Science of the Total Environment 649 (2019) 515–525

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Data-driven mapping of the spatial distribution and potential changes of frozen ground over the Tibetan Plateau

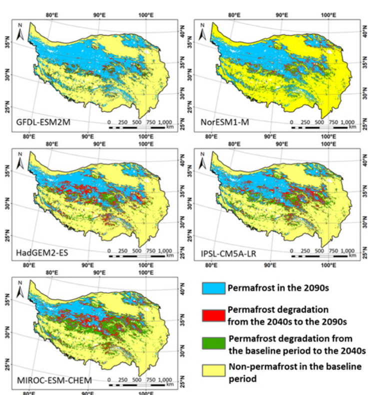
Taihua Wang, Dawen Yang *, Beijing Fang, Wencong Yang, Yue Qin, Yuhan Wang

State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China

HIGHLIGHTS

- Statistics and machine learning methods are compared and applied on frozen ground estimation of the Tibetan Plateau.
- Permafrost and maximum thickness of seasonally frozen ground distribution are mapped over the Tibetan Plateau.
- Around 26% and 44% of the current permafrost is projected to disappear by the 2040s and 2090s under RCP 4.5 scenario.
- Decreases in maximum seasonal frozen depth are larger at higher elevation compared with the decreases at lower elevation.

GRAPHICAL ABSTRACT



Permafrost degradation processes projected by 5 General Circulation Models under Representative Concentration Pathway 4.5

ARTICLE INFO

Article history:
Received 21 June 2018
Received in revised form 24 August 2018
Accepted 26 August 2018
Available online 28 August 2018

Editor: Jay Gan

Keywords:
Machine learning
Tibetan Plateau
Permafrost degradation
Maximum thickness of seasonally frozen ground
Climate change

ABSTRACT

Frozen ground degradation profoundly impacts the hydrology, ecology and human society on the Tibetan Plateau (TP) and the downstream regions. The spatial distribution and potential changes of permafrost and maximum thickness of seasonally frozen ground (MTSFG) on the TP is of great importance and needs more in-depth studies. This study maps the permafrost and MTSFG distribution in the baseline period (2003−2010) and in the future (2040s and 2090s) with 1-km resolution. Logistic regression (LR), support vector machine (SVM) and random forest (RF) are validated using 106 borehole observations and proved to be applicable in estimating permafrost distribution. According to the majority voting results of the three algorithms, 45.9% area of the TP is underlain by permafrost in the baseline period, and respectively 25.9% and 43.9% of the current permafrost will disappear by the 2040s and the 2090s projected by mean of the projections from the five General Circulation Models under the Representative Concentration Pathway 4.5 scenario. SVM performs better in spatial generalization than RF based on the results of nested cross validation. According to the MTSFG results derived from SVM, the most dramatic decrease in MTSFG will occur in the southwestern TP, which is projected to exceed 50 cm in the 2090s compared with the baseline period. This study introduces the statistics and machine learning algorithms to frozen ground estimation on the TP, and the high resolution permafrost and MTSFG maps produced by this study can provide useful information for future studies on the third pole region.

* Corresponding author.
E-mail address: yangdw@tsinghua.edu.cn (D. Yang).

https://doi.org/10.1016/j.scitotenv.2018.08.369

1. Introduction

Permafrost is defined as the sub-surface material having a temperature of less or equal to 0 °C during at least two consecutive years (Gruber, 2012). The Tibetan Plateau (TP), also known as the third pole of the globe, has the largest area of alpine permafrost in the world (Qiu, 2008; Immerzeel et al., 2010; Cuo et al., 2015). As climate warms, the active layer thickness increases and the permafrost disappearance have been reported on the TP (Li et al., 2008; Wu et al., 2015; Luo et al., 2016). In addition to permafrost, the regions with absence of permafrost include seasonally frozen ground region which freezes and thaws annually, as well as unfrozen ground which never freezes. The maximum thickness of seasonally frozen ground (MTSFG) decreased continuously in the past decades both on the TP and in other regions of the world, which reflects the comprehensive effects of changes in temperature, soil moisture conditions and site-specific properties and can affect the decomposition of soil organic matter and greenhouse gas exchanges between the land surface and the atmosphere (Frauenfeld and Zhang, 2011; Peng et al., 2017).

Frozen ground can influence the eco-hydrological processes on the TP via runoff generation, groundwater-surface water interaction, soil moisture conditions as well as other means (Cheng and Wu, 2007; Jin et al., 2009; Yang et al., 2010; Wang et al., 2018b). Understanding the spatial pattern of frozen ground provides the basis for quantitatively assessing the eco-hydrological processes on the TP, which calls for more reliable frozen ground maps at finer resolution. In a changing climate, frozen ground changes can greatly affect the regional hydrology in the TP and water resources security in the downstream, where more than 1.4 billion people feed on the rivers originating from the TP (Immerzeel et al., 2010). The projections of future frozen ground changes can enable policy makers to develop appropriate strategies prior to the anticipated changes (Guo and Wang, 2016). Although there are already studies projecting global permafrost changes in the future (Slater and Lawrence, 2013; Guo and Wang, 2016), very limited studies focus on the TP (Lu et al., 2017) and the spatial resolution is not high enough for hydrological application. Therefore, more studies are needed to project the future frozen ground changes on the TP, including both permafrost and MTSFG changes at finer resolution.

Borehole observations provide solid information for mapping frozen ground distribution. However, they are hard to obtain, very expensive, and restricted to the point scale (Zhang et al., 2005; Frauenfeld and Zhang, 2011; Wu et al., 2015). Consequently, simple empirical models are developed to map the large scale permafrost and MTSFG distribution on the TP (Zhao et al., 2017). For example, Li and Cheng (1999) developed an elevation model to map permafrost distribution by determining the lower limit of permafrost; Nan et al. (2002) established a mean annual ground temperature (MAGT) model to determine the permafrost across the TP. These models, however, are established based on limited and unevenly distributed in-situ observations, and they are derived from the static variables which are hard to be used in future projections.

Meanwhile, semi-empirical models with simplified physical processes applicable to high-latitude permafrost were used for the frozen ground simulation on the TP, including the surface frost number model, temperature at top of permafrost (TTOP) model, Stefan model (Zou et al., 2017), etc. However, if the underlying assumptions of the models are not met, their spatial generalization ability needs to be carefully assessed (Walvoord and Kurylyk, 2016). Process-based models have also been applied to simulate frozen ground on the TP at larger scale, which considered both the thermal and hydrological processes, including CLM4 model (Guo et al., 2012), GIPL2 model (Qin et al., 2017a), GBEHM model (Qin et al., 2017b; Gao et al., 2018), Noah model (Wu et al., 2018), etc. Nevertheless, the process-based models still have limitations in describing the complicated cryospheric and eco-hydrological processes using explicit equations due to limited understanding of the cold region environment. They often make simplifications to certain physical processes, introduce plenty of parameters, and are usually computational intensive (Walvoord and Kurylyk, 2016).

Although in-situ observations like boreholes and testing pits are still rare on the TP, nowadays more and more spatial datasets including topography, temperature and soil properties are available with the assistance of new technologies like remote sensing. This makes the data-driven methods including statistics and machine learning methods potentially applicable in estimating the frozen ground on the TP (Li et al., 2011; Lary et al., 2016). The statistics and machine learning algorithms can be promising when the relationships between the explained and explanatory variables are hard to explicitly described, but they might also encounter problems like overfitting and underfitting (Domingos, 2012), and the performances of different algorithms might differ for a given problem hence need to be evaluated. The statistics and machine learning methods have been adopted in many fields of geoscience, and have also been introduced into cryospheric researches recently. Deluigi et al. (2017) used different statistics and machine learning algorithms to map the potential permafrost distribution in Western Swiss Alps. Shi et al. (2018) used the decision tree method to get the 1-km permafrost map on the TP. The performance of different categories of statistics and machine learning algorithms in frozen ground estimation on the TP, however, have never been evaluated and compared. Besides, these methods have not yet been applied to map the MTSFG distribution as well as the future frozen ground changes on the TP, which are important for the third pole region.

This study adopts statistics and machine learning algorithms to map the current and future frozen ground distribution, including permafrost and MTSFG on the TP. The objectives of this study are to (1) evaluate and compare the performances of different statistics and machine learning methods in frozen ground estimation on the TP; (2) generate high-resolution permafrost and MTSFG distribution maps on the TP; (3) project the response of frozen ground to the future climate change on the TP.

2. Methods

In this study, statistics and machine learning methods are adopted for two objectives: first, classification for mapping permafrost distribution, more specifically, whether permafrost exists in a certain grid on the Tibetan Plateau (TP); second, regression for predicting MTSFG distribution in non-permafrost region. This study selects logistic regression (LR), support vector machine (SVM) and random forest (RF) for classification. These algorithms are selected because they belong to three specific sub-domains of learning techniques, respectively linear parametric learning, non-parametric learning and ensemble learning (Deluigi et al., 2017). LR, despite its name, is only applicable for classification. Therefore, only SVM and RF are adopted and compared for regression. The three selected algorithms are introduced in supplementary material (see Text S1). Detailed procedures for addressing the two problems will be introduced in Sections 2.1 and 2.2.

2.1. Mapping the permafrost distribution

The entire Tibetan Plateau (TP) is classified into two frozen ground types: permafrost and non-permafrost regions in this study. All estimates exclude glaciers and lakes within the TP boundary. Different types of permafrost, i.e., continuous and discontinuous permafrost, are not distinguished. Seasonally frozen ground and unfrozen ground are both regarded as non-permafrost regions. The whole procedure for establishing the classification models and mapping the current and future permafrost distribution is illustrated as Fig. 1. Six variables that affect the thermal state of frozen ground are selected as the input for training the models: the mean annual air temperature (MAAT), the ratio of mean annual ground surface temperature to mean annual air