A DECISION TREE ALGORITHM FOR FREEZE/THAW CLASSIFICATION
OF SURFACE SOIL USING SSM/I

Rui Jin and Xin Li

Cold and Arid Regions Environmental and Engineering Research Institute,
Chinese Academy of Sciences
Lanzhou 730000, China
E-mail: jinrui@lzb.ac.cn

ABSTRACT
A decision tree algorithm was developed to classify the
surface soil frozen/thawed status based on the microwave
emission and scattering characteristics of the frozen/thawed
soil, and the cluster analysis of samples from the frozen soil,
the thawed soil, the desert and the snow. The decision tree
works with five SSM/I channels (19V, 19H, 22V, 37V, 85V)
and three dominant indices, including the scattering index,
the 37GHz vertical polarization brightness temperature and
the 19GHz polarization difference. The algorithm also takes
the scattering effect of the desert and the precipitation into
consideration. The classification results validated by the
4cm-depth soil temperature and the Map of the
Geocryological Regionalization and Classification in China
showed satisfied accuracy.

Index Terms—freeze/thaw, SSM/I, decision tree

1. INTRODUCTION
There is about 50×10^6 km^2 land surface would experience
transition between the frozen and the thawed status annually
on earth[1]. The soil freeze/thaw status has an important
effect on the energy and mass exchange between land and
atmosphere, the surface runoff, the carbon cycle and so on
[2]. The timing, duration and areal extent of the near-surface
soil freeze/thaw status can be taken as the indicator for
climate change because of its sensitivity to environmental
temperature [3].

Many literatures about the freeze/thaw classification
were published during the 1980s ~1990s. The frozen soil
has different microwave emission and scattering characteristics from the thawed soil: ①lower thermodynamic temperature ②higher emissivity, ③stronger volume scatter darkening, namely the brightness
temperature decrease with increased frequency. These
characteristics can be used to develop proper indices for
identifying the soil freeze/thaw status. The dual indices
algorithms including 37GHz brightness temperature and
spectral gradient were widely used, which based on the
characteristics of lower brightness temperature and volume
scatter darkening effect of the frozen soil [3~5]. However,
the desert, the precipitation and the snow are often
misclassified into the frozen soil because they have similar
volumetric scattering characteristics in the cold environment.
Therefore, it’s necessary to exclude these disturbed land
types when implementing the large-scale soil frozen/thawed
classification. At the same time, the decision tree algorithms
had developed to identify the global snow cover and the
precipitation [6]. However, the frozen soil type was not
being considered in the classification.
A new decision tree algorithm using SSM/I brightness
temperature has been developed based on the microwave
emission and scattering characteristics of the frozen/thawed
soil. The long time series of the frozen/thawed soil dataset
can be obtained using this decision tree, which can provide
basic soil information for research on the climate and the
cryosphere interaction, the carbon cycle, the hydrological
process and the land surface process et al [7, 8]. The
decision tree can also be used as the pre-classification
process before land surface parameter retrieval using remote
sensing data.

2. METHODS
The resolution of passive microwave remote sensing is very
course, how to select “pure” pixels are of critical importance
to decide the thresholds in the decision tree. We have
selected four types of samples including the frozen soil, the
thawed soil, the desert and the snow with the aid of
reference data including MODIS snow cover product, the
Map of Geocryological Regionalization and Classification in China, Chinese Land Use Map, and 4cm-deep soil
temperatures observed in situ. The criteria for selecting
samples are as follows.
(1) frozen soil: 4cm-deep soil temperature is less than 0°C
and is distributed in the seasonally frozen ground and
permafrost regions.
(2) thawed soil: 4cm-deep soil temperature is larger than 0°C and is not in the seasonally frozen ground and permafrost regions.

(3) desert: in the hinterland of desert.

(4) snow: 75% area of the SSM/I pixel is covered by the snow. The snow coverage is obtained by the upscaling of MODIS snow product.

The cluster analysis is done based on three indices: ① Scattering Index (SI) [6], it is used to distinguish the strong scatterer and non-scatterer or the absorber. ② 37GHz vertical polarization brightness temperature T37V is used due to its high correlation with the soil temperature. ③ 19GHz Polarization Difference (PD= T19V – T19H) is used to reflect the surface roughness. The mean value and the 2 times standard deviation of each index for each type show the discriminability (Fig. 1).

The thresholds in the decision tree are determined as follows (Fig. 2): ① The PD of the desert is 36.28±2.22 (mean±2×standard deviation), which is obviously larger than other types because of the near-spectral surface of desert. Threshold of PD>30 can identify most of desert, and the remainder can be further identified in the sub-branches of the decision tree by using PD>25.② Both the frozen soil and the snow are strong scatterers with higher SI, but the SI of the thawed soil is smaller (4.76±2×5.51) because of its absorbing characteristics. So the threshold SI>10 can be used to separate stronger scatterers and non-scatters. ③ The T37V of the frozen soil is 232.57±2×9.40, while the thawed soil is 259.1±2×5.33. The threshold of T37V<252 is used to separate the cold and the warm soil. ④ The Grody’s method, 165+0.49×T85V<=T22V, is adopted to identify the deep convective precipitation with the ice particle and the strong scatter effect, and the rule 254<=T22V<=258 and SI<=2 is applied to identify the general precipitation [9]. Furthermore, the index T85V/T19V <0.9 is used to identify the hail clouds and rainstorm.

The snow has similar microwave characteristics as the frozen soil because of its low temperature and dielectric constant [10]. The SI of the snow with depth more than 10 cm (calculated by $SD = 0.66(T_{19H} - T_{19V})$, [11]) is above 30, sometimes even reaches 80 due to stronger volume scattering at 85GHz. However, the SI of shallow snow is generally between 0 and 20, which is the same as that of the frozen soil. Since the snow cover in China is often patchily distributed and shallow, it is difficult to play much role as a heat insulator for the underlying soil. It can be considered that the soil under the snow cover is frozen at most time [12]. Therefore, we do not further distinguish the frozen soil and snow in the decision tree.

3. VALIDATION

First step, the results were validated by the 4cm-deep soil temperature observations obtained from the CEOP (Coordinated Energy and Water Cycle Observations Project). The total number of in-situ observations is 1695 and the number of misclassification is 219, so the classification accuracy reaches 87%. Among the 219 misclassification, 18 are for desert. This case mainly happens in the regions with flat and dry surface, and when the surface temperature is high during June to September. If the desert is masked out based on a land use map, this kind of misclassification can be avoided. Analyzed from the viewpoint of error distribution with 4-cm temperature, 40% misclassifications are in the range of -0.5°C~0.5°C and 73% in -2.0°C ~2.0°C, say, near the freezing point. From the viewpoint of error distribution with time, 33% of misclassifications are in April and May and 38% in September and October, i.e. the transition period between the cold and the warm seasons. These misclassifications are
partially due to the coarse spatial resolution of SSM/I. Because of the heterogeneity within the pixel, the in situ and pixel observations are somewhat not comparable.

Second step, a map-to-map validation is done. A new frozen soil map is created from the daily freeze/thaw results obtained above and the reference data is the Map of Geocryological Regionalization and Classification in China [13]. Only the pixels which are frozen > 15 days are used in the comparison. If a pixel is frozen < 15 days, it will be considered as intermittent frozen soil and is not comparable with the permafrost or seasonally frozen soil area in the reference map. The results show that the total classification accuracy is 91.66% and the Kappa coefficient is 80.5%. Additionally, the south boundary of freeze/thaw in the new map is matched well with the south limit of the seasonally frozen soil in the reference map.

Fig. 2 Flow chart of decision tree for soil freeze/thaw classification

4. CONCLUSION

A decision tree algorithm was developed to identify the surface soil freeze/thaw status with an accuracy of 87%. The algorithm can be extended for the SMMR (1978–1987) and AMSR-E (2002 till now) data to obtain the long time series of the soil freeze/thaw dataset and to provide an important dataset for the climate change, the agricultural production, the hydrologic cycle and carbon cycle researches.

5. ACKNOWLEDGEMENTS

This work is supported by NSFC (National Natural Science Foundation of China) project (grant number: 40701113), the CAS (Chinese Academy of Sciences) Action Plan for West Development Project (grant number: KZCX2-XB2-09) and Chinese Meteorological Administration project (GYHY(QX)2007-6-18).

6. REFERENCES