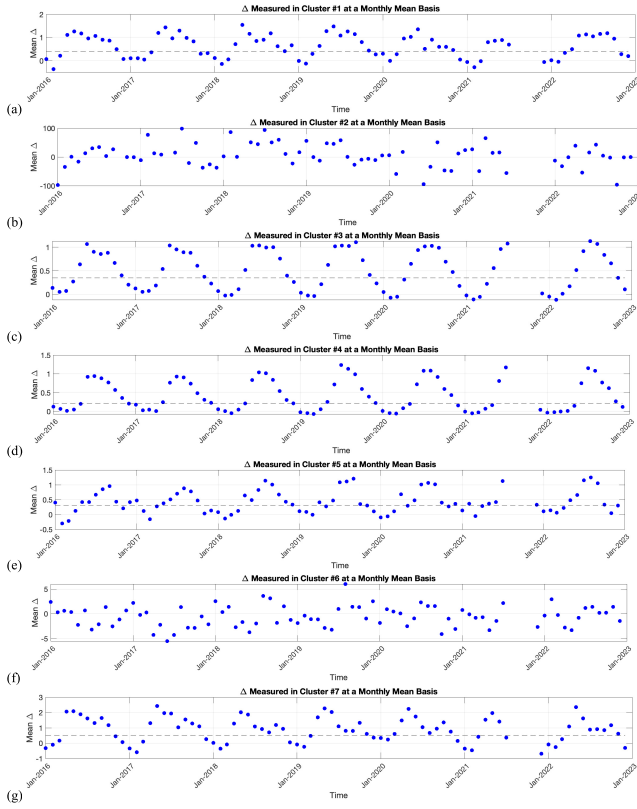
Fig. 5. Monthly mean discriminator Δ value as a function of time computed at the different clusters: (a) Cluster #1. (b) Cluster #2. (c) Cluster #3. (d) Cluster #4. (e) Cluster #5. (f) Cluster #6 and (g) Cluster #7. Each point corresponds to a monthly mean, resulting from adding cells within one cluster, ~ 1800 pixels corresponding to the same cluster each month.
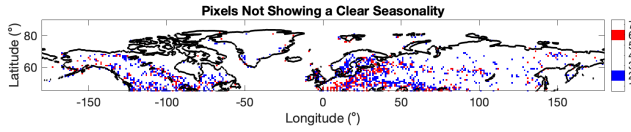

Fig. 6. View of the SMAP-R reflectivity Γ_0 for k-means cluster #2 and #6, for the terrestrial cryosphere at latitudes over 45° N latitude.

SMAP-R derived product at 36 km. From the [TT, TF, FT, FF] values in each confusion matrix, we can then compute:

5) the accuracy, which measures the overall correctness of the classification model, as

$$\text{Accuracy} = 100*\frac{\text{TT} + \text{TF}}{(\text{TT} + \text{TF} + \text{FT} + \text{FF})}[\%] \quad (6)$$

6) the precision, which quantifies the ability of the model to make accurate thawed predictions, as

$$\text{Precision} = 100*\frac{\text{TT}}{(\text{TT} + \text{FT})}[\%] \quad (7)$$

7) the sensitivity, which measures the model's ability to correctly identify all relevant instances of the thawed class, as

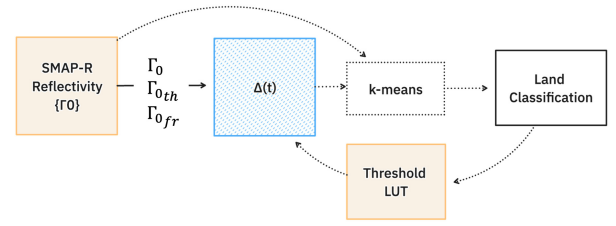$$\text{Sensitivity} = 100*\frac{\text{TT}}{(\text{TT} + \text{FF})}[\%] \quad (8)$$

Fig. 7. Flow diagram summarizing the implemented F/T Classification based on $\Delta(t)$.

8) the specificity, which measures the ability of the model to correctly identify frozen instances, as

$$\text{Specificity} = 100*\frac{\text{TF}}{(\text{TF} + \text{FT})}[\%]. \quad (9)$$

Table II and Table III show the performance results for the algorithm in Fig. 7, for years 2016 to 2018.

The complexity of the landscape makes the simple clustering methodology applied not complete to understand the F/T conditions of some pixels, as shown through the poor performance in Table II and Table III. Additionally, the threshold value for some clusters in Table I could not be manually found. For this reason, we implement a random forest decision algorithm, that ingests the discriminator itself together with full polarimetric information contained in the SMAP-R dataset.

C. F/T Classification Based on a Random Forest Decision Algorithm

An expanded methodology is proposed where we implement a machine learning approach that takes as inputs the full polarimetric information of SMAP-R, along with the reflectivity Γ_0, the discriminator in (5), and the land classification built from k-means clustering of the discriminator as shown in Fig. 4. The output target is obtained from the current official validated product from the SMAP radiometer. We select data from the seven years from frozen (January to March) and thaw (July to September) periods. For each period, we select pixels that remain constant in the same state (either frozen or thawed) for the duration of that period. We collocate SMAP-R measurements, i.e., full polarimetric GNSS-R data, during the same periods to obtain all input measurements belonging to each state. Then we use random forest methodology to build a decision tree, using the official SMAP radiometer F/T product as our target product. This is depicted in the diagram in Fig. 8. The SMAP official product is therefore used as targeted values, and its only use in the input specification is related to pixel collocation—to select pixels of guaranteed constant state that will then be used for training the random forest. Similar to what was done in [28] we used a randomization of multiple decision trees to construct the classification algorithm. This algorithm is based on a machine learning approach known as random forest [29], which builds multiple decision trees based on different combinations of input variables. The population of training data is 142 729 of a total of 438 835 available samples: i.e., 32.52% of the samples randomly selected over a few years of data are used for training, only the first four years,

TABLE II

Set of Confusion Matrices for Years 2016, 2017, and 2018. Each Confusion Matrix Shows the % of Classified Values for Each Target Category. Target = F/T SMAP Official Product Classification, Predicted = F/T SMAP-R Derived Product

2016	Predicted = 0 (thawed)	Predicted = 1 (frozen)	2017	Predicted = 0 (thawed)	Predicted = 1 (frozen)	2018	Predicted = 0 (thawed)	Predicted = 1 (frozen)
Target = 0 (thawed)	41.36%	58.64%	Target = 0 (thawed)	41.89%	58.11%	Target = 0 (thawed)	42.86%	57.14%
Target = 1 (frozen)	51.91%	48.09%	Target = 1 (frozen)	51.35%	48.65%	Target = 1 (frozen)	52.32%	47.68%

TABLE III

Accuracy, Precision, Sensitivity, and Specificity Obtained for Years 2016, 2017, and 2018

Year	Accuracy	Precision	Sensitivity	Specificity
2016	44.47%	41.36%	48.12%	41.34%
2017	45.05%	41.89%	48.30%	42.24%
2018	45.16%	42.86%	47.23%	43.29%

TABLE IV

Training/Validation and Testing Distribution

Year	Testing Population	Random Distribution of Training/Validation Population
2016	69961	22772 (32.55%)
2017	70049	45103 (64.39%)
2018	68873	42235 (61.32%)
2019	61422	32619 (53.11%)
2020	67411	0 (0%)
2021	40678	0 (0%)
2022	60441	0 (0%)
Total	438835	142729 (32.52%)

leaving the last three for testing. This sample includes only constantly frozen and constantly thawed pixels over periods of 1 month. The F/T state distribution for the used 32.52% of the total samples is:

1) Population of constant 0 is 77 995 out 142 729, i.e., 54.65%.
2) Population of constant 1 is 64 734 out 142 729, i.e., 45.35%.

The random forest uses 20 randomizations to build forest trees using 20 different combinations of observables from all available. For each one then, we use 70% of the available data for training the random forest and 30% for validation of that random forest. The 70% is selected 20 times randomly. This generates 400 forest trees.

Then data from every year is entered into those 400 random forest trees and the final class is selected using mode function in MATLAB that provides the more frequent value in a sample.

Population of testing data includes 438 835 samples, distributed as shown in Table IV.

Samples are unevenly distributed over the years because we chose to pick them randomly.

D. Performance of the Random Forest Decision Algorithm

The performance is assessed following the methodology in Section III-B, through computation of the confusion matrices, where the F/T classification from the SMAP radiometer product is used as target (reference true), i.e., Target = F/T

TABLE V

Set of Confusion Matrices for Each Year. Showing the % of Classified Values for Each Target Category. Target = F/T SMAP Official Product Classification, Predicted = F/T SMAP-R Derived Product

2016	Predicted 0 (thawed)	Predicted 1 (frozen)	2017	Predicted 0 (thawed)	Predicted 1 (frozen)
Target 0 (thawed)	99.07%	0.93%	Target 0 (thawed)	99.1%	0.90%
Target 1 (frozen)	5.42%	94.58%	Target 1 (frozen)	5.27%	94.73%
2018	Predicted 0 (thawed)	Predicted 1 (frozen)	**2019**	Predicted 0 (thawed)	Predicted 1 (frozen)
Target 0 (thawed)	99.11%	0.89%	Target 0 (thawed)	99.19%	0.81%
Target 1 (frozen)	5.54%	94.46%	Target 1 (frozen)	5.39%	94.61%
2020	Predicted 0 (thawed)	Predicted 1 (frozen)	**2021**	Predicted 0 (thawed)	Predicted 1 (frozen)
Target 0 (thawed)	99.21%	0.79%	Target 0 (thawed)	98.89%	1.11%
Target 1 (frozen)	5.50%	94.50%	Target 1 (frozen)	4.58%	95.42%
2022	Predicted 0 (thawed)	Predicted 1 (frozen)			
Target 0 (thawed)	99.08%	0.92%			
Target 1 (frozen)	4.86%	95. 14%			

TABLE VI

Yearly Values of Accuracy, Precision, Sensitivity, and Specificity Obtained for the SMAP-R F/T Product

Year	Accuracy	Precision	Sensitivity	Specificity
2016	97.15%	99.07%	96.08%	98.71%
2017	97.22%	99.10%	96.15%	98.75%
2018	97.04%	99.11%	95.69%	98.85%
2019	96.97%	99.19%	95.13%	99.10%
2020	97.22%	99.21%	96.11%	98.87%
2021	97.27%	98.89%	96.08%	98.69%
2022	97.08%	99.08%	95.15%	99.08%

SMAP official product classification at 36 km and Predicted = F/T SMAP-R derived product at 36 km. Also, the accuracy, precision, sensitivity, and specificity metric for the F/T classification algorithm is calculated following Section III-C. The resulting performance is shown in Table V and Table VI.

The overall performance of the algorithm proposed shows a mean accuracy of 97.13%, a mean precision of 99.09%, a mean sensitivity of 95.77%, and a mean specificity of 98.86%. This accuracy is related to the SMAP F/T official product, i.e., SMAP-R F/T product accuracy is comparable
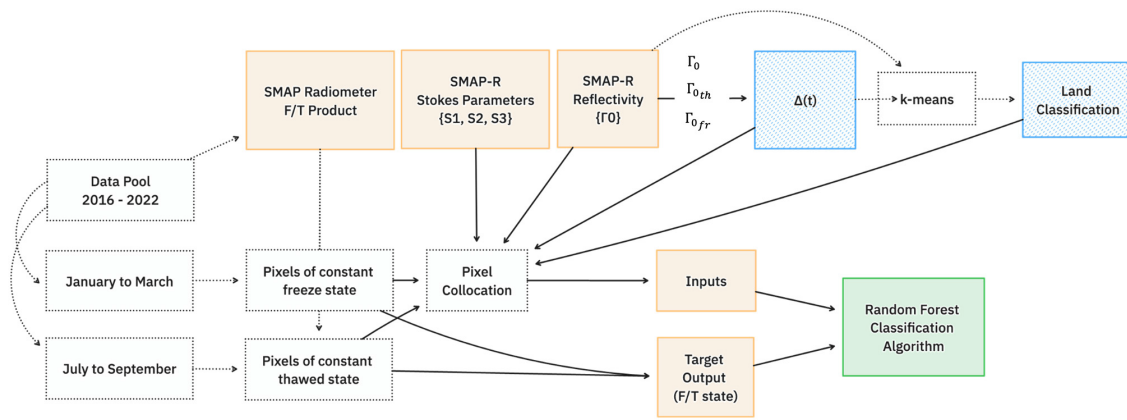
Fig. 8. Improved algorithm diagram flow. Note the use of the SMAP-R full polarimetric GNSS-R data, i.e., the Stokes parameters, in the algorithm.
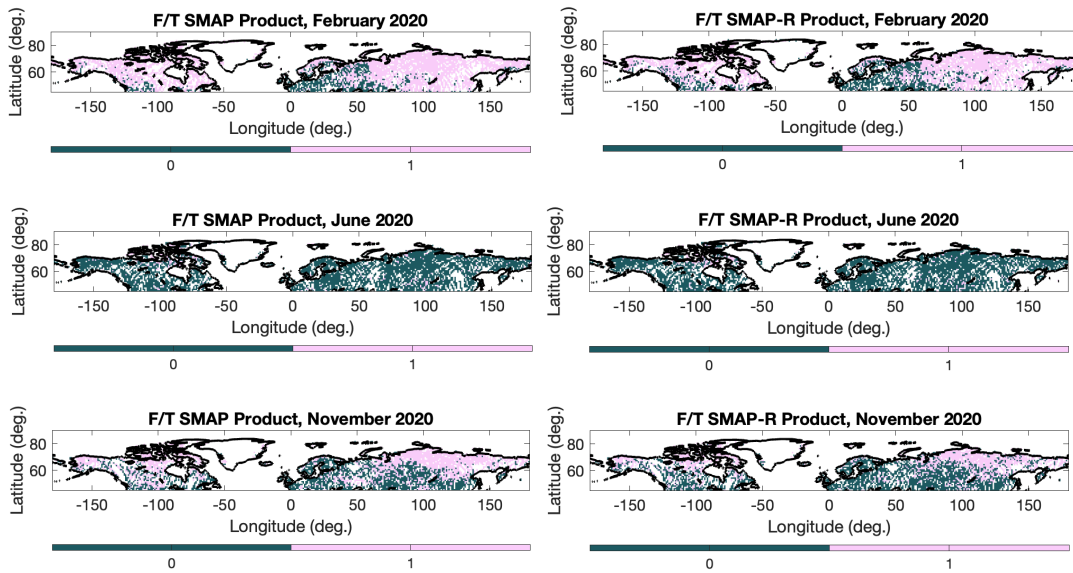

Fig. 9. (Left) F/T classification comparison for 2020 showing the months of February, June, and November from top to bottom: (Right) F/T SMAP official product, and F/T SMAP-R product.
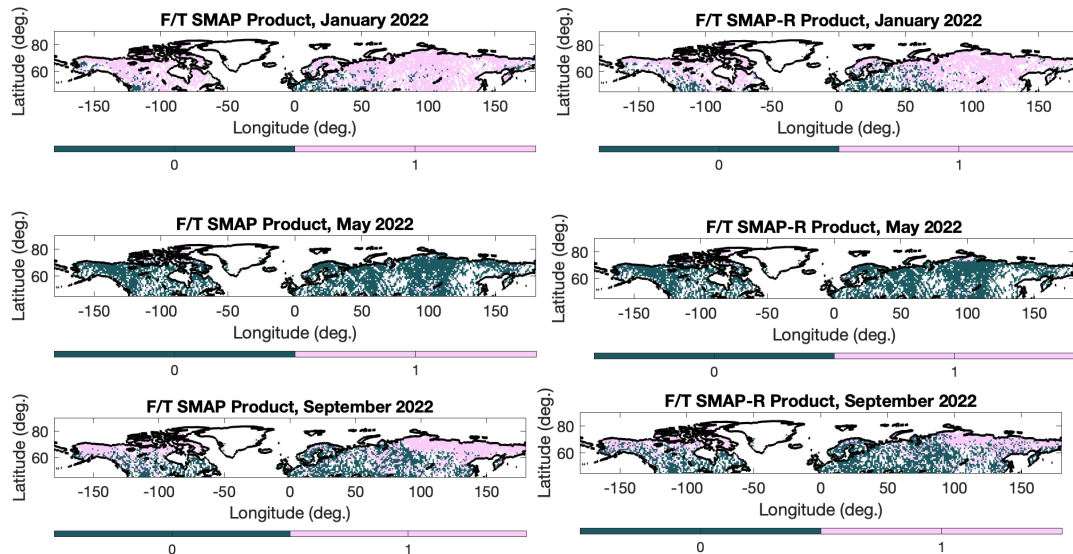

Fig. 10. (Left) F/T classification comparison for 2022 showing the months of January, May, and September from top to bottom: (Right) F/T SMAP official product, and F/T SMAP-R product.

to that of the validated SMAP F/T official product. The value of the random forest comes from the use of full polarimetric GNSS-R data (use of the four Stokes parameters)

which provide a full reconstruction of the Earth polarimetric signatures and enables therefore building a decision based on those parameters as well as the discriminator and the land
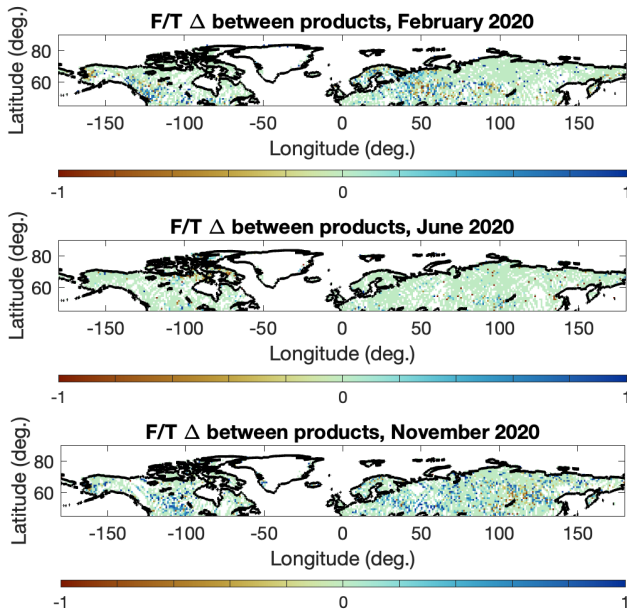
Fig. 11. F/T classification difference between products for the year 2020 showing the months of February, June, and November from top to bottom.
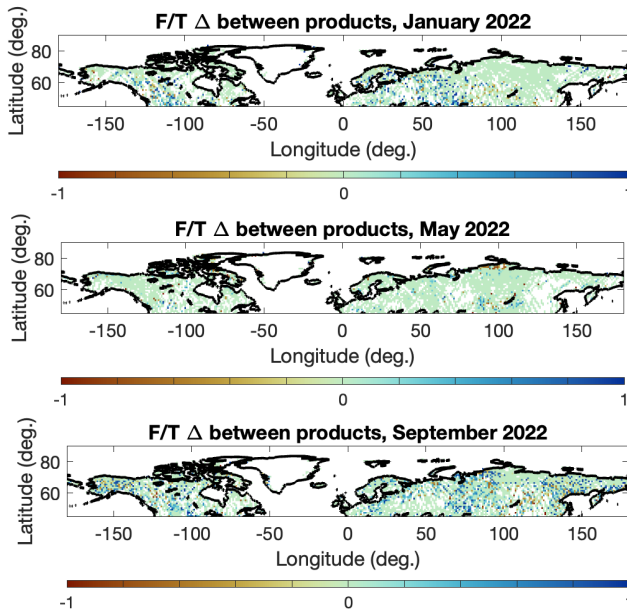

Fig. 12. F/T classification difference between products for the year 2022 showing the months of January, May, and September from top to bottom.

classification, rather than using the discriminator alone to obtain a threshold handpicked based on seasonality per cluster. Next, the section shows the resulting F/T classification maps using the Random Forest Decision Algorithm and compares them to the official SMAP F/T radiometer products.

IV. F/T MAPPING ERROR ASSESSMENT

This section presents the results obtained for the methodology described in Fig. 8, where monthly data are input to the Random Forest Decision Algorithm for F/T classification. The results are shown in the next figures. In Fig. 9, the SMAP F/T official product and the SMAP-R F/T product derived from this study are shown for 2020. Note that only a few months,

February–June–November, are shown, for the sake of reducing the number of plots in the manuscript. Similarly, Fig. 10 shows the results for January 2022, May 2022, and September 2022. In those maps 0 denotes thawed state and 1 frozen state.

The difference between the F/T SMAP official product and the F/T SMAP-R derived product for January 2020, May 2020, and September 2020 are shown in Fig. 11, where most of the pixels show a value of 0. Instances where the F/T SMAP official product classifies pixels as frozen, and the F/T SMAP-R derived product classifies pixels as thawed are positive 1. Instances where the F/T SMAP official product classifies pixels as thawed, and the F/T SMAP-R derived product classifies frozen as thawed are negative 1. Similarly, Fig. 12 shows the results for the difference between the F/T SMAP official product and the F/T SMAP-R derived product for January 2022, May 2022, and September 2022.

V. CONCLUSION

The manuscript describes the development of an algorithm for classifying the frozen and thawed states of the terrestrial cryosphere based on SMAP-R data following a similar approach to the one originally developed by the SMAP mission for its radar. The algorithm uses a seasonal threshold approach, where a discriminator Δ is calculated based on the reflectivity Γ_0 measured at different times. A threshold T is used to classify the surface as frozen or thawed and different T values are assigned to specific landscape types. The algorithm includes a land classification step using k-means clustering to group samples based on their mean Γ_0 and Δ values. The algorithm further refines the threshold values based on the evolution of mean Δ values over time within each cluster, leading to a T LUT for the different clusters.

Since some clusters do not exhibit clear seasonality, a machine learning decision algorithm based on random forest is introduced to incorporate polarimetric information for improved classification. The developed algorithm combines statistical and machine learning techniques to classify frozen and thawed states of the terrestrial cryosphere using SMAP-R data, obtaining an overall performance with a mean accuracy of 97.13%, a mean precision of 99.09%, a mean sensitivity of 95.77%, and a mean specificity of 98.86%, when compared to the validated SMAP FT official product.

This manuscript has proven the capability and added value of full polarimetric GNSS-R measurements, such as the currently available from the SMAP-R instrument, to provide F/T classifications of the cryosphere. We have obtained estimates of surface binary F/T state for the region north of 45° N latitude, which includes the boreal forest zone, with a mean spatial classification accuracy of 97.13% at 36 km spatial resolution and 30-day average intervals. While the spatial resolution and the average intervals do not meet the SMAP mission's initial objective for the F/T product, this manuscript intends to prove the utility of polarimetric GNSS-R data to assess FT classifications. The work developed achieves a binary F/T classification, and therefore does not capture transitional values. As it was shown in [30], the mean polarimetric signatures of different areas can also serve to observe and assess transitional periods.

While this manuscript's goal is not to address F/T transition conclusions, there is potential for this type of measurement to help determine if the F/T state transition occurs earlier or later in the year. This information would enable the determination of a trend in the time of the transition and could potentially link it to global warming effects. For these measurements to be meaningful, higher temporal resolutions are needed. Future constellations of polarimetric GNSS-R instruments are planned by private companies, which will enable meaningful maps below the 1-week time scale. Regarding SMAP-R data the average interval can be improved using 15-day to 30-day temporal moving windows through the year, or using daily sparse measurements. These data can be combined with the SMAP radiometer to improve the accuracy of the assessments, but in that case, the SMAP-R algorithm should be first assessed using in situ data. The intent is, therefore, not to develop an operational F/T product but to demonstrate that polarimetric GNSS-R signals can produce accuracies like those of a radiometer. Future missions such as NASA-ISRO Synthetic Aperture Radar (NASAR) will also be an asset in adding high spatial resolution measurements, that will add relevant backscatter information to the brightness temperatures from radiometers, such as SMAP and SMOS, as well as to the forward-scattering information from polarimetric GNSS-R, such as SMAP-R measurements or the measurements from the future planned constellations of satellites from Muon Space, HydroGNSS and YUNYAO.

Acknowledgment

References

[1] J. Kimball, "Application of the NASA Scatterometer (NSCAT) for determining the daily frozen and nonfrozen landscape of Alaska," *Remote Sens. Environ.*, vol. 75, no. 1, pp. 113–126, Jan. 2001, doi: 10.1016/s0034-4257(00)00160-7.

[2] K. C. McDonald, J. S. Kimball, E. Njoku, R. Zimmermann, and M. Zhao, "Variability in springtime thaw in the terrestrial high latitudes: Monitoring a major control on the biospheric assimilation of atmospheric CO_2 with spaceborne microwave remote sensing," *Earth Interact.*, vol. 8, no. 20, pp. 1–23, Dec. 2004, doi: 10.1175/1087-3562(2004)8<1:VISTIT>2.0.CO;2.

[3] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing: Active and Passive*, vol. 2. Norwood, MA, USA: Artech House, 1986.

[4] J. Way et al., "The effect of changing environmental conditions on microwave signatures of forest ecosystems: Preliminary results of the March 1988 Alaskan aircraft SAR experiment," *Int. J. Remote Sens.*, vol. 11, no. 7, pp. 1119–1144, Jul. 1990, doi: 10.1080/01431169008955084.

[5] R. S. Dunbar et al. (Dec. 9, 2014). *Algorithm Theoretical Basis Document (ATBD) SMAP Level 3 Radar Freeze/Thaw Data Product (L3_FT_A)*. [Online]. Available: https://smap.jpl.nasa.gov/documents/

[6] S. Dunbar et al. (Oct. 15, 2016). *Algorithm Theoretical Basis Document (ATBD) SMAP Level 3 Radiometer Freeze/Thaw Data Products (L3_FT_P and L3_FT_P_E)*. [Online]. Available: https://smap.jpl.nasa.gov/documents/

[7] C. Derksen et al., "Retrieving landscape freeze/thaw state from soil moisture active passive (SMAP) radar and radiometer measurements," *Remote Sens. Environ.*, vol. 194, pp. 48–62, Jun. 2017, doi: 10.1016/j.rse.2017.03.007.

[8] K. Rautiainen et al., "SMOS prototype algorithm for detecting autumn soil freezing," *Remote Sens. Environ.*, vol. 180, pp. 346–360, Jul. 2016, doi: 10.1016/j.rse.2016.01.012.

[9] Y. Kim, J. S. Kimball, K. C. McDonald, and J. Glassy, "Developing a global data record of daily landscape freeze/thaw status using satellite passive microwave remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 949–960, Mar. 2011, doi: 10.1109/TGRS.2010.2070515.

[10] Y. Kim, J. S. Kimball, J. Glassy, and J. Du, "An extended global Earth system data record on daily landscape freeze–thaw status determined from satellite passive microwave remote sensing," *Earth Syst. Sci. Data*, vol. 9, no. 1, pp. 133–147, Feb. 2017, doi: 10.5194/essd-9-133-2017.

[11] K. Rautiainen, D. Comite, J. Cohen, E. Cardellach, M. Unwin, and N. Pierdicca, "Freeze–Thaw detection over high-latitude regions by means of GNSS-R data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 4302713, doi: 10.1109/TGRS.2021.3125315.

[12] T. Zhao, L. Zhang, L. Jiang, S. Zhao, L. Chai, and R. Jin, "A new soil freeze/thaw discriminant algorithm using AMSR-E passive microwave imagery," *Hydrological Processes*, vol. 25, no. 11, pp. 1704–1716, May 2011, doi: 10.1002/hyp.7930.

[13] T. Zhao et al., "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 Space Sci.*, vol. 4, no. 8, pp. 472–484, Aug. 2017, doi: 10.1002/2017ea000277.

[14] H. Carreno-Luengo, S. Lowe, C. Zuffada, S. Esterhuizen, and S. Oveisgharan, "Spaceborne GNSS-R from the SMAP mission: First assessment of polarimetric scatterometry over land and cryosphere," *Remote Sens.*, vol. 9, no. 4, p. 362, Apr. 2017, doi: 10.3390/RS9040362.

[15] C. Chew et al., "SMAP radar receiver measures land surface freeze/thaw state through capture of forward-scattered L-band signals," *Remote Sens. Environ.*, vol. 198, pp. 333–344, Sep. 2017, doi: 10.1016/J.RSE.2017.06.020.

[16] J. F. Munoz-Martin, N. Rodriguez-Alvarez, X. Bosch-Lluis, and K. Oudrhiri, "Stokes parameters retrieval and calibration of hybrid compact polarimetric GNSS-R signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022, Art. no. 5113911, doi: 10.1109/TGRS.2022.3178578.

[17] J. F. Munoz-Martin, X. Bosch-Lluis, N. Rodriguez-Alvarez, and K. Oudrhiri, "Calibration strategy for compact polarimetric GNSS-R instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023, Art. no. 5103513, doi: 10.1109/TGRS.2023.3266602.

[18] J. F. Munoz-Martin, N. Rodriguez-Alvarez, X. Bosch-Lluis, and K. Oudrhiri, "Analysis of polarimetric GNSS-R Stokes parameters of the Earth's land surface," *Remote Sens. Environ.*, vol. 287, Mar. 2023, Art. no. 113491, doi: 10.1016/j.rse.2023.113491.

[19] N. Rodriguez-Alvarez, J. F. Munoz-Martin, X. Bosch-Lluis, and K. Oudrhiri, "A hybrid compact polarimetry GNSS-R analysis of the Earth's cryosphere," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, 2023, Art. no. 4301615, doi: 10.1109/TGRS.2023.3280363.

[20] M. J. Unwin et al., "An introduction to the HydroGNSS GNSS reflectometry remote sensing mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6987–6999, 2021, doi: 10.1109/JSTARS.2021.3089550.

[21] N. Fu and F. Li, "An introduction of GNSS reflectometer remote sensing mission from yunyao aerospace technology Co., Ltd," in *Proc. IEEE Specialist Meeting Reflectometry using GNSS other Signals Opportunity (GNSS+R)*. China: IEEE, Sep. 2021, pp. 77–81, doi: 10.1109/GNSSR53802.2021.9617716.

[22] J. F. Munoz-Martin, N. Rodriguez-Alvarez, X. Bosch-Lluis, and K. Oudrhiri, "Detection probability of polarimetric GNSS-R signals," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3236601.

[23] N. Rodriguez-Alvarez, J. F. Munoz-Martin, X. Bosch-Lluis, K. Oudrhiri, D. Entekhabi, and A. Colliander, "The first polarimetric GNSS-reflectometer instrument in space improves the SMAP mission's sensitivity over densely vegetated areas," *Sci. Rep.*, vol. 13, no. 1, Mar. 2023, Art. no. 1, doi: 10.1038/s41598-023-30805-7.

[24] R. K. Raney, "Polarimetric portraits," *Earth Space Sci.*, vol. 8, 2021, Art. no. e2021EA001768, doi: 10.1029/2021ea001768.

[25] A. S. Belward, J. E. Estes, and K. D. Kline, "The IGBP-DIS global 1-Km land-cover data set discover: A project overview," *Photogramm. Eng. Remote Sens.*, vol. 65, no. 9, pp. 1013–1020, Sep. 1999.

[26] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.

[27] X. Xu, R. S. Dunbar, C. Derksen, A. Colliander, Y. Kim, and J. S. Kimball, "SMAP L3 radiometer global and Northern Hemisphere daily 36 km EASE-grid freeze/thaw state, version 3 [Data Set]," NASA Nat. Snow Ice Data Center Distrib. Act. Arch. Center, Boulder, CO, USA, 2020. Accessed: May 6, 2024. [Online]. Available: https://doi.org/10.5067/ZJOKL452HRLD

[28] N. Rodriguez-Alvarez, E. Podest, K. Jensen, and K. C. McDonald, "Classifying inundation in a tropical wetlands complex with GNSS-R," *Remote Sens.*, vol. 11, no. 9, p. 1053, May 2019, doi: 10.3390/rs11091053.

[29] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[30] N. Rodriguez-Alvarez, J. F. Munoz-Martin, X. Bosch-Lluis, and K. Oudrhiri, "Initial evaluation of freeze/thaw state and sea ice detection using the SMAP-R dataset," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 3876–3879, doi: 10.1109/IGARSS46834.2022.9883525.

Nereida Rodriguez-Alvarez (Senior Member, IEEE) received the B.S., and M.S. degrees in telecommunications and the Ph.D. degree in remote sensing from the Universitat Politecnica de Catalunya, Barcelona, Spain, in 2005, 2007, and 2011, respectively.

She has collaborated with the National Oceanographic and Atmospheric Administration (NOAA), Boulder, CO, USA the Atmospheric Science Department, Colorado State University, Fort Collins, CO, USA; and Purdue University, West Lafayette, IN, USA, as Post-Doctoral Researcher. She was part of the Cyclone Global Navigation Satellite System (CYGNSS) Science Team for three years. Since 2018, she has been with the Planetary Radar and Radio Science Group, Jet Propulsion Laboratory, Pasadena, CA, USA. Currently, she continues to explore applications of GNSS-R in earth science, while she expands her knowledge to Planetary science to sense the properties of Solar System bodies. She specializes in bistatic radar techniques for sensing the earth employing signals of opportunity, through Global Navigation Satellite System–Reflectometry (GNSS-R) and has contributed to the retrieval of land geophysical parameters and ocean surface winds. She has engaged with Goldstone Solar System Radar (GSSR) signal processing activities. While at JPL, she has made relevant accomplishments on wetlands, polar sea ice, and floods for what she has received the NASA Early Career Public Achievement Award in 2019.

Joan Francesc Munoz-Martin (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in telecommunications engineering from the Universitat Politècnica de Catalunya (UPC BarcelonaTech), in 2014, 2017, and 2021, respectively.

He joined the UPC NanoSat Lab, Barcelona, Spain, in 2013, and he has contributed to several CubeSat-based missions for earth Observation ever since. While at UPC, he actively participated in the software development for OBDH, TT&C, and ground segment, for the 3Cat-X family (-2, -4, and -5), and he led the development of the remote sensing payloads for 3Cat-4, 3Cat-5, and 3Cat-7 (FMPL-1, -2, and -3). He has expertise collaborating with the European Space Agency with the BEXUS 19, Fly Your Satellite II, and the 3Cat-5/FSSCat mission, and in the GNSS-R experiment of the MOSAiC campaign in the Arctic Sea. In 2021, he moved to the Jet Propulsion Laboratory, Pasadena, CA, USA, as a Researcher. During his Postdoctoral stay, he contributed to the Goldstone Solar System Radar (GSSR) signal processing activities and continued its activities in microwave radiometry and GNSS-R, and designing new Earth Observation instruments. He is currently a Technologist with the Signal Processing and Networks Group, the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. His research expertise is in remote sensing using GNSS-R and radiometry. His research interests include instrument design and calibration for microwave radiometers and bistatic radars and other novel instruments using quantum technologies.

Xavier Bosch-Lluis (Senior Member, IEEE) received the B.Sc. and M.S. degrees in telecommunications engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2003 and 2005, respectively, and the Ph.D. degree in remote sensing from the Department of Signal Theory and Communications, UPC, in 2011.

Following his doctoral studies, he was a Postdoctrol with the Department of Electrical and Computer Engineering, Microwave Systems Laboratory, Colorado State University, Fort Collins, CO, USA. Currently, he is a Technologist with the Signal Processing and Networks group at the Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA, USA. Here, he has taken on significant roles in innovative projects such as the Cold Atom Laboratory (CAL), Quantum Gravimetry Gradient (QGG), and the use of high-temperature superconductors in space for quantum remote sensing. Additionally, he played a pivotal role in the Microwave Temperature and Humidity Profiler (MTHP), where he led the development and demonstration of this advanced atmospheric sounder, enhancing our understanding of atmospheric dynamics through high-resolution measurements. He has made relevant accomplishments the demonstrating the first hyperspectral multi-incidence-angle microwave sounder resulting in improved estimation of the planetary boundary layer.

Dr. Bosch-Lluis has received the NASA Early Career Public Achievement Award in 2024.

Kamal Oudrhiri is currently the Deputy Section Manager of the Communications Architectures and Research Section, the Jet Propulsion Laboratory (JPL), Pasadena, CA, USA. He has over 17 years of experience in Technical Management, Radio Science, End-to-End Systems Engineering, Cognizant Engineering, and Deep Space Projects operations, specializing in planetary radar and radio science. Managed multidisciplinary teams through the development and delivery of the Cold Atom Laboratory (CAL), and Ultra Stable Oscillators for Mars Reconnaissance Orbiter (MRO) and Gravity Recovery and Interior Laboratory (GRAIL) projects within cost and schedule. He is the CAL Mission Manager. Oudrhiri has served in key roles on multiple NASA missions: His research focuses on planetary atmospheric characterization, gravity science, and bistatic radar.

Dr. Oudrhiri has received three NASA JPL Mariner & Voyager Honor Awards and multiple NASA Group Achievement Awards. He was a recipient of the People Leadership Award, which recognizes exceptional people-leadership skills critical to the ongoing success of NASA exploration missions. The Mars Exploration Rovers (MER), the Cassini mission to Saturn, the GRAIL lunar mission, the Mars Science Laboratory (MSL), the Juno mission to Jupiter, and the New Horizons mission to Pluto.