Assessment of Four Near-Surface Soil Freeze/Thaw Detection Algorithms Based on Calibrated Passive Microwave Remote Sensing Data Over China

Wanwan Shao1 and Tingjun Zhang 1

1Key Laboratory of Western China’s Environmental Systems (MOE), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, China

Abstract
The near-surface soil freeze/thaw (F/T) cycle affects the surface energy balance, hydrological processes, and soil greenhouse gas release. Passive microwave remote sensing data are the most widely used method for determining the near-surface soil F/T status. While many algorithms have been developed for this purpose, their performance has never been compared using same and large in situ data set. Here, we evaluate and inter-compare the classification results of the four most widely used algorithms using a large ground truth data set covering China. Based on the ground observations, our evaluation shows a wide range of near-surface soil F/T detection performance, with Cohen’s kappa coefficient ranging from 0.42 to 0.72 and an overall accuracy between 73.8% and 86.2%. We suggest performing parameter calibration for the decision tree algorithm and the discriminant function algorithm before applying them to areas outside of the training sites. All these four algorithms exhibited remarkable uncertainty in the detection of the onset and offset of near-surface soil freezing with root mean squared errors of more than 27 days. These results suggest that careful cautions should be taken when outputs of these algorithms are used for investigations of long-term changes.

1. Introduction
The phase transition between water and ice occurs in about 60% of near-surface soils globally (Kim et al., 2011). The near-surface soil freeze/thaw (F/T) cycle is characterized as a toggle switch for land surface processes because of its strong influence on surface energy balance (Zhao et al., 2017), hydrological processes (Guo et al., 2011), vegetation dynamics (Michelsen et al., 2002), carbon dioxide emissions (Schimel & Clein, 1996), and ecosystem function (Freppaz et al., 2007). The seasonally frozen ground is spatially and temporally variable (Kim et al., 2011, 2012). During the past few decades, many studies have shown that rising near-surface air temperature (T_s) has contributed to earlier onset of near-surface soil thaw and has decreased the duration of the frozen period (Kim et al., 2012; McDonald et al., 2004; Peng et al., 2016; Wang et al., 2015).

There are a number of methods used to assess whether near-surface soil is frozen or thawed. These can be classified as (i) traditional methods, which include interpolation of geological maps, in situ measurements (e.g., Peng et al., 2016), and extrapolation based on terrain features known to be closely associated with permafrost or seasonally frozen ground, such as rock glaciers (e.g., Boeckli et al., 2012); (ii) geophysics, most notably ground penetrating radar (e.g., Cao et al., 2017; Ma et al., 2015); (iii) numerical modeling (e.g., Freppaz et al., 2007; Guo et al., 2018); and (iv) remote sensing (e.g., Bartsch et al., 2007; Kim et al., 2017; Leverington & Duguy, 1996). While the traditional and geophysical methods can describe the near-surface soil F/T cycle well at the site scale, they are inadequate for larger-scale studies and trend detection, especially in mountains and in remote areas where observations are sparse. Numerical modeling is challenging due to the limited availability of reliable forcing data, such as snow distribution and soil conditions. Over the past several decades, satellite remote sensing techniques, validated by ground truth data, have been used successfully to investigate the near-surface soil F/T cycle from local to global scales and have shown numerous advantages over other methods. Passive microwave remote sensing has emerged as one of the most widely used systems for near-surface soil F/T cycle investigations due, in part, to its temporal continuity (from 1978 to present), global coverage, and short revisit period (typically twice daily). Passive microwave remote sensing detects the electromagnetic radiation spontaneously emitted by matter based on its temperature, composition, and physical structure (Zhang et al., 2004). The discrimination between frozen and thawed soil...
is made possible by the strong contrast in electromagnetic properties between liquid and solid (ice) phase of water which is typically present in near-surface soil. Algorithms that use passive microwave remote sensing data for near-surface soil F/T detection can be broadly grouped into several categories: (i) dual-index algorithms and their variants, which exploit both the spectral gradient and the relationship between brightness temperature \( T_b \) and air temperature \( T_a \) or ground surface temperature \( T_g \) (Jin et al., 2015; Kim et al., 2011, 2017); (ii) temporal change detection algorithms, which use the differences in \( T_b \) between different frequencies (Chen et al., 2019; Smith et al., 2004); (iii) decision tree algorithms, which use the characteristics of the volume scattering effect and the lower \( T_g \) of frozen ground (Jin et al., 2009); and (iv) model-based discriminant algorithms, which train discriminant models based on observations or use combined models that consider different surface conditions, such as soil moisture and snow cover (Hu et al., 2017; Zhao et al., 2011, 2017). These methods are mainly based on the signatures associated with frozen soil, including lower physical (or thermal) temperature, higher surface emissivity, larger optical thickness, and lower \( T_b \) (Jin et al., 2015; Zhang et al., 2004).

Generally, these algorithms have been evaluated to some extent. However, some of the evaluations are limited to small data sets that included only a few sites or short time series. This may hinder the application of some of these algorithms to novel study areas outside of the test sites. Despite increasing efforts to investigate near-surface soil F/T changes, the detection algorithms based on passive microwave remote sensing have not been inter-compared using unique and large validation data sets. This is an essential step in describing the state-of-the-art for near-surface soil F/T detection as well as for identifying any possible biases. It is also critical for identifying priorities when updating these methods or the satellite and sensor designs in the future. Additionally, even though sensor architectures may be consistent as one instrument replaces its predecessor, the calibration of individual instruments is not identical (Sapiano et al., 2013). Therefore, inter-calibration among historic data sets is important for deriving the near-surface soil F/T cycle as the differences between the instruments could introduce uncertainty or inconsistency in the raw data and derived results. However, these differences are often either ignored (e.g., Jin et al., 2015) or calibrated using simple techniques such as linear regression based on periods of overlapping data (e.g., Kim et al., 2012, 2017). The objectives of this study are to (1) evaluate and inter-compare four widely used near-surface soil F/T detection algorithms and (2) discuss their bias and potential uncertainties.

2. Data

2.1. Calibrated Enhanced Resolution Brightness Temperature

In this study, we used the Calibrated Enhanced Resolution Brightness Temperature (CETB) data product obtained from the National Snow & Ice Data Center (NSIDC). The CETB was recently developed by the National Aeronautics and Space Administration (NASA) using the most mature available Level-2 records from three passive microwave sensors on multiple platforms between 1978 and 2017 (Figure 1) (Brodzik & Long, 2018). More specifically, the CETB comprises historic data from the Special Sensor Microwave Imager/Sounder (SSMIS), Special Sensor Microwave/Imager (SSM/I), and Scanning Multichannel Microwave Radiometer (SMMR) instruments which have been reprocessed to a higher standard. Compared to the original data sets, the CETB includes improved cross-sensor calibration and quality checking, enhanced resolution (up to 3.125 km), better quality control, improved projection grids, and local time-of-day processing. The newest version of the CETB data (v1.3) used here is physically consistent based on multiple independent inter-calibration techniques (Brodzik & Long, 2018; Sapiano et al., 2013). The CETB can therefore provide suitable \( T_b \) long-term data records for studying the near-surface soil F/T cycles.

Although spatial resolutions of up to 3.125 km are available for the CETB, only data with 25 km resolution were used here in order to make this contribution more comparable with previous studies. The CETB provides continuous global \( T_b \) measurements twice per day, corresponding to the ascending and descending tracks separated by the overpass time of the radiometers.

The effects of atmospheric conditions (e.g., aerosols and water vapor) and clouds on microwave remote sensing data quality were assumed to be minimal, and hence, no correction for atmospheric effects was applied. To obtain a continuous spatiotemporal record of \( T_b \), any missing values caused by orbital gaps between satellite overpasses were filled using a two-step procedure. First, a temporal linear interpolation of adjacent \( T_b \) was conducted with a window size of 11 days to fill small gaps (Jin et al., 2015). This step significantly improved the data completeness (Table 1). The large gaps were then filled by exploiting the linear relationship between \( T_b \) and \( T_g \) using downscaled meteorological data from ERA-Interim (for details see section...
2.4). This second step was conducted annually (Figure A1). Figure A3 gives an overview of the interpolation approach used to fill the \( T_b \) gaps. To further improve the reliability and continuity of CETB, whenever overlaps in data occurred between multiple platforms, we chose the platform with the greater completeness and longer observation period. Furthermore, newer satellite platforms were favored over older ones. The resulting time series used data from the Nimbus-7 of SMMR from October 1978 to July 1987, the F08 of SSM/I from July 1987 to December 1991, the F11 of SSM/I from December 1991 to February 2000, the F15 of SSM/I from February 2000 to October 2005, and the F16 of SSMIS from November 2005 to June 2017. This produced a continuous record of \( T_b \) spanning 1978–2017 (Table 1 and Figure 1a). However, because the \( T_b(85V) \) was only available for the SSM/I sensor, the temporal coverage for this frequency only starts in July 1987 and between February 2000 and 2016; \( T_b \) records from F15 were used instead.

2.2. Land Use Map and Topography

To derive land cover masks, we used the Multi-source Integrated Chinese Land Cover (MICLCover) map because it has better performance (overall accuracy >70%) compared to other popularly used maps (Ran et al., 2012). The original International Geosphere-Biosphere Programme (IGBP) land cover classification used in MICLCover was aggregated into five major types: forest (including evergreen needle leaf forest, evergreen broad leaf forest, deciduous needle leaf forest, deciduous broad leaf forest, and mixed forest), shrub (including closed shrub lands and open shrub lands), grass (including woody savannas, savannas, grasslands, permanent wetlands, and croplands), bare land (including urban and built-up, cropland/natural vegetation mosaic, snow and ice, and barren or sparsely vegetated), and water by following Xiao et al. (2018). This was done to account for the biased distribution of CMA (China Meteorological Administration) stations that resulted from using the original categories. The MICLCover map was resampled from its original 1 km spatial resolution to match the NSIDC EASE-Grid format using majority resampling. The Global Digital Elevation Model version 2 (GDEM2) with a spatial resolution of 1 arc second (~30 m) was first aggregated to 3 arc seconds to avoid the noise in the original data set and then resampled using bi-linear interpolation.
Table 1
Metadata for the Original Calibrated Enhanced Resolution Brightness Temperature (CETB) Data Set and the Subset Used in This Study

| Sensor | Swath (km) | Platform | Temporal coverage [dd/mm/yyyy] | Completeness (%) | Original | Interpolated |
|--------|------------|----------|--------------------------------|------------------|-----------|-------------|
| SMMR   | 780        | NIMBUS7  | 25/10/1978–20/08/1987          | 01/01/1979–08/07/1987 | 14.6      | 57.6        |
| SSM/I  | 1,394      | F08      | 09/07/1987–31/12/1991          | 09/07/1987–02/12/1991 | 45.9      | 76.7        |
|        |            | F10      | 08/12/1990–14/11/1997          | —                | 53.5      | —           |
|        |            | F11      | 03/12/1991–16/05/2000          | 03/12/1991–02/05/1995 | 50.9      | 91.6        |
|        |            |          | 03/12/1991–22/02/2000          |                  | —         | —           |
|        |            | F13      | 03/05/1995–19/11/2009          | 03/05/1995–31/10/2005 | 64.5      | 98.7        |
|        |            | F14      | 07/05/1997–23/08/2008          | —                | 65.8      | —           |
|        |            | F15      | 23/02/2000–31/12/2016          | 23/02/2000–31/12/2016 | 55.0      | 99.7        |
| SSMIS  | 1,707      | F16      | 01/11/2005–31/12/2016          | 01/11/2005–31/12/2016 | 80.8      | 99.6        |
|        |            | F17      | 01/03/2008–31/12/2016          | —                | 76.4      | —           |
|        |            | F18      | 08/03/2010–31/12/2016          | —                | 81.0      | —           |
|        |            | F19      | 27/11/2014–02/09/2016          | —                | 79.7      | —           |
| Overall| —          | —        | 25/10/1978–31/12/2016          | 01/01/1979–31/12/2016 | 59.9      | 88.1        |

Note. The interpolated completeness is derived using adjacent $T_b$ (Figure A3), and only the selected data are present here. Please note that different temporal coverage was used for $T_b(85V)$ (underlined) because it is not available for the sensor of SMMR and SSMIS.

to match the 25 km NSIDC EASE-Grid format (version 2, Brodzik et al., 2012). The 25 km DEM is then used for $T_g$ downscaling. Finally, before any soil F/T classification took place, desert regions were masked from the original CETB data based on “The map of desert distribution in 1:2,000,000 in China” (Chao, 1984).

2.3. Ground Observations

We obtained $T_g$ records at 0 cm depth from the China Meteorological Administration (CMA) to (i) derive the $T_g$ threshold for the dual-index algorithm and (ii) evaluate the classification results for all near-surface soil F/T detection algorithms. $T_g$ was measured with a white mercury ball thermometer with a diameter of about 3 mm and was partially buried in order to reduce the influences radiation and atmosphere (Wang et al., 2015). The satellite data from the a.m. and p.m. passes were found to be more closely related to daily minimum and maximum $T_g$, respectively, than to the daily mean (Xie et al., 2013). The daily maximum was not used here because no frozen ground was found for any of the sites in the summer. Therefore, only data from the a.m. satellite pass was used for evaluation. It is important to note that when the ground was covered by snow, the snow surface temperature was measured instead of the soil temperature (Wang et al., 2015). We make the assumption that the near-surface soil is frozen when it is covered by snow, as snow cover can only be maintained when the near-surface soil temperature is below 0 °C (Zhang et al., 2004). The assumption is reasonable as the classification accuracy of snow-covered period is found to be better than that of snow-free period based on the measurements; however, these results are not presented here.

Regions with little F/T activity were excluded from the study. To do this, the freezing probability was calculated for each site by dividing the number of days with $T_g < 0$ by the total number of days. Sites were removed if their overall freezing probability during 1979–2016 was less than 0.1. Sites in grid cells containing more than 50% water or in desert were also removed. This resulted in a total of 526 stations with ~ 15.0 × 10^6 observations (Figure 2). These stations extend over a large area of China and cover all land use types with a distribution of about 14.1% in forest, 4.7% in shrub areas, 31.0% in grassland, and 51.7% in bare land (see section 2.3 for the description of the land use map). In addition, the CMA data set represents a wide range of elevations from about 0 to 5,600 m (Figure 3a). The distribution of aspects for the stations was approximately uniform (26.8% on the east slope, 24.1% on the south slope, 24.5% on the west slope, and 24.5% on the north slope); however, the majority (63.5%) of stations were located in relatively flat areas with slopes of less than 10° (Figure 3b). The ground observations used here are therefore considered to be representative...
enough to conduct the evaluation and inter-comparisons among different near-surface soil F/T detection algorithms over most areas of China but may have increased uncertainty in steep terrain. Note that no sites are located within the same grid.

2.4. Reanalysis and Downscaling

As discussed below, $T_a$ measurements were required to derive $T_b$ thresholds for the modified seasonal threshold algorithm (MSTA). For this, ERA-Interim, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), was used. ERA-Interim uses a fully coupled atmosphere-ocean-land model, making it suitable to describe large-scale climate, and it has a similar temporal coverage as the CEBT (January 1979 to current) (Dee et al., 2011). In order to match with the CEBT data, the original ERA-Interim grids
Table 2

| Algorithm | SDIA | MSTA | DTA | DFA |
|-----------|------|------|-----|-----|
| Frequencies | 18/19V, 37V | 19V, 37V | 19V, 22V, 37V, 85V | 37V, 18H/19H |
| Inputs | Observed Tg | Grid Ta | Observed Tg | Observed/simulated Tb |
| Spatial variable | Land-cover-wise | Grid-cell-wise | No | No |
| Temporally variable | Yes | Yes | No | No |
| Temporal coverage | 1978–2016 | 1978–2016 | 1987–2016 | 1978–2016 |
| Improvements | Seasonal threshold | Downscaled Ta & SG | Simplified | — |
| Reference | Jin et al. (2015) | Kim et al. (2017) | Jin et al. (2009) | Zhao et al. (2011) |

Note. SDIA = seasonal dual-index algorithm; MSTA = modified seasonal threshold algorithm; DTA = decision tree algorithm; DFA = discriminant function algorithm; Ta = near-surface air temperature; Tb = brightness temperature; Tg = ground surface temperature; SG = spectral gradient.

with a spatial resolution of ∼79 km were downscaled to 25 km based on the elevation correction method (Gao et al., 2012). This downscaling method has been shown to perform well for fine-scale T_a (Cao et al., 2017, Cao, Zhang, et al., 2019; Cao, Quan, et al., 2019).

3. Methods

3.1. Near-Surface Soil F/T Classification Algorithms

This study is focused on assessing only the most widely used near-surface soil F/T algorithms rather than presenting a review and synthesis of all such algorithms. Table 2 summarizes the four that were selected for this study: the seasonal dual-index algorithm (SDIA), MSTA, decision tree algorithm (DTA), and the discriminant function algorithm (DFA). Each of these algorithms is discussed below. The temporal change detection algorithm was not used here because it is not able to detect repeated F/T cycles in autumn and spring.

3.1.1. Seasonal Dual-Index Algorithm

The vertically polarized Tb at 37GHz, written as Tb(37V), is strongly correlated with Ta and Tg at depths less than 5 cm (Jin et al., 2015; Zhang et al., 2003). However, the presence of snow and/or thick vegetation cover reduces the strength of this correlation. Additionally, when soil freezes, the volume-scattering effect becomes obscured in the lower frequency channel, which contributes to a negative spectral gradient (SG, Zhao et al., 2011). The combination of Tb(37V) and the 19 to 37 GHz SG can therefore be used to assess whether near-surface soils are frozen or thawed (Zhang et al., 2004). The near-surface soil freeze status thaw status of SDIA (FTSDIA) can be expressed as

\[
FT_{SDIA} = \begin{cases} 
1, & T_b(37V) \leq T_{cutoff} \text{ & } SG \leq SG_{cutoff} \\
0, & \text{otherwise} 
\end{cases}
\]

(1)

where a value of 1 means the near-surface soil is classified as frozen. The variables T_{cutoff} [K] and SG_{cutoff} [K GHz⁻¹] are empirically defined threshold or cutoff values for T_b and the SG, respectively. The spectral gradient between 19 and 37 GHz is calculated as

\[
SG = \frac{T_b(37V) - T_b(19V)}{37 - 19}.
\]

(2)

The value of T_b is equal to the product of the physical temperature and the emissivity of the measured surface. Emissivity is affected by both the sensor (e.g., frequency, polarization, and incidence angle of the satellite observation) and the surface conditions (e.g., roughness, vegetation cover, and soil moisture) (Smith et al., 2004). For this reason, the T_b cutoff is derived annually or seasonally for each land cover type rather than being treated as a constant (detailed T_{cutoff} derivation and evaluation is described in section 3.2).

We used weighted nonlinear regression to handle the strong sensitivity of the algorithm to T_{cutoff} (Kim et al., 2017). Values of T_g nearer to 0 °C were assigned a greater weight because of the significance of the freeze-thaw boundary. Weights were derived from a cosine function with a T_g range from −30 to 70 °C, which covered 99.6% of all T_g observations (Figure A1). Figure 4 shows the derived T_{cutoff} used for SDIA. The SG_{cutoff} was assigned a constant value of 0 K GHz⁻¹.
Figure 4. Changes of the mean $T_{\text{cutoff}}$ for seasonal dual-index algorithm (SDIA) based on 10-fold cross-validation. The dashed orange line separates the results of different sensors, while the gray line represents the boundaries of platforms.

3.1.2. Modified Seasonal Threshold Algorithm

MSTA is a variant of the seasonal threshold algorithm which treats $T_b$ thresholds as both spatially and temporally variable. The algorithm may reduce the potential influence of spatial and temporal variations in climate (e.g., precipitation) and land surface conditions (e.g., soil moisture and vegetation cover) on F/T classification and promote greater consistency in large area performance over the long-term record (Kim et al., 2017). As above, the $T_b$ threshold was derived empirically using the weighted, nonlinear relationship between $T_a$ and $T_b$ (see section 2.4) rather than simply interpolating the 2 m surface air temperature from ERA-Interim. The improved representation of $T_a$ is thought to better capture the relationship between $T_b$ and $T_a$ and contribute to higher classification accuracy. Additionally, the SG was introduced here to further improve classification. In other words, MSTA was effectively a variant of the dual-index algorithm with grid-cell-wise $T_b$ thresholds and determined as equation (1), where a value of 1 corresponds to frozen status as above.

3.1.3. Decision Tree Algorithm

Jin et al. (2009) developed a DTA to distinguish between frozen near-surface soil, thawed near-surface soil, precipitation, and desert. There are three critical indices used in the decision tree: the scattering index, $T_b(37V)$, and the polarization difference at 19 GHz. When the ground is frozen, the effect of volume scattering darkening is stronger at higher frequencies. Hence, the scattering index is approximated as the difference between the observed $T_b(85V)$ and the simulated value assuming ideal conditions and no scattering. This can be expressed as

$$SI = T_b(85V)_{\text{sim}} - T_b(85V)_{\text{ps}}$$

where $T_b(85V)_{\text{sim}}$ is the simulated $T_b(85V)$, while $T_b(85V)_{\text{ps}}$ is obtained from the passive microwave sensor. $T_b(85V)_{\text{sim}}$ is estimated based on an empirical relationship using $T_b(19V)$ and $T_b(22V)$ (Grody, 1991):

$$T_b(85V)_{\text{sim}} = 450.2 - 0.506 \times T_b(19V) - 1.874 \times T_b(22V) + 0.00637 \times (T_b(22V))^2$$

$T_b(37V)$ was used as an indicator for soil physical temperature as it had the best agreement with $T_g$ compared to other channels. The 19 GHz polarization difference is used to determine surface roughness and to improve identification of desert. For simplicity, the desert areas were masked using the available desert inventory, and the precipitation class was reclassified as thawed ground. The simplified DTA (FT$\text{DTA}$) is shown in Figure 6 and can be written as

$$FT_{\text{DTA}} = \begin{cases} 1, & SI \geq 5 \text{ & } T_b(37V) \leq 252 \\ 0, & \text{otherwise} \end{cases}$$

SHAO AND ZHANG 7 of 19
Figure 5. Spatial variability of $T_{\text{cutoff}}$ values for the modified seasonal threshold algorithm (MSTA) in 2009. Please note that this example provides the threshold over China regardless of their frozen probability, although we only used the stations with frozen probability $\geq 0.1$ during evaluation.

where a value of 1 corresponds to frozen status, as above.

3.1.4. Discriminant Function Algorithm

DFA is a more recently developed method (Zhao et al., 2011). The criteria and discriminant functions were obtained from both radiometer observations and model simulations. In order to simulate near-surface soil $T_b$ for different surface covers, a combined microwave emission model was first developed by considering various factors, especially the vegetation and snow cover. Then a F/T discriminant algorithm was estimated using simulated $T_b(37V)$ and $T_b(19H)$ by assuming that (i) the $T_g$ is strongly related to $T_b(37V)$ and (ii) the emissivity, which increases considerably when water becomes ice, can be estimated using the ratio of $T_b(19H)$ to $T_b(37V)$. The DFA uses Fisher linear discriminant analysis to classify near-surface soil F/T status. Parameters are validated based on the simulated $T_b$ and the measured ground F/T. This can be written as

$$D_F = 1.47 \times T_b(37V) + 91.69 \times Qe(19H) - 226.77,$$

$$D_T = 1.55 \times T_b(37V) + 86.33 \times Qe(19H) - 242.41,$$

where $D_F$ and $D_T$ are discriminant scores for frozen and thawed near-surface soil, respectively. Here, $Qe$ refers to quasi-emissivity, which is an approximation to the true emissivity given by

$$Qe(19H) = \frac{T_b(19H)}{T_b(37V)}.$$

The near-surface soil is classified as frozen if $D_F$ is greater than $D_T$. Otherwise, it is classified as thawed:

$$FT_{\text{DFA}} = \begin{cases} 1, & D_F > D_T, \\ 0, & D_F \leq D_T. \end{cases}$$

To make the evaluation possible, the near-surface soil F/T status based on DFA was conducted using the $T_b$ from SMMR, SSM/I, and SSMIS, even though it was developed using the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E). The $T_b$ from the platforms of SMMR, SSM/I, and SSMIS show good agreement with those obtained from AMSR-E, and values were calibrated to AMSR-E based on their linear relationship in order to remove any bias (Figure A2).
Figure 6. Decision tree algorithm flow chart revised from Jin et al. (2009). $T_b(85V)_{\text{sim}}$ and $T_b(85V)_{\text{ps}}$ are the simulated and observed values of $T_b(85V)$, respectively.

3.2. Evaluation

To evaluate the performance of the algorithms, the classification results were compared to observations from the CMA. Observed soil F/T status was based on the measured $T_g$. The near-surface soil was considered frozen when $T_g \leq 0 ^\circ C$. Because the $T_{\text{cutoff}}$ value for SDIA must be derived using observational data, we conducted a 10-fold cross-validation to assess its performance. The CMA stations were first divided into four subsets based on land use types. Each time, $\sim 90\%$ of the stations for each land use type were randomly selected from the subsets for deriving $T_b$ thresholds, and the remaining $10\%$ were used for evaluation. Following Zhang et al. (2003), evaluations of the maps with categorical types were conducted using classification accuracies ($acc$).

\[ F_{\text{acc}} = \frac{F_F}{F_F + F_T} \times 100\% , \]

\[ T_{\text{acc}} = \frac{T_T}{T_T + T_F} \times 100\% , \]

\[ O_{\text{acc}} = \frac{F_F + T_T}{F_F + F_T + T_T + T_F} \times 100\% . \]

The subscripts on the right-hand side of the equations identify the classified value. For instance, $F_F$ is the number of frozen observations which were correctly classified as frozen, and $F_T$ is the number of frozen observations which were misclassified as thawed. To avoid the impact of uneven distribution of observation numbers for near-surface soil F/T status, the Cohen's kappa coefficient ($\kappa$), which measures inter-rater
Table 3
Evaluation of Near-Surface Soil F/T Algorithms Derived From Calibrated Passive Microwave Remote Sensing Data During 1979–2016

| Algorithm | SDIA | MSTA | DTA | DFA |
|-----------|------|------|-----|-----|
|           | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ |
| January   | 79.8 | 90.8 | 80.5 | 0.30 | 81.3 | 85.7 | 81.6 | 0.30 | 57.5 | 98.9 | 60.3 | 0.15 | 93.8 | 62.1 | 91.8 | 0.45 |
| February  | 78.2 | 91.7 | 79.7 | 0.42 | 79.3 | 88.9 | 80.4 | 0.42 | 51.5 | 98.5 | 57.3 | 0.20 | 90.7 | 69.3 | 88.2 | 0.51 |
| March     | 72.1 | 90.0 | 77.1 | **0.52** | 70.1 | 91.9 | 76.2 | 0.51 | 35.6 | 97.7 | 53.4 | 0.23 | 80.3 | 76.4 | **79.2** | 0.52 |
| April     | 54.6 | 89.8 | **75.0** | **0.46** | 45.6 | 94.7 | 74.1 | 0.43 | 18.0 | 98.1 | 65.5 | 0.18 | 59.3 | 83.9 | 73.6 | 0.44 |
| May       | 47.5 | 93.0 | 86.5 | **0.42** | 35.3 | 96.7 | **87.9** | 0.39 | 12.2 | 99.0 | 87.3 | 0.17 | 52.1 | 89.9 | 84.5 | 0.40 |
| June      | 46.3 | 96.2 | 94.6 | **0.32** | 23.6 | 98.9 | 96.6 | 0.28 | 7.6 | 99.5 | **97.0** | 0.11 | 51.3 | 94.3 | 93.0 | 0.28 |
| July      | 28.8 | 97.6 | 96.9 | **0.16** | 8.9 | 99.6 | 98.6 | 0.12 | 1.7 | 99.7 | **98.8** | 0.02 | 38.5 | 96.2 | 95.6 | 0.15 |
| August    | 34.0 | 97.0 | 96.0 | **0.20** | 14.0 | 99.4 | 98.1 | 0.18 | 1.5 | 99.7 | **98.3** | 0.02 | 44.7 | 95.3 | 94.4 | 0.19 |
| September | 43.0 | 93.9 | 89.3 | **0.36** | 24.4 | 97.7 | 91.1 | 0.29 | 6.6 | 99.5 | **91.8** | 0.10 | 51.8 | 90.6 | 87.1 | 0.35 |
| October   | 55.4 | 91.5 | **76.5** | **0.49** | 42.8 | 95.5 | 73.6 | 0.41 | 13.6 | 99.3 | 64.6 | 0.15 | 62.0 | 85.9 | 76.0 | **0.49** |
| November  | 72.2 | 93.0 | 77.8 | 0.54 | 67.4 | 94.1 | 74.6 | 0.49 | 27.9 | 99.2 | 47.4 | 0.17 | 82.3 | 79.6 | **81.6** | 0.57 |
| December  | 76.7 | 92.6 | 87.3 | 0.36 | 77.5 | 93.9 | 87.7 | 0.35 | 48.3 | 99.2 | 53.6 | 0.16 | 91.5 | 68.3 | **89.2** | 0.49 |
| Overall   | 71.0 | 94.3 | 84.1 | 0.67 | 68.1 | 97.1 | 84.3 | 0.67 | 40.0 | 99.2 | 73.8 | 0.42 | 81.8 | 89.6 | **86.2** | 0.72 |

Note. The largest O<sub>acc</sub> and κ for each sensor are indicated in bold font.

agreement for categorical items, was introduced here for classification evaluation (Landis & Koch, 1977). The value of (κ) is defined as

\[ κ = \frac{p_o - p_e}{1 - p_e}, \]  

where \( p_e \) and \( p_o \) are the probability of random agreement and disagreement, respectively, and can be calculated as

\[ p_e = \frac{(F_F + F_T) \times (F_T + T_T) + (F_T + T_F) \times (T_F + T_T)}{(F_F + F_T + T_F + T_T)^2}, \]

\[ p_o = \frac{F_F + T_T}{F_F + F_T + T_F + T_T}. \]

In addition to the evaluation metrics discussed above, a detailed evaluation was also conducted using the root mean squared error (RMSE) and mean bias (BIAS) of four indices: the number of freezing days (NFD), the onset and the offset of freezing, and the total duration of near-surface soil freezing. The onset (offset) of the near-surface soil freeze was defined as the earliest (latest) date after 1 July when the near-surface soil remained frozen for five successive days.

4. Results
4.1. Broad Evaluation

Our results showed a wide range of performance for the four algorithms with overall κ between 0.42 and 0.72 and an O<sub>acc</sub> between 73.8% and 86.2% (Table 3). DFA achieved the best performance overall with the

Table 4
Overall Performance of the Four Algorithms in Detecting the Number of Freezing Days (NFD), Onset, Offset, and Duration of Near-Surface Soil Frozen Status During 1979–2016

| Platform | SDIA | MSTA | DTA | DFA |
|----------|------|------|-----|-----|
|          | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ | F<sub>acc</sub> | T<sub>acc</sub> | O<sub>acc</sub> | κ |
| SMMR     | 59.0 | 95.5 | 78.5 | 0.56 | 56.5 | 98.0 | 78.6 | 0.56 | — | — | — | — | 88.5 | 84.0 | **86.1** | 0.72 |
| SSM/I    | 76.1 | 93.7 | 86.0 | 0.71 | 70.9 | 97.1 | 85.6 | 0.70 | 37.4 | 99.3 | 72.1 | 0.39 | 81.4 | 90.7 | **86.6** | 0.73 |
| SSMIS    | 72.3 | 94.3 | 85.1 | 0.68 | 72.9 | 96.5 | **86.6** | **0.72** | 40.7 | 99.1 | 74.6 | 0.43 | 77.0 | 91.8 | 85.6 | 0.70 |

Note. The T<sub>b</sub> (85V) required by DTA is not available for the sensor of SMMR. The largest O<sub>acc</sub> and κ for each sensor are indicated in bold font.
Figure 7. Annual changes of near-surface soil F/T detection algorithm performance. The results are not available in 1990 for DTA because of missing $T_B(85V)$ values in that year. The dashed orange lines represent a change in sensor, and the dashed gray line represents a change in the platform.

Highest $\kappa$ and $O_{acc}$ of 0.72 and 86.2%. While the SDIA, MSTA, and DFA show similar performance ($\kappa$ ranges from 0.68 to 0.73) for the SSM/I and SSMIS sensors, significant improvement is achieved for SDIA and MSTA compared to the SMMR sensor (Table 4 and Figure 7). The DTA showed the worst performance and had significantly lower $\kappa$ and $O_{acc}$ values than the other algorithms.

All four algorithms were better at detecting the thawed (from May to September) and frozen seasons (from November to February) than the transition season (April and October). As a consequence, the algorithms show a clear seasonal variation of overall performance with a mean $O_{acc}$ of 93.1 ± 4.7% in the thawed season and 74.4 ± 11.6% in the frozen season (Figure 8).
4.2. Detailed Evaluation

Not surprisingly, the sites with lower probabilities of being frozen normally have better performance due to the better detection of the thawed season compared to the transition season. Consequently, the performance generally decreases for more northerly sites. Most (92.6%) of the sites calculated using DFA have a $\kappa \geq 0.4$, and it decreases to between 38.8% and 82.1% for the other algorithms. The DTA only shows suitability for a few sites, most of which are northerly (latitude $> 30^\circ$ N) and have a relatively cold climate.

Because all algorithms underestimate the freeze status during the transition season, the NFD was consistently underestimated, with a BIAS of between $-8.1$ and $-95.3$ days.

SDIA, MSTA, and DFA showed remarkable error in the duration of frozen status, and the overall RMSE was between 51.8 and 57.8 days (Table 5). The pronounced uncertainty on the frozen duration is attributed to both the error in predicting onset, which has a BIAS of $-0.57$–$48.5$, and the error in predicting offset, which has a BIAS between $-42.7$ and $0$. The detection of NFD, onset, offset, and hence frozen duration for SDIA

| Table 5 |
|---|---|---|---|---|---|---|
| Overall Performance of the Four Algorithms in Detecting the Number of Freezing Days (NFD), Onset, Offset, and Duration of Near-Surface Soil Frozen Status During 1979–2016 |
| Index (days) | SDIA | MSTA | DTA | DFA |
| --- | --- | --- | --- | --- |
| NFD | 59.5 $-34.3$ | 52.8 $-38.8$ | 105.8 $-95.3$ | 52.8 $-8.1$ |
| Onset | 33.4 $10.2$ | 30.9 $18.5$ | 56.2 $48.5$ | 32.0 $-0.57$ |
| Offset | 31.3 $-8.5$ | 27.1 $-13.6$ | 53.5 $-42.7$ | 31.0 $0.00$ |
| Duration | 56.4 $-18.6$ | 51.8 $-32.0$ | 104.0 $-91.2$ | 57.8 $0.57$ |

Note. The smallest RMSE and BIAS for each index are indicated in bold font.
Figure 10. The range of bias in each algorithm for the number of frozen days (NFD), onset, offset, and duration of soil frozen for near-surface soil. Red dots correspond to mean values. The box edges mean the 25th and 75th percentile, while bar represent the 5% and 95% percentile, respectively.

and MSTA show significant differences between sensors with the SMMR having the worst performance. On the other hand, the DFA algorithm shows similar yet large bias for all the sensors (Figure 10).

5. Discussion

5.1. Inter-comparison of Algorithms

With the exception of DTA, all the algorithms were suitable for detecting near-surface soil F/T status at a large scale and had $O_{acc}$ around 85% and $\kappa$ between 0.67 and 0.72.

In general, DFA was the best-performing algorithm and had the highest overall agreement with observations. One of the reasons for this is the better performance of the SMMR sensor compared to the SDIA and MSTA sensors (Table 4 and Figure 7). There is, however, a temporary decline in accuracy from 2006 to 2009 associated with a peak in $T_b$. The $T_b(37V)$ and $T_b(19H)$ during this period were between 2.4–3.6 K and 1.9–3.1 K higher than 2010 values, respectively (Figure 11a). Based on the sensitivity test, DFA performance depends on $T_b$ (Figure 11b). This is because the parameters of DFA in equations (8) and (9) are derived from statistically based discriminant functions. In other words, better performance could only be expected in areas where the mean $T_b$ values are similar to the training data sets used by Zhao et al. (2011). Parameter calibration is therefore recommended before applying DFA to other regions.

SDIA and MSTA show similar performance with a remarkable improvement from the SMMR sensor to SSM/I and SSMIS. There are several explanations for this; the first is the difference in $T_b$ accuracy between sensors. Njoku (1980) reported that the $T_b$ error of the SMMR sensor was around 1–2.5 K, whereas the error for SSM/I and SSMIS was only 0.5 K (Brodzik & Long, 2018). Second, the narrow sensor swath contributes to the decreased performance of SMMR (Table 1) because interpolated grid cells are expected to perform...
Figure 11. (a) Temporal variability of DFA and (b) sensitivity of DFA performance with respect to $T_b(37V)$ and $T_b(19H)$. In (b), the origin (black point, $\kappa = 0.72$) corresponds to the DFA’s overall accuracy in Table 3 with the mean $T_b(37V)$ and $T_b(19H)$ of 260.6 and 251.4 K over the entire period, respectively. The red points correspond to the measurements in (a) between 2006 and 2009. The black line is a contour of $\kappa = 0.72$.

more poorly than measured ones even though a rigorous procedure was used to fill the $T_b$ gap (Figure A3). For example, the $\kappa$ for SMMR derived from MSTA using the original (no interpolation) $T_b$ was 0.61, but it decreased to 0.55 for the no-gap $T_b$. Since the data completeness for SMMR is 20–42% lower than the other platforms, worse performance is expected. Although SDIA and MSTA achieve similar performance in the four indices based on detailed evaluation, MSTA can be more easily applied in areas with limited observations because its model parameters are derived from globally available grid data rather than from observations.

Our results showed that DTA is less suitable for near-surface soil F/T status detection than the three other algorithms. In particular, the DTA significantly underestimated the freeze status with about 60% of observations incorrectly classified as thawed. This is very likely caused by an underestimated threshold for $T_b(37V)$, which was fit using a small data set comprising only eight stations (Jin et al., 2009). In contrast, the $T_{c_{\text{cutoff}}}$ for SDIA was about 259.7 ± 1.8 K for bare land, 261.4 ± 1.9 K for grass, 261.8 ± 1.6 K for shrub, and 261.6 ± 1.8 K for forests. These values are about 10 K higher than that used in DTA (252 K). This is also demonstrated by a sensitivity test (Figure 12). Figure 12 shows that the performance of DTA is highly sensitive to the values of selected $T_b(37V)$ cutoff. Based on the sensitivity test, DTA achieves its best performance in the $T_b(37V)$ range of 263–265 K. For this reason, more validation data are recommended before applying DTA to areas outside of the test sites in order to estimate more reliable parameters.
5.2. Uncertainties of Near-Surface Soil F/T Detection Algorithms

In order to take full advantage of our observations, we assumed that the near-surface soil was frozen when snow cover was present. However, when the $T_g$ is only slightly below 0 °C, the presence of liquid water under the snow will yield dielectric properties that resemble thawed conditions. This would decrease the accuracy of detection, especially during the snow-melt season.

The passive microwave data were found to be less sensitive to near-surface soil F/T status in the F/T transition season with worse performance with respect to freezing onset and offset detection. This is because near-surface soil freezing may occur sporadically for some length of time during the F/T transition season, and detection mismatching would result in greater uncertainty. The mismatch between observation depth and passive microwave detection depth will also contribute to poor performance in assessing freezing onset, offset, and NDF. The reference data set used as ground truth was derived from measurements at a depth of 0 cm whereas the detection depth of passive microwave sensors is expected to be on the order of several centimeters. The greater detection depth will lead to a delayed or lowered sensitivity of F/T detection in the transition seasons due to the time lag caused by heat diffusion. Although soil temperatures from multiple depths, such as 0 cm (Jin et al., 2015), 4 cm (Jin et al., 2009; Zhao et al., 2011), and 5 cm (Zhang et al., 2003, 2004), were used, the detection depth for passive microwave sensors is site-specific due to the strong spatial variation of fine-scale soil properties, such as moisture content, as well as surface characteristics, such as roughness and vegetation (Zheng et al., 2018). Recently, Zheng et al. (2019) reported that vertical $T_b$ has better agreement with the soil moisture measured at 2.5 cm than that measured at 5 cm based on in situ measurements. A truly systematic evaluation of these effects would require a comprehensively designed set of observations. We hope that, over time, more investigations of passive microwave detection depth in various soil conditions will help to establish correlations of trusted values with soil characteristics.

Although the calibrated $T_b$ data sets are used here to reduce the potential uncertainties caused by the changes of sensors, the mismatch of overpass time remains unsolved in CETB. While the overpass time of the SSM/I and SSMIS sensors (5–10 a.m.) is highly aligned with the minimum $T_g$ measured at 08:00, the overpass of the SMMR sensor is about midnight. This is thought to reduce the classification accuracy of SDIA for the SMMR sensor by weakening the relationship between $T_g$ and $T_b$.

5.3. Reliability of Change Detection

Passive microwave remote sensing data sets have been widely used for deriving long-term near-surface soil F/T changes. For example, Jin et al. (2015) reported that the frozen period has shortened by 20 days during the period 1978–2008 based on estimates of onset delay and offset advance. Similarly, a delay of ~14 days to the first day of soil freezing and an advance of ~10 days to the last day was found over the period 1978–2007.
on the Qinghai-Tibet Plateau (Li et al., 2012). However, our results suggest that the RMSE of onset and offset dates could be up to 41.6 and 35.8 days, respectively. This is the case even for the algorithm-sensor pair with the best performance. The uncertainty based on the calibrated \( T_b \) data set is about 1.8–6.2 times the magnitude of the estimated change over the last several decades.

We also evaluate the global FT data records, daily freeze/thaw Earth System Data Record (FT ESDR, https://nsidc.org/data/nsidc-0477), proposed by Kim et al. (2017) in order to avoid the possible bias induced by different \( T_b \) gap filling and aggregation (or resampling) methods. Similarly, FT ESDR is found to have a RMSE of about 32 and 26 days for the onset and offset, respectively, over China.

For this reason, we argue that remarkable uncertainty still exists in the detection of long-term near-surface soil F/T status, and the results of previous studies may not be reliable. Recently, a noisy-edge detection method has been successfully applied to investigate the near-surface soil F/T changes over the Alaska, and the results show a RMSE of 1 ± 7 (Chen et al., 2019). This provides a promising method for long-term near-surface soil F/T changes in the future as soon as it is calibrated more widely.

6. Conclusions

We conducted a detailed evaluation of near-surface soil F/T status detection based on passive microwave remote sensing data. Specifically, we evaluated and inter-compared the four most widely used algorithms (SDIA, MSTA, DTA, and DFA) using a large number of ground truth observations across China. Our results support the following conclusions:

- In general, the SDIA, MSTA, and DFA algorithms are suitable for near-surface soil F/T status investigations and have overall accuracies between 84.1% and 86.2% and \( \kappa \) of 0.67–0.72. However, the current parameters for DFA may limit its performance in areas with different \( T_b \) from the training data set. DTA is not transferable to the areas outside the test stations used for its development, and more validation is recommended before applying it to a large area.
- All these four algorithms show a reduced ability to capture the near-surface soil freeze status in the F/T transition seasons and exhibit significant uncertainties in onset, offset, and NFD.
- Investigations of long-term changes to near-surface soil F/T using coarse-grid passive microwave remote sensing data may be subject to remarkable uncertainty.

Appendix A

Figure A1. The weighting scheme used in the linear regression model for \( T_{\text{cutoff}} \) estimation. \( T_g \) is used for SDIA, while \( T_a \) is used for MSTA.
Figure A2. Comparison of $T_b$ from SSM/I as well as SSMIS and from AMSR-E during 02/06/2002–05/10/2011 based on all the original $T_b$ values (about $1.1 \times 10^6$ individual values). The solid line is the linear regression relationship, while the dashed line is the 1:1 reference line.

Figure A3. Examples of the two-step $T_b$ gap filling scheme. The blue numbers in the bottom right of the panels describe data completeness.
References

Bartsch, A., Kidd, R. A., Wagner, W., & Bartalis, Z. (2007). Temporal and spatial variability of the beginning and end of daily spring freeze/thaw cycles derived from scatterometer data. Remote Sensing of Environment, 106(3), 360–374. https://doi.org/10.1016/j.rse.2006.09.004

Brocki, L., Brenning, A., Gruber, S., & Noetl, J. (2012). A statistical approach to modelling permafrost distribution in the European Alps or similar mountain ranges. The Cryosphere, 6(1), 125–140. https://doi.org/10.5194/tc-6-125-2012

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H. (2012). EASE-Grid 2.0: Incremental but significant improvements for earth-gridded data sets. ISPRS International Journal of Geo-Information, 1(1), 32–45. https://doi.org/10.3390/ijgi1010032

Brodzik, M. J., & Long, D. G. (2018). NASA MEaSUREs Calibrated Passive Microwave Daily EASE-Grid 2.0 Brightness Temperature ESDR (CETB) Algorithm Theoretical Basis Document. NSIDC. Boulder, CO: MEaSUREs Project White Paper.

Cao, B., Gruber, S., & Zhang, T. (2017). REDCAPP (v1.0): Parameterizing valley inversions in air temperature data downscaled from reanalyses. Geoscientific Model Development, 10(8), 2905–2923. https://doi.org/10.5194/gmd-10-2905-2017

Cao, B., Gruber, S., Zhang, T., Li, L., Peng, X., Wang, K., et al. (2017). Spatial variability of active layer thickness detected by ground-penetrating radar in the Qilian Mountains, Western China. Journal of Geophysical Research: Earth Surface, 122, 574–591. https://doi.org/10.1002/2016JF004018

Cao, B., Quan, X., Brown, N., Stewart-Jones, E., & Gruber, S. (2019). GlobSim (v1.0): Deriving meteorological time series for point locations from multiple global reanalyses. Geoscientific Model Development, 12(11), 4661–4679. https://doi.org/10.5194/gmd-12-4661-2019

Cao, B., Zhang, T., Wu, Q., Sheng, Y., Zhao, L., & Zou, D. (2019). Permafrost zonation index map and statistics over the Qinghai-Tibet Plateau based on field evidence. Permafrost and Periglacial Processes, 30(3), 178–194. https://doi.org/10.1002/ppp.2006

Chao, S.-c. (1984). The sandy deserts and the gobi of China. In F. El-Baz (Ed.), Deserts and arid lands (pp. 95–113). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-6608-0_5

Chen, X., Liu, L., & Bartsch, A. (2019). Detecting soil freeze/thaw onsets in Alaska using SMAP and ASCAT data. Remote Sensing of Environment, 220(May 2018), 59–70. https://doi.org/10.1016/j.rse.2018.10.010

Dee, D. P., Uppala, S. M., Simmons, a. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597. https://doi.org/10.1002/qj.828

Feppaz, M., Williams, B. L., Edwards, A. C., Scallenbge, R., & Zanini, E. (2007). Simulating soil freeze/thaw cycles typical of winter alpine conditions: Implications for N and P availability. Applied Soil Ecology, 35(1), 247–255. https://doi.org/10.1016/j.apsoil.2006.03.012

Gao, L., Bernhardt, M., & Schulz, K. (2012). Elevation correction of ERA-Interim temperature data in complex terrain. Hydrology and Earth System Sciences, 16(12), 4661–4673. https://doi.org/10.5194/hess-16-4661-2012

Grody, N. C. (1991). Classification of snow cover and precipitation using the special sensor microwave imager. Journal of Geophysical Research, 96(D4), 7423. https://doi.org/10.1029/91JD00045

Guo, D., Li, D., & Hua, W. (2018). Quantifying air temperature evolution in the permafrost region from 1901 to 2014. International Journal of Climatology, 38(1), 66–76. https://doi.org/10.1002/joc.5161

Guo, D., Yang, M., & Wang, H. (2011). Characteristics of land surface heat and water exchange under different soil freeze/thaw conditions over the central Tibetan Plateau. Hydrological Processes, 25(16), 2531–2541. https://doi.org/10.1002/hyp.8025

Hu, T., Zhao, T., Shi, J., Wu, S., Liu, D., Qin, H., & Zhao, K. (2017). High-resolution mapping of freeze/thaw status in China via fusion of MODIS and AMSR2 data. Remote Sensing, 9(12), 1339. https://doi.org/10.3390/rs9121339

Jin, R., Li, X., & Che, T. (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

Jin, R., Zhang, T., Li, X., Yang, X., & Ren, Y. (2015). Mapping surface soil freeze-thaw cycles in China based on SSMR and SSM/I brightness temperature from 1978 to 2008. Arctic, Antarctic, and Alpine Research, 47(2), 322–329. https://doi.org/10.1657/AAAR00C-1304

Kim, Y., Kimball, J. S., Glassy, J., & Du, J. (2017). An extended global Earth system data record on daily landscape freeze-thaw status 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., Kimball, J. S., McDonald, K. C., & Glassy, J. (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., Kimball, J. S., Zhang, K., & McDonald, K. C. (2012). Satellite detection of increasing Northern Hemisphere non-frozen seasons from 1979 to 2008: Implications for regional vegetation growth. Remote Sensing of Environment, 121, 472–487. https://doi.org/10.1016/j.rse.2012.02.014

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data published by: International Biometric Society Stable. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310

Leverington, D. W., & Duguay, C. R. (1996). Evaluation of three supervised classifiers in mapping depth to late-summer frozen ground, Central Yukon Territory. Canadian Journal of Remote Sensing, 22(2), 163–174. https://doi.org/10.1007/BF0089992.1996.10874650

Li, X., Jin, R., Pan, X., Zhang, T., & Guo, J. (2012). Changes in the near-surface soil freeze-thaw cycle on the Qinghai-Tibetan Plateau. International Journal of Applied Earth Observation and Geoinformation, 17(1), 31–42. https://doi.org/10.1016/j.jag.2012.03.002

Ma, Y., Zhang, Y., Zubrzycki, S., Guo, Y., & Farhan, S. B. (2015). Hillslope-scale variability in seasonal frost depth and soil water content investigated by GPR on the southern margin of the sporadic permafrost zone on the Tibetan Plateau. Permafrost and Periglacial Processes, 26(4), 321–334. https://doi.org/10.1002/ppp.1844

McDonald, K. C., Kimball, J. S., Njoku, E., Zimmermann, R., & Zhao, M. (2004). Variability in springtime thaw in the terrestrial high latitudes: Monitoring a major control on the biospheric assimilation of atmospheric CO2 with spaceborne microwave remote sensing. Earth Interactions, 8(20), 1–23. https://doi.org/10.1175/1087-3562(2004)8<1:VISITET>2.0.CO;2

Michelsen, A., Larsen, K. S., & Jonasson, S. (2002). Repeated freeze/thaw events and their effects on biological processes in two arctic ecosystem types. Applied Soil Ecology, 21(3), 187–195.

Njoku, E. G. (1990). Antenna pattern correction procedures for the Scanning Multichannel Microwave Radiometer (SMMR). Boundary-Layer Meteorology, 18(1), 79–98. https://doi.org/10.1007/BF00171912

Peng, X., Frauenfeld, O. W., Cao, B., Wang, K., Wang, H., Su., et al. (2016). Response of changes in seasonal soil freeze/thaw state to climate change from 1950 to 2010 across China. Journal of Geophysical Research: Earth Surface, 121, 1984–2000. https://doi.org/10.1002/2015JF003876
Ran, Y. H., Li, X., Lu, L., & Li, Z. Y. (2012). Large-scale land cover mapping with the integration of multi-source information based on the Dempster\-Shafer theory. *International Journal of Geographical Information Science*, 26(1), 169–191. https://doi.org/10.1080/13658816.2011.577745

Sapiano, M. R. P., Berg, W. K., McKague, D. S., & Kummerow, C. D. (2013). Toward an intercalibrated fundamental climate data record of the SSM/I sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3), 1492–1503. https://doi.org/10.1109/TGRS.2012.2206601

Schimel, J. P., & Clein, J. S. (1996). Microbial response to freeze-thaw cycles in tundra and taiga soils. *Soil Biology and Biochemistry*, 28(8), 1061–1066. https://doi.org/10.1016/0038-0717(96)00083-1

Smith, N. V., Saatchi, S. S., & Randerson, J. T. (2004). Trends in high northern latitude soil freeze and thaw cycles from 1988 to 2002. *Journal of Geophysical Research D: Atmospheres*, 109(D12), 1–14. https://doi.org/10.1029/2003JD004472

Wang, K., Zhang, T., & Zhong, X. (2015). Changes in the timing and duration of the near-surface soil freeze/thaw status from 1956 to 2006 across China. *The Cryosphere*, 9(3), 1321–1331. https://doi.org/10.5194/tc-9-1321-2015

Xiao, X., Zhang, T., Zhong, X., Shao, W., & Li, X. (2018). Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data. *Remote Sensing of Environment*, 210(August 2017), 48–64. https://doi.org/10.1016/j.rse.2018.03.008

Xie, Y., Jin, R., & Yang, X. (2013). Algorithm development of monitoring daily near surface freeze/thaw cycles using AMSR-E brightness temperature (in Chinese). *Remote Sensing Technology and Application*, 28(2), 182. https://doi.org/10.11873/j.jissn.1004-0323.2013.2.182

Zhang, T., Armstrong, R. L., & Smith, J. (2003). Investigation of the near-surface soil freeze-thaw cycle in the contiguous United States: Algorithm development and validation. *Journal of Geophysical Research*, 108(D22), 2003JD003530. https://doi.org/10.1029/2003JD003530

Zhao, T., Barry, R. G., & Armstrong, R. L. (2004). Application of satellite remote sensing techniques to frozen ground studies. *Polar Geography*, 28(3), 163–196. https://doi.org/10.1080/0789610186

Zhao, T., Shi, J., Hu, T., Zhao, L., Zou, D., Wang, T., et al. (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., Zhang, L., Jiang, L., Zhao, S., Chai, L., & Jin, R. (2011). A new soil freeze/thaw discriminant algorithm using AMSR-E passive microwave imagery. *Hydrological Processes*, 25(11), 1704–1716. https://doi.org/10.1002/hyp.7930

Zheng, J., RoyChowdhury, T., Yang, Z., Gu, B., Wullschleger, S. D., & Graham, D. E. (2018). Impacts of temperature and soil characteristics on methane production and oxidation in Arctic tundra. *Biogeosciences*, 15(21), 6621–6635. https://doi.org/10.5194/bg-15-6621-2018

Zheng, J., Thornton, P. E., Painter, S. L., Gu, B., Wullschleger, S. D., & Graham, D. E. (2019). Modeling anaerobic soil organic carbon decomposition in Arctic polygon tundra: Insights into soil geochemical influences on carbon mineralization. *Biogeosciences*, 16(3), 663–680. https://doi.org/10.5194/bg-16-663-2019