Simulation of the present and future projection of permafrost on the Qinghai-Tibet Plateau with statistical and machine learning models

Jie Ni [1, 2], Tonghua Wu [1,3], Xiaofan Zhu [1], Guojie Hu [1], Defu Zou [1], Xiaodong Wu [1], Ren Li [1], Changwei Xie [1], Yongping Qiao [1], Qiangqiang Pang[1], Junming Hao [1, 2, 4] and Cheng Yang [1, 2]

[1] Cryosphere Research Station on the Qinghai-Tibet Plateau, State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, Gansu 730000, China; [2] University of Chinese Academy of Sciences, Beijing 100049, China; [3] Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou 511458, China; [4] School of Civil Engineering, Lanzhou University of Technology, Lanzhou, 730050, China.

Correspondence to:

Tonghua Wu (thuawu@lzb.ac.cn)

Key Points:

- The combined statistical method with machine learning is efficient to obtain the thermal regime of permafrost on the QTP.
- The present permafrost area on the QTP is ~1.04 × 10^6 km^2, and the average MAGT and ALT are -1.35 ± 0.42°C and 2.3 ± 0.60 m, respectively.
- The future changes of permafrost are projected to be pronounced due to climate change, but region-specific.

Abstract

The comprehensive understanding of the occurred changes of permafrost, including the changes of mean annual ground temperature (MAGT) and active layer thickness (ALT), on the Qinghai-Tibet Plateau (QTP) is critical to project permafrost changes due to climate change. Here, we use statistical and machine learning (ML) modeling approaches to simulate the present and future changes of MAGT and ALT in the permafrost regions of the QTP. The results show that the combination of statistical and ML method is reliable to simulate the MAGT and ALT, with the root-mean-square error of 0.53°C and 0.69 m for the MAGT and ALT, respectively. The results show that the present (2000–2015) permafrost area on the QTP is 1.04×10^6 km^2 ($0.80–1.28 \times 10^6$ km^2), and the average MAGT and ALT are -1.35 ± 0.42°C and 2.3 ± 0.60 m, respectively. According to the classification system of permafrost stability, 37.3% of the QTP permafrost is suffering from the risk of disappearance. In the future (2061–2080), the near-surface permafrost area will shrink significantly under different Representative Concentration Pathway scenarios (RCPs). It is predicted that the permafrost area will be reduced to 42% of the present area under RCP8.5. Overall, the future changes of MAGT and ALT are pronounced and region-specific. As a result, the combined statistical method with ML requires less parameters and input variables for simulation permafrost thermal regimes and could present an efficient way to figure out the response of permafrost to climatic changes on the QTP.

Keywords: permafrost; mean annual ground temperature; active layer; climate change; Qinghai-Tibet Plateau

1. Introduction

Frozen ground is an important component of the cryosphere, which exerts strong influences on regional ecology, hydrology and infrastructure engineering (Westermann et al., 2015; Wang et al., 2018a). The Qinghai-Tibet Plateau (QTP) is underlain by typical high-altitude permafrost region, which is undergoing more dramatic climatic warming than its surrounding regions (Wang et al., 2019a). A growing number of studies have reported the present status and predicted degradation of permafrost under various global warming scenarios (Pang et al., 2010, 2012; Zhang and Wu, 2012a; Guo and Wang, 2017; Xu et al., 2017a; Wang et al., 2018a). The degradation of permafrost may trigger the release of organic carbon into the atmosphere (Cheng and Wu 2007; Wu et al., 2017a; Chang et al., 2018; Wang et al., 2018b; Ran et al., 2018). It is also a potential threat to engineering construction and maintenance. However, most of these studies are based on linear statistical models and equilibrium models, and mainly focused on identifying the extent of permafrost, while researches on the present and future change of ground thermal regimes (including: the mean annual ground temperature, MAGT, and the active layer thickness, ALT) are relatively rare (Zhang et al., 2012a; Wang et al., 2019a). The changes of MAGT and ALT could affect the ecosystem of the QTP by altering the ground ice evolution, hydrological processes, vegetation dynamics and carbon cycling, etc. (Yang et al., 2010a; Wu et al., 2016; Niu et al., 2019; Hu et al., 2020). Therefore, it is of great importance to investigate present and future changes of the MAGT and ALT in the permafrost region (Qin et al., 2017; Zhang et al., 2018).

Permafrost is a thermally-defined subsurface phenomenon (Westermann et al 2015). Satellite sensors could obtain limited surface information, and only portion of the microwave remote sensing could penetrate several centimetres underground (Zhao et al., 2011; Michaelides et al., 2018; Qu et al., 2019). In general, it is difficult to use remote sensing to directly obtain information on changes in the physical state of permafrost (Yang et al., 2019). The current research on permafrost thermal regime is mostly focus on either *in situ* observing or modeling using atmospheric circulation models (Westermann et al., 2015). Most of the existing modeling frameworks require ground-based measurements as model inputs, while the *in situ* observations of permafrost are relatively sparse and highly non-uniform in cold regions. The long-term and continuous *in situ* observation sites for permafrost on the QTP are mostly located along the Qinghai-Tibet Highway and Railway, and other regions are less well distributed (Hu et al., 2015; Qin et al., 2017; Zheng et al., 2019). The absence of observation data would greatly weakens the accuracy of simulation results. Therefore, it is challenging to select reliable modeling approaches with limited data to obtain the occurrence of permafrost and its projection due to climate change.

At present, the simulation studies on the ALT and soil thermal state of the QTP fall into two categories, including equilibrium models and mechanistic transient models. (Riseborough et al., 2008; Qin et al., 2017; Aalto et al., 2018). The most commonly used equilibrium models include Stefan formula (Zhang and Wu 2012a; Xu et al., 2017a), Kudryavtsev formula (Pang et al., 2009; Wang et al., 2020a), the N factor (Nan et al., 2012), and the Temperature at the Top of the Permafrost model (TTOP) (Zou et al., 2017). The form of the equilibrium model is relatively simple and requires fewer driving data for input (Riseborough et al., 2008; Pang et al., 2009). However, this type of model tend to show poor portability. In contrast, mechanistic transient models consider more details of the hydrothermal exchange processes between the atmosphere and ground. Examples of this model include the Community Land Model (CLM; Oleson et al., 2010; Fang et al., 2016; Chen et al., 2017), Noah (Gao et al., 2015; Chen et al., 2015), the Geomorphology-based Eco-hydrological Model (GBEHM; Zheng et al., 2019), the SHAW model (Guo et al., 2011; Liu et al., 2013), and the CoupModel (Zhang et al., 2012b; Hu et al., 2013). Nevertheless, the processes of these models are complex and often insufficiently account for the hydrothermal dynamics, with the understanding of the soil physical mechanisms increase, the parameterization processes will become more complex (Harris et al., 2009; Hu et al., 2015; Guo and Wang, 2016). In addition to the transient models mentioned above, in recent years, the fine-scale tightly coupled hydro-thermal modeling of permafrost has also made great progress (e.g., models like ATS, Jafarov et al., 2018; and SUTRA, Walvoord et al., 2019, etc.), These models are typically based on a multidimensional solution to address fully coupled surface/subsurface permafrost thermal hydrology, which have played an important role to study the permafrost of local scale and microtopography (Painter et al., 2016).

Physics-based mechanistic models are currently the popular methods to study the permafrost, and the simulation results can show high accuracy. However, even with significant improvements in computer technology and algorithm simulation (Westermann et al., 2016), the current modeling still exists a trade-off between modeling resolution and size of the geographical domain (Etzelmüller, 2013). Especially in the case of lack of data and insufficient computing resources, the extensive application of physics-based mechanistic models would be limited. Whereas, the combined statistical method with machine learning (ML) can make up these deficiencies. In recent years, their great power in permafrost modeling has been confirmed (Xu et al., 2017b; Chadburn et al., 2017; Aalto et al., 2018). The main purpose of statistical and ML model is to identify the relationship between a dependent variable and one or more explanatory variables (Wheeler et al., 2013). They can easily explain environmental conditions related to topography and land cover, whereas these factors may be difficult to express with physical parameters (Etzelmüller, 2013). Due to the good coupling between air temperature (often characterized by mean annual air temperature or cumulative temperature sums) and ground thermal regime (Chadburn et al., 2017; Aalto et al., 2018), the subsurface (<10–20 m) soil thermal conditions respond well to climate change at the decadal scale (Aalto et al., 2018). In addition, precipitation type (e.g., snow, rain and sleet) and local environmental predictors (e.g., topography, underlying surface condition and soil texture condition) have great impacts on soil hydrothermal dynamics and the surface radiation budget (Lee et al., 2013; Zhu et al., 2019).

Therefore, in this study, we employed statistical and ML methods to investigate the MAGT and ALT across the QTP. The objective is to verify the applicability of the combined method on the QTP and quantitatively assess the present and future status of QTP permafrost. Firstly, we identified the critical factors which determining the occurrence of permafrost. Secondly, we used the combined modeling approaches integrated with field observation data, meteorological data and geospatial environmental predictors to calculate the present MAGT and ALT. Thirdly, the present results were benchmarked against *in situ* measurements of ALT and ground temperatures. Finally, the optimal modeling framework was used to predict future MAGT and ALT forced by different RCPs. The projection of the MAGT and ALT can serve as a useful reference and provide important information for the study of climate change, hydrology, ecology, and geohazards resulted from permafrost degradation on the QTP.

2. Data and Methods

2.1. Data sources

1) Ground temperature data

The MAGT is an important factor that reflects the thermal state of permafrost, and is defined as the ground temperature at the zero annual amplitude depth (ZAA, i.e., the depth at which the annual temperature variation < 0.1°C) (Qin, 2016). Due to the harsh environment of the QTP, some boreholes are measured manually using a multimeter once each year (Qin et al., 2017). Most MAGTs, however, are not easily accessible from the ZAA. In these cases, the temperature at or closest to 10 m below the ground surface was used (Nan et al., 2002; Liu et al., 2017). All disturbed measurement sites (e.g., sites submerged by the rising waters of a lake) were removed. Ultimately, 84 MAGT sites (Figure 1) were selected from both field station observations (Cryosphere Research Station on the Qinghai-Tibet Plateau, Chinese Academy of Sciences, available at http://www.crs.ac.cn/) and the related literatures (Wu et al., 2012a; Qin et al., 2017; Wang et al., 2017). We selected the period from 2000 to 2015 as the reference period, and all observations obtained were during this period. Some sites were based on one year of observation, while others were based on the average of several years, from which we calculated the long-term average value.

2) Active layer thickness data

In order to better fit the ALT, we attempted to collect a large amount of observed data from relevant literatures (Wu et al., 2012a; Qin et al., 2017; Wang et al., 2017). An additional portion of the active layer data came from field pit detection. A total of 77 ALT observation sites (Figure 1) were selected. The time node selection and disturbance data processing for ALT were the same as those used for the MAGT. Based on the distribution of MAGT and ALT observation sites, we divided them into five typical regions, the Wenquan typical region (WQIR), Xikunlun typical region (XKLIR), Gaize typical region (GZIR), Aerjin typical region (AEJIR) and Qinghai-Tibet Highway typical region (G109IR), which represent the permafrost regions of the eastern, western, southern, northern and central areas of the QTP, respectively.

3) Meteorological data

In order to obtain climate data for the reference periods (2000–2015), the China Meteorological Forcing Dataset (CMFD) (available at http://www.tpedatabase.cn/; Yang et al., 2010b; Yang et al., 2010b; He et al., 2020) with temporal and spatial resolutions of 3 hours and 0.1° × 0.1°, respectively, was utilized in this study. The time scale of the dataset covered the studying period. According to the study of He et al. (2020), the CMFD was established by merging Princeton reanalysis data, GLDAS data, GEWEX-SRB radiation data, and TRMM precipitation data, as well as the regular meteorological observations made by the China Meteorological Administration. The accuracy of CMFD is between the observation data and the remote sensing data (Yang et al., 2010b), and it has been widely used due to its high reliability (Xue et al., 2013; Xu et al., 2017a; Wang et al., 2019a).

In the study, we used air temperature and precipitation data from the CMFD to calculate the two key predictors, including the thawing indices (thawing degree days, TDD) and the freezing indices (freezing degree days, FDD), which play essential roles in the studies of the frozen ground. As useful indicators, they have been widely applied in the permafrost region to predict the ALT (Zhang et al., 2005; Nelson et al., 1997; Peng et al., 2018; Shiklomanov and Nelson, 2002) and permafrost distribution (Nelson and Outcalt, 1987). In addition, we also calculated the other two predictors, including the solid precipitation (i.e., precipitation with a temperature below 0°C, Sol_pre), and liquid precipitation (i.e., precipitation with a temperature above 0°C, Liq_pre).

For future conditions, the BCC-CSM 1.1 climate change modeling data was used (available at http://www.worldclim.org/). It was downscaled GCMs data from CMIP5 (IPCC Fifth Assessment). BCC-CSM1.1 is the version 1.1 of the Beijing Climate Center Climate System Model, the coupling was realized using the flux coupler version 5 developed by the National Center for Atmosphere Research (NCAR) (Wu et al., 2019). It was a fully coupled model with ocean, land surface, atmosphere, and sea-ice components, and was often used to simulate the response of global climate to rising greenhouse gas concentrations, the performance is satisfactory in China (Zhang and Wu, 2012b; Xin et al., 2018). In this study, we chose the monthly average air temperature and precipitation over the time period 2061–2080 under three Representative Concentration Pathways (RCPs): RCP2.6, RCP4.5, and RCP8.5 (Moss et al., 2010; Taylor et al 2012). The four predictors (TDD, FDD, Sol_pre, and Liq_pre) were recalculated in the same way for each time period and RCP scenario.

4) Geospatial environmental predictors

The geospatial environmental predictors were mainly derived from topographic data and regional environmental data. The Shuttle Radar Topography Mission (SRTM) data for a 1-km spatial resolution digital elevation model (DEM) (Reuter et al., 2007) were selected to calculate the predictors of elevation (Ele) and potential incoming solar radiation (PISR) (McCune and Keon, 2002). Soil organic matter is also an important factor affecting the ALT of permafrost. Due to the low decomposition rate of organic matter, high soil organic carbon has been accumulated in the permafrost regions (Ping et al., 2008). The adiabatic properties of organic matter relative to minerals will reduce the heat exchange between ground and air (Mölders and Romanovsky, 2006; Nicolsky et al., 2007; Paquin and Sushama, 2015). Moreover, the organic matter can also affect the thermal properties and the amount of unfrozen water of soil (Romanovsky and Osterkamp, 2000; Nicolsky et al., 2009). In order to consider the influence of the regional organic matter content (Wu et al., 2012b), soil organic carbon content information (SOC, ton·ha^{-1}) from global SoilGrids 1-km data (available at https://soilgrids.org; Hengl et al., 2014) was also used in our prediction analysis. Finally, all of the data layers were resampled to the matching spatial resolution (0.1°×0.1°) and cropped to the study area (QTP).

5) Glacier and lake data

The spatial distributions of the glaciers and lakes on the QTP were collected from the Second Glacier Inventory Dataset of China and the Chinese Cryosphere Information System provided by the Cold and Arid Regions Science Data Center (http://westdc.westgis.ac.cn).

2.2. Model description

Statistical models are general methods in the study of geography. It is usually built on some theoretical assumptions, and the data need to obey or approximately conform to a specific spatial distribution before the model can obtain credible results. However, ML algorithm is a general approximation algorithm, which generally does not require theoretical assumptions. The spatial analysis algorithm based on ML does not need a prior knowledge but a set of training data to learn the patterns of the geoscience system (Lary et al., 2016). Based on the above characteristics, we chose two statistical models and two ML algorithms to fit the present and future MAGT and ALT in this paper. The generalized linear modeling (GLM) and the generalized additive modeling (GAM) are traditional statistical methods used to simulate the thermal regimes of permafrost (Nan et al., 2002; Zhang et al., 2012a). And the two ML algorithms are the generalized boosting method (GBM) and random forest (RF). In this study, all the four models were executed based on the R software program. The detailed information and characteristics of the models are as follows:

1) Generalized linear model

The generalized linear model (GLM) is an extension of a linear model that can deal with the nonlinear relationships between explanatory variables and response variables (Nelder and Wedderburn, 1972):

$$g\left(\mu\left(x\right)\right)=\beta_0+\beta_1\left(x_1\right)+\beta_2\left(x_2\right)+\ldots+\beta_i\left(x_i\right) \tag{1}$$

where $g\left(\mu\right)$ is the link function connecting the estimated mean to the distribution of the response variable (here is MAGT and ALT), $\mu = E_{\left(y/x_1,x_2,x_3,\ldots,x_i\right)}$, E is the expected value, β_0 is the intercept component, β_i is the regression coefficient to be estimated and x_i is the predictor. For MAGT and ALT, GLM was based on first and second order polynomials and identity–link function.

2) Generalized additive model

Generalized additive model (GAM) is semi-parametric extensions of GLM that specify smoothing functions to fit nonlinear response curves to the data (Hastie and Tibshirani, 1986):

$$g\left(\mu(x)\right)=\beta_0+f_1(x_1)+f_2(x_2)+\ldots+f_i(x_i) \tag{2}$$

where $g(\mu)$ is the link function connecting the estimated mean to the distribution of the response variable (here is MAGT and ALT), μ = $E_{\left(y/x_1,x_2,x_3,\ldots,x_i\right)}$, E is the expected value, β_0 is the intercept component, f_i is a smoothing function for each explanatory variable and x_i is the predictor. To associate the MAGT and ALT with environmental predictors, the maximum smoothing function was set to three which were subsequently optimized by the model fitting function.

3) Generalized boosting method

The generalized boosting method (GBM, based on the R package dismo) is a sequential integration modeling method that combines a large number of iteratively fitted classification trees into a single model, using cross-validation methods to estimate the optimal number of trees, and thereby improving prediction accuracy (Elith et al., 2008). GBMs automatically incorporate interactions between predictors and are capable of modeling highly complex nonlinear systems (Aalto et al., 2018). GBMs (with Gaussian–Markov error assumption) were fitted using the gbm.step function, including the main parameters of the learning rate, tree complexity, bagging fraction, maximum number of trees, and others.

4) Random forest

Random forest (RF, implemented in the R package randomForest.) is a ML algorithm based on a classification tree, which forms a "forest" by generating a large ensemble of regression trees. The model uses a bootstrap sampling method to extract multiple samples from the original samples, conduct decision tree modeling for each sample, and then combine the prediction of multiple decision trees in order to obtain the final prediction result through a voting process. The model is characterized by strong applicability, effective avoidance of over-fitting and insensitivity to missing data and multivariate collinearity (Breiman et al., 2001; Hutengs and Vohland 2016). It is an effective empirical approach in the nonlinear-regression systems and its superiority has been proved useful by a large number of applications in the earth system (Lary et al., 2016).

To study the effects of predictors on MAGT and ALT, our models were designed using the following specifications:

$$\mathrm{MAGT} = f_1(TDD) + f_2(FDD) + f_3(Sol_{pre}) + f_4(Liq_{pre}) + f_5(PISR) + f_6(SOC) + f_7(Lon) + f_8(Lat) + f_9(Ele) \tag{3}$$

$$\mathrm{ALT} = f_1(TDD) + f_2(FDD) + f_3(Sol_{pre}) + f_4(Liq_{pre}) + f_5(PISR) + f_6(SOC) + f_7(Lon) + f_8(Lat) + f_9(Ele) \tag{4}$$

The independent variables in these equations are same, while the corresponding $f_i(x_i)$ in each equation is different. In order to fully consider the advantages and disadvantages of the above four models and to reduce the uncertainty, we used an ensemble approach. This method puts the average of the four models as the new results. The optimal model was determined by comparing the key parameters of the final five results. Model performance was assessed using a repeated cross-validation (CV) scheme. Based on a total of 84 boreholes and 70 ALT observation sites, the models gave the simulated results after 10 times fitting processes using a random sample of 90% of the observation data and verification processes using the remaining 10%. After each CV run for all models, the predicted and observed values of MAGT and ALT were compared in the terms of the root-mean-square error (RMSE), mean difference (cf. bias), and R-squared (R^2).

3. Results

3.1. Reliability assessment of MAGT and ALT

The simulation results were compared with the *in situ* observation data using cross-validation. A comparison of the five results (Figure 2) reveals that there was no significant bias between the simulated values and the available borehole data on the QTP, but the RMSE and R^2 of the ensemble method imply that it was more reliable than the other four methods. The consistency between the measured and simulated MAGT at most sites for the five models was less than 1°C. Among these models, the ensemble method performed optimally, with a simulation accuracy for 80 sites of < 1°C, which account for 95% of the total sites. It exhibited a strong positive correlation between the simulated and observed MAGT (R^2 = 0.73, p < 0.001). Overall, the ensemble method (Figure 2(e)) displayed the highest accuracy among the models in forecasting the MAGT. For this reason, the ensemble model was selected to simulate the present MAGT and future trends.

Similarly, the simulated ALT results were compared with the *in situ* observation data using the same statistical method. For ALT, the best fitting result was RF (Figure 3(d)), which exhibited the highest R^2 and the lowest RMSE values of 0.51 and 0.69 m, respectively. Although the GLM method exhibited a smaller bias, the difference between the two methods was not large. Overall, the validations for the five results did not differ significantly. Based on further comparison of Figures 2 and 3, it can be seen that the fitting accuracy of MAGT was better than that of ALT, with R^2 values of the corresponding optimal fitting results of 0.73 and 0.51, respectively. This is due to the fact that the spatial heterogeneity of the ALT is larger than that of the MAGT on the QTP, and the ALT will fluctuate greatly during climate change within a short period (Cao et al., 2017).

We calculated the error distribution for five typical regions separately (Table 1). Overall, the distribution of RMSE and bias on the QTP was relatively uniform, with the exception of the RMSE in the AEJIR. The reason for this may be that there are relatively few observation sites in the northern part of the whole investigated regions, and the simulating accuracy has high sensitivity to single points and poor regional representation. In addition, permafrost along the G109 Highway is greatly affected by human activities, and there are more observation sites in this region. Compared with the error statistics of the entire QTP, the RMSE of MAGT in the G109IR was relatively small, while the RMSE of ALT was relatively large. Thus, we may conclude that MAGT is relatively less affected by human activities, while ALT is more affected by disturbance and displays great spatial heterogeneity. In terms of bias, the region with the largest bias was GZIR. The reason is that GZIR located in the transition zone between permafrost and seasonally frozen ground, and the accuracy of the results would be affected to some extent.

3.2. MAGT and ALT during the reference period

Using the collected borehole data, we fitted the meteorological factors and geographical environmental factors to obtain the MAGT distribution of the permafrost regions on the QTP (Figure 4). We extracted the MAGT of the QTP below 0 °C as an average range of permafrost (Chen et al., 2015), which indicating suitable conditions for permafrost, with a total area of 1.04×10^6 km^2 (excluding glaciers and lakes). Considering the heterogeneity and uncertainty of ground temperature on the QTP, the minimum permafrost extent is 0.8×10^6 km^2 (the area within MAGT ≤ -0.5°C), and the maximum extent is 1.28×10^6 km^2 (the area within MAGT $\leq +0.5$°C). Compared with the pan-Arctic permafrost, the permafrost temperature on the QTP is relatively high (Obu et al., 2019). Nearly half of the permafrost temperature area on the QTP exceed -1.0ºC and the average temperature is -1.35 ± 0.42 ºC. The permafrost temperature is not only affected by latitude, but also by altitude. As illustrated in Figure 4, the lower-temperature permafrost on the QTP generally occurs in high-altitude mountains, and the ground temperature gradually rises with decreasing altitude, with the lowest value distributes in the Kunlun Mountain and its surrounding regions. In general, the MAGT on the QTP was found to be higher in the southern region (GZIR) than that in the northern region (AEJIR), and higher in the eastern region (WQIR) than that in the western region (XKLIR).

Based on permafrost extent, the spatial distribution of the ALT for the entire QTP was obtained (Figure 5). The statistical results indicated that the average ALT is 2.3 ± 0.60 m on the QTP, and the ALT value of ~ 90% of the permafrost region ranged from 1.6 to 3.0 m. Geographically, the ALT in the eastern part of the QTP is relatively thinner (generally no more than 2 m) with slight variations. The ALT along the Qinghai-Tibet Highway and in the central and western plateau is highly heterogeneous. The overall ALT pattern is thin in the mountains, thick on the plains, thin in the hinterlands, and thick along the periphery of the permafrost. The maximum value appears along the southern boundary of the permafrost and the surrounding sporadic permafrost (generally ≥ 3.2 m). In general, MAGT and ALT exhibit a consistent trend in spatial distribution, with a correlation coefficient of 0.44. The smaller value of MAGT corresponds to thinner ALTs.

3.3. The projection of MAGT and ALT

In view of a strong statistical rule of MAGT and ALT in climatic factors (e.g., TDD and FDD) and topographic factors (e.g., Lon, Lat, and Ele), most studies have begun to use similar statistical methods to investigate the present and future development trends of the periglacial climate realm (Koven et al., 2013; Aalto et al., 2017, 2018; Zhang et al., 2019). In this study, the optimal fitting model for the present state was employed to forecast MAGT and ALT under different future climate scenarios. For ALT, the spatial domain was limited to the area with simulated MAGT ≤ 0°C during each associated period and/or RCP scenario.

Due to climate change, the permafrost temperature exhibits an obvious rising trend on the QTP. We simulated the future change of permafrost on the QTP after half a century. The results revealed that the future changes of MAGT and ALT are predicted to be pronounced, but region-specific (Figure 6). The forecasted average MAGT over the QTP permafrost regions will increase from -1.35°C (present status) to -0.66°C by 2061–2080 (RCP2.6) and to 0.25°C for RCP8.5 (Table 2). The ALT, however, was only predicted to increase from 2.3 m (2000–2015) to 2.7 m (2061–2080) for RCP8.5. The reason for the consistency or small change of the ALT is that, the section of the permafrost with a MAGT near 0°C is forecasted to degrade to seasonally frozen ground, and this part of the permafrost usually corresponds to a thicker active layer. Additionally, the uncertainties related to the forecasts of MAGT and ALT under different RCPs in the future were given. And, the uncertainties are characterized by the range of MAGT value and ALT value. As can be seen in Figure

7, even under the different RCPs scenarios, the fluctuation range of MAGT and ALT is basically the consistent.

Over the next half century, the near-surface permafrost areas are predicted to continue to decrease by 0.13×10^6 km^2 (12%), 0.42×10^6 km^2 (40%) and 0.60×10^6 km^2 (58%) on the QTP by 2070 (2061–2080), under the RCP2.6, RCP4.5 and RCP8.5 scenarios, respectively. The result is basically consistent with the projected change by Chang et al. (2018) (Figure 8). Permafrost is in non-equilibrium under the influence of climate change, and there may be no permafrost that is driven by the current climate. In fact, it may be that permafrost is degrading, so the distribution range of the simulation results may be underestimated (Zhao et al., 2019). The changes in MAGT and ALT are not only related to the changes in temperature and precipitation but also closely related to hydrothermal parameters and surface energy balance (Guo and Wang, 2016; Hu et al., 2019). Based on the existing observation data and improved soil physics, the estimated changes in previous studies are generally larger than that of actual change (Lawrence et al., 2012; Cheng et al., 2019; Wang et al., 2019b).

4. Discussion

In order to project the possible future changes of permafrost, we simulated MAGT and ALT changes under the present state and future scenarios based on statistical and ML methods. The results show that under different RCPs, significant degradation of the QTP permafrost may occur (e.g., MAGT rising and ALT thickening); in particular, under RCP8.5, more than half of the near-surface permafrost will disappear, and regional differences were observed. In this section, to further verify the feasibility of our results, we compared our simulated MAGT and ALT with those of previous studies and then analyzed the vulnerability of permafrost to climate change under the present state. Based on these findings, we proposed urgent action should be taken to adapt climate change. Finally, the model performance and potential sources of the uncertainty in this study were discussed.

4.1. Comparisons with previous results

The most likely permafrost area on the QTP is 1.04×10^6 km^2 (the region where MAGT < 0°C, Figure 4), or about 45.4% of the total QTP land surface area. Our results were compared with the permafrost distribution map of the QTP for the period 2003–2012 based on the TTOP model, which was basically consistent with the new permafrost area (1.06×10^6 km^2, Zou et al., 2017). The two results showed substantial consistency, with a kappa coefficient of 0.63 (Table 3). However, there were still certain spatial differences (Figure 9). These differences mainly occurred at the southern margin of the continuous permafrost and the islands of permafrost in the south eastern QTP.

For the results of MAGT and ALT, a similar study showed relatively large deviations at the hemispheric scale (the RMSEs of MAGT and ALT were 1.6°C and 0.89 m, respectively; Aalto et al., 2018). In their study, an interesting discovery was mentioned, for both MAGT and ALT: after considering the area north of 60°N, the uncertainty was greatly reduced. This is primarily due to the fact that the permafrost on the QTP is quite different from that of the pan-Arctic region. The QTP is the dominant by the high-altitude permafrost, while the pan-Arctic is mainly the high-latitude permafrost. Compared with the pan-Arctic region, the active layer on the QTP is thicker, the ground temperature is higher, and the spatial heterogeneity is greater (Nicolsky et al., 2017; Cao et al., 2017; Qin et al., 2017). Therefore, combining the QTP permafrost and the pan-Arctic permafrost hemispherically will inevitably reduce the accuracy of the results.

We further compared the simulated results of MAGT and ALT with previous studies on the QTP. Table 4 summarizes the error statistics among different types of permafrost models (i.e., equilibrium model, transient model and statistical model). We can find that for the R-value, our method combined of the statistical and ML has the similar accuracy with the transient model. Although the RMSE of ALT in our study is the largest among all models, the differences are not significant. Moreover, the RMSE of MAGT in our study shows relatively smaller error. Meanwhile, from the overall spatial distribution of the ALT, although there are some differences in the spatial details, the distribution pattern of our result is comparable with the presented recently (Zhao and Wu, 2019; Wang et al., 2020b). In generally, our model can obtain a relatively higher simulation accuracy.

We qualitatively analyzed the main reasons for these differences as follows. Firstly, there are differences in accuracy among different types of models, such as the equilibrium models and mechanistic transient models. Secondly, there is a slight gap between the research period and the data used for verification. Permafrost is often viewed as a product of long-term climate change, which is slowly changing (Zhang et al., 2007); this may also lead to differences between the results. Finally, the 0.1° resolution of our model can't capture all of regional information on climate change, which may limit the model's ability to capture detailed changes in the permafrost to some extent, especially in the boundary of the permafrost region (Etzelmüller, 2013; Guo and Wang, 2016). Therefore, the ability to capture the permafrost edge information should be further improvement. Overall, by comparing with previous studies on the QTP, that our method is relatively simple and effective, and thus could be a useful tool to evaluate the permafrost conditions with a high accuracy on the QTP.

4.2. Permafrost vulnerability

According to Figure 4, the ground temperature of the entire QTP permafrost is relatively high. In order to analyze the vulnerability of the QTP permafrost to climate warming, the permafrost region with MAGTs ranging from -0.5 to 0.5°C was extracted (Figure 10). According to the permafrost stability classification (Cheng and Wang, 1982), permafrost in this range is classified as unstable region. It can be observed that 0.49×10^6 km^2 of the permafrost area over the QTP is in danger at present, which accounting for 37.3% of the maximum permafrost area. This unstable permafrost primarily distributed in the transition region of permafrost and seasonally frozen ground.

As a result of the global warming and increased anthropogenic activity, the QTP has experienced an approximately 3-fold warming increase over the past 50 years (Wan et al., 2018). Under the influence of this accelerated warming, the permafrost region adjacent to the seasonally frozen ground is becoming increasingly fragile (Qin et al., 2017). This part of the permafrost is generally in the process of ice-water phase transformation. A comparison with Figure 6, reveals that this region is consistent with the areas in which permafrost will disappear under future RCPs, but it also greatly affected by the local ground ice content, underlying surface types, and other related factors (Nelson et al., 2001; Yang et al., 2010c).

The Qinghai-Tibet Engineering Corridor (QTEC, the region that contains the Qinghai-Tibet Highway and Railway, pipelines, electric transmission lines, and so on) is an important conduit connecting mainland China and the QTP. Under the influence of intensifying global climate change and frequent human activities, the ecological environment along the QTEC is fragile, and the permafrost in the QTEC has degraded significantly and the alpine ecosystem is facing new challenges (Niu et al., 2018). Based on Figure 10, the statistical results show that 757 km of the QTEC crosses through the permafrost region (at its maximum extent), accounting for nearly 40% of its total length (from Xining to Lhasa). Of this, approximately half of the QTEC faces the risk of the permafrost disappearing, and the other half may experience varying degrees of permafrost degradation in the future. This will result in huge economic losses and threaten associated infrastructures along the QTEC.

Recent studies have shown that several cryosphere tipping points are dangerously close (IPCC, 2019), and the permafrost in the Arctic has begun to thaw irreversibly and release carbon dioxide and methane, but the inevitable effects could still be mitigated by reducing greenhouse gas emissions (Lenton et al., 2019). The stability and resilience of the QTP permafrost is in peril. We should take urgent action to reduce greenhouse gas emissions, and put them as the priority of the present and future work. In order to effectively mitigate the degradation of permafrost, all the emission reduction measures should be reflected in words even in actions.

4.3. Model performance and uncertainty analysis

Our study integrated field observation data, meteorological data, geospatial environmental predictors and multiple statistical models to study MAGT and ALT changes in the present and future QTP permafrost regions. Based on the CV analysis, the reliability of both predictions displayed relatively low uncertainty. For MAGT, the benefits of using the ensemble modeling approach were obvious, i.e., the average of the four methods yielded the best simulation result. For ALT, large errors still existed among the ensemble modeling approach after CV, which indicating that the method does not always produce the most reliable results. The simulation accuracy of ALT is lower than that of MAGT, and the result can only represent the general change trend of ALT. The main reason for this is that, the spatial heterogeneity of ALT on the QTP is large, with the change rate of ALT per unit (100 m^2) reaching 80%, thus resulting in the relatively low R^2 values and large RMSEs (Cao et al., 2017). Additionally, our model predicts the equilibrium state of permafrost and does not consider the lag time associated with the formation and degradation of permafrost (Xu et al., 2017b). Compared with previous studies, although our results show great reliability, there are still some uncertainties embedded in the predictions, including the measurement accuracy of the data, the equilibrium assumption in the statistical modeling and the influence of other factors (Aalto et al., 2018).

Due to the limitations of the observation data, we had to use one-year or multi-year averages to represent the present state and to fit the model. MAGT and ALT changed during this period, however, in particular, ALT changed greatly at the inter-annual scale. We did our best to collect datasets with MAGT and ALT, but the number of sample points used for training was still limited, and the model was still highly sensitive to single observations. To some extent, this also indicates that the number of observation sites on the QTP is too sparse to represent the present large spatial heterogeneity of the plateau.

When calculating the input factors of the model, in the future warming scenarios, the TDD and FDD were calculated based on the monthly mean air temperature instead of the daily mean air temperature. This approximate calculation method will bring some unavoidable errors, especially when the temperature is close to 0 °C (Wu et al., 2011; Shi et al., 2019). Additionally, we simply take 0°C temperature as the critical temperature threshold between solid precipitation and liquid precipitation, while, in most cases, snowfall events even occur in some regions on the QTP when the air temperature is > 4°C, but not 0 °C (Wang et al., 2016).

In this study, some key soil parameters, including soil texture, soil moisture content and bulk density, were excluded from the analyses in the model due to missing data, which exerted strong influence on water and heat transfer in the active layer as well as the change in permafrost temperature (Wu et al., 2017b; Du et al., 2020). The PISR and SOC in permafrost region are not static. However, it was assumed to be the fixed value in our model. With the further research on the key predictors of the permafrost region, we will add more dynamic datasets to our model. In summary, we used statistical and ML models combined with easily accessible data to simulate the present and future dynamics of permafrost on the QTP. By comparison and verification, our model can obtain high precision results through a relatively simple calculation process.

5. Conclusions

In this study, the method combined of statistical and ML was used to obtain the key permafrost metrics in both the present and a half-century in the future (2061–2080) on the QTP. Based on the comparison with *in situ* observation data and previous researches, we found that this method was reliable for simulating the changes in MAGT and ALT. We demonstrated the permafrost degradation from a quantitative perspective. Our results can provide a scientific basis for the study of climate change in permafrost. The main conclusions are listed as follows:

1) A combination method of statistical and ML models is efficient to capture the changes in the thermal state of the permafrost on the QTP.

2) The present (2000–2015) permafrost area on the QTP is approximate to be 1.04 × 10^6 km^2. The average MAGT and ALT of the permafrost region amount to -1.35 ± 0.42ºC and 2.3 ± 0.60 m, respectively.
3) In the future (2061–2080), the maximum permafrost area may be reduced to 0.44 × 10^6 km^2. The future changes of MAGT and ALT are forecast to be pronounced, but region-specific.
4) The unstable permafrost mainly distributed at the edge of the permafrost region, and approximately half permafrost in the QTEC will be at risk of disappearing in the future.

Acknowledgements

This work was financially supported by the Natural Science Foundations of China (41690142; 41771076; 41961144021; 42071093), and the CAS "Light of West China" Program. The logistical supports from the Cryosphere Research Station on the Qinghai-Tibet Plateau are especially appreciated. Datasets for this research are available at https://data.mendeley.com/datasets/hbptbpyw75/1. We also thank the three anonymous reviewers for their constructive suggestions.

References

Aalto, J., Harrison, S., & Luoto, M. (2017). Statistical modelling predicts almost complete loss of major periglacial processes in Northern Europe by 2100. *Nature Communications*, 8, 515. https://doi.org/10.1038/s41467-017-00669-3

Aalto, J., Karjalainen, O., Hjort, J., & Luoto, M. (2018). Statistical Forecasting of Current and Future Circum-Arctic Ground Temperatures and Active Layer Thickness. *Geophysical Research Letters*, 45, 4889-4898. https://doi.org/10.1029/2018GL078007

Boeckli, L., Brenning, A., Gruber, S., & Noetzli, J. (2012). Permafrost distribution in the European Alps: calculation and evaluation of an index map and summary statistics. *The Cryosphere*, 6(4), 807-820. https://doi.org/10.5194/tc-6-807-2012

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Cao, B., Gruber, S., Zhang, T., Li, L., Peng, X., Wang, K., ... & Guo, H. (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(3), 574-591. https://doi.org/10.1002/2016JF004018

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., & Westermann, S. (2017). An observation-based constraint on permafrost loss as a function of global warming. *Nature Climate Change*, 7(5), 340-344. https://doi.org/10.1038/nclimate3262

Chang, Y., Lyu, S., Luo, S., Li, Z., Fang, X., Chen, B., Li, R., & Chen, S. (2018). Estimation of permafrost on the Tibetan Plateau under current and future climate conditions using the CMIP5 data. *International Journal of Climatology*, 38(15), 5659-5676. https://doi.org/10.1002/joc.5770

Chen, B., Luo, S., Lyu, S., Fang, X., & Chang Y. (2017). Land surface characteristics in soil freezing and thawing process on the tibetan plateau based on community land model (in Chinese with English abstract). *Journal of Glaciology and Geocryology,* 39(04), 760-770.

Chen, H., Nan, Z., Zhao, L., Ding, Y., Chen, J., & Pang, Q. (2015). Noah modelling of the permafrost distribution and characteristics in the West Kunlun area, Qinghai-Tibet Plateau, China. *Permafrost and Periglacial Processes*, 26(2), 160-174. https://doi.org/10.1002/ppp.1841

Cheng, G., &Wang, S. (1982). On the zonation of high-altitude permafrost in China (in Chinese with English abstract). *Journal of Glaciology and Geocryology*, 4(2), 1-17.

Cheng, G., & Wu, T. (2007). Responses of permafrost to climate change and their environmental significance, Qinghai-Tibet Plateau. *Journal of Geophysical Research: Earth Surfac*e, 112(F2). https://doi.org/10.1029/2006JF000631

Cheng, G., Zhao, L., Li, R., Wu, X., Sheng, Y., Hu, G., ... & Wu, Q. (2019). Characteristic, changes and impacts of permafrost on Qinghai-Tibet Plateau. *Chinese Science Bulletin*, 64(27), 2783-2795. https://doi.org/10.1360/tb-2019-0191

Du, Y., Li, R., Zhao, L., Yang, C., Wu, T., Hu, G., ... & Ma, J., 2020. Evaluation of 11 soil thermal conductivity schemes for the permafrost region of the central Qinghai-Tibet Plateau. *Catena*, 193, 104608. https://doi.org/10.1016/j.catena.2020.104608

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Etzelmüller, B. (2013). Recent advances in mountain permafrost research. *Permafrost and Periglacial Processes*, 24(2), 99-107. https://doi.org/10.1002/ppp.1772

Fang, X., Luo, S., Lyu, S., Chen, B., Zhang, Y., Ma, D., & Chang, Y. (2016). A simulation and validation of CLM during freeze-thaw on the Tibetan Plateau. *Advances in Meteorology*, 2016, 1-15. http://dx.doi.org/10.1155/2016/9476098

Gao, Y., Li, K., Chen, F., Jiang, Y., & Lu, C. (2015). Assessing and improving Noah-MP land model simulations for the central Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, 120(18), 9258-9278. https://doi.org/10.1002/2015JD023404

Guo, D., & Wang, H. (2016). CMIP5 permafrost degradation projection: A comparison among different regions. *Journal of Geophysical Research: Atmospheres*, 121(9), 4499-4517. https://doi.org/10.1002/2015JD024108

Guo, D., & Wang, H. (2017). Permafrost degradation and associated ground settlement estimation under 2°C global warming. *Climate Dynamics.* 49, 2569-2583. http://dx.doi.org/10.1007/s00382-016-3469-9

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

Harris, C., Arenson, L. U., Christiansen, H. H., Etzelmüller, B., Frauenfelder, R., Gruber, S., ... & Isaksen, K. (2009). Permafrost and climate in Europe: Monitoring and modelling thermal, geomorphological and geotechnical responses. *Earth-Science Reviews, 92*(3-4), 117-171. https://doi.org/ 0.1016/j.earscirev.2008.12.002

Hastle, T. J., & Tibshirani, R. (1986). Generalized additive models (with discussion). *Statistical Science,* 1, 297-318.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X. (2020). The first high-resolution meteorological forcing dataset for land process studies over China. *Scientific Data*, 7, 25, https://doi.org/10.1038/s41597-020-0369-y.

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., ... & Gonzalez, M. R. (2014). SoilGrids1km-global soil information based on automated mapping. *PloS one*, 9(8), e105992. https://doi.org/10.1371/journal.pone.0105992

Hu, G., Zhao, L., Li, R., Wu, T., Xiao, Y., Jiao, K., ... & Jiao, Y. (2013). The water-thermal characteristics of frozen soil under freeze-thaw based on CoupModel (in Chinese with English abstract). *Scientia Geographica Sinica, 33*(3), 356-362. https://doi.org/10.13249/j.cnki.sgs.2013.03.356

Hu, G., Zhao, L., Li, R., Wu, T., Wu, X., Pang, Q., ... & Shi, J. (2015). Modeling hydrothermal transfer processes in permafrost regions of Qinghai-Tibet Plateau in China (in Chinese with English abstract). *Chinese geographical science, 25*(6), 713-727. https://doi.org/10.1007/s11769-015-0733-6

Hu, G., Zhao, L., Li, R., Wu, X., Wu, T., Zhu, X., ... & Hao, J. (2019). Simulation of land surface heat fluxes in permafrost regions on the Qinghai-Tibetan Plateau using CMIP5 models. *Atmospheric research*, 220, 155-168. https://doi.org/10.1016/j.atmosres.2019.01.006

Hu, G., Zhao, L., Li, R., Wu, X., Wu, T., Chang, W., ... & Hao, J. (2020). Thermal properties of active layer in permafrost regions with different vegetation types on the Qinghai-Tibetan Plateau. *Theoretical and Applied Climatology*. 139, 1-11. https://doi.org/10.1007/s00704-019-03008-2

Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, 178, 127-141. https://doi.org/10.1016/j.rse.2016.03.006

IPCC. (2019). Summary for policymakers. In: Pörtner HO, Roberts DC, Masson-Delmotte V et al (eds) IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. In press. https://report.ipcc.ch/srocc/pdf/SROCC_FinalDraft_FullReport.pdf

Jafarov, E. E., Coon, E. T., Harp, D. R., ... & Romanovsky, V. E. (2018). Modeling the role of preferential snow accumulation in through talik development and hillslope groundwater flow in a transitional permafrost landscape. *Environmental Research Letters*, 13(10). https://doi.org/10.1088/1748-9326/aadd30

Koven, C. D., Riley, W. J., & Stern, A. (2013). Analysis of Permafrost Thermal Dynamics and Response to Climate Change in the CMIP5 Earth System Models. *Journal of Climate*, 26(6), 1877-1900. https://doi.org/10.1175/JCLI-D-12-00228.1

Lary, D. J., Alavi, A. H., Gandomi A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3-10.

Lawrence, D. M., Slater, A. G., & Swenson, S. (2012). Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4. *Journal of Climate*, 25(7), 2207-2225. https://doi.org/10.1175/JCLI-D-11-00334.1

Lee, W. L., Liou, K. N., & Wang, C. C. (2013). Impact of 3-D topography on surface radiation budget over the Tibetan Plateau. *Theoretical and applied climatology*, 113(1-2), 95-103. https://doi.org/10.1007/s00704-012-0767-y

Lenton, T. M., Rockström, J., Gaffney, O., Rahmstorf, S., Richardson, K., Steffen, W., & Schellnhuber, H. J. (2019). Climate tipping points-too risky to bet against. *Nature*, 575(7784), 592. https://doi.org/10.1038/d41586-019-03595-0

Liu, G., Zhao, L., Li, R., Wu, T., Jiao, K., & Ping, C. (2017). Permafrost warming in the context of step-wise climate change in the Tien Shan Mountains, China. *Permafrost and Periglacial Processes*, 28(1), 130-139. https://doi.org/ 10.1002/ppp.1885

Liu, Y., Zhao, L., & Li, R. (2013). Simulation of the soil water thermal features within the active layer in Tanggula Region, Tibetan Plateau, by using SHAW model (in Chinese with English abstract). *Journal of Glaciology and Geocryology*, 35(2), 280-290.

McCune, B., & Keon, D. (2002). Equations for potential annual direct incident radiation and heat load. *Journal of vegetation science*, 13(4), 603-606. https://doi.org/10.1658/1100-9233(2002)013[0603:EFPADI]2.0.CO;2

Michaelides, R. J., Schaefer, K., Zebker, H. A., Parsekian, A. D., Liu, L., Chen, J., ... & Schaefer, S. R. (2019). Inference of the impact of wildfire on permafrost and active layer thickness in a discontinuous permafrost region using the remotely sensed active layer thickness (ReSALT) algorithm. *Environmental Research Letters*, 14(3). https://doi.org/10.1088/1748-9326/aaf932

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., ... & Meehl, G. A. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282), 747. https://doi.org/10.1038/nature08823

Molders, N., & Romanovsky, V. E. (2006). Long-term evaluation of the Hydro-Thermodynamic Soil-Vegetation Scheme's frozen ground/permafrost component using observations at Barrow, Alaska. Journal of Geophysical Research. 111, D04105. https://doi.org/10.1029/2005JD005957

Nan, Z., Li, S., & Liu, Y. (2002). Mean Annual Ground Temperature Distribution on the Tibetan Plateau: Permafrost Distribution Mapping and Further Application (in Chinese with English abstract). *Journal of Glaciology and Geocryology.* 24, 142-148.

Nan, Z., Li S., Cheng G., & Huang P. (2012). Surface frost number model and its application to the Tibetan plateau (in Chinese with English abstract). *Journal of Glaciology and Geocryology*, 34(1), 89-95.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370-384. https://doi.org/10.1201/9780203753736

Nelson, F. E., & Outcalt, S. I. (1987). A computational method for prediction and regionalization of permafrost. *Arctic and Alpine Research*. 19, 279–288.

Nelson, F. E., Shiklomanov, N. I., Mueller, G. R., Hinkel, K.M., Walker, D.A., & Bockheim, J.G. (1997). Estimating active-layer thickness over a large region: Kuparuk River Basin, Alaska, U.S.A. *Arctic and Alpine Research*. 29, 367–378.

Nelson, F. E., Anisimov, O. A., & Shiklomanov, N. I. (2001). Subsidence risk from thawing permafrost. *Nature,* 410(6831), 889. https://doi.org/10.1038/35073746

Nicolsky, D. J., Romanovsky, V. E., Alexeev, V. A., & Lawrence, D. M. (2007). Improved modeling of permafrost dynamics in a GCM land-surface scheme. *Geophysical Research Letters*, 34(8). https://doi.org/10.1029/2007GL029525

Nicolsky, D., Romanovsky, V., Panteleev, G. (2009). Estimation of soil thermal properties using in-situ temperature measurements in the active layer and permafrost. *Cold Regions Science and Technology*, 55(1):120-129.

Nicolsky, D. J., Romanovsky, V. E., Panda, S. K., Marchenko, S. S., & Muskett, R. R. (2017). Applicability of the ecosystem type approach to model permafrost dynamics across the Alaska North Slope. *Journal of Geophysical Research: Earth Surface*, 122(1), 50-75. https://doi.org/10.1002/2016JF003852

Niu, F., Gao, Z., Lin, Z., Luo, J., & Fan, X. (2019). Vegetation influence on the soil hydrological regime in permafrost regions of the Qinghai-Tibet Plateau, China. *Geoderma*, 354, 113892. https://doi.org/10.1016/j.geoderma.2019.113892

Niu, F., Yin, G., Luo, J., Lin, Z., & Liu, M. (2018). Permafrost distribution along the Qinghai-Tibet Engineering Corridor, China using high-resolution statistical mapping and modelling integrated with remote sensing and GIS. *Remote Sensing*, 10(2), 215. https://doi.org/10.3390/rs10020215

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., ... & Khomutov, A. (2019). Northern Hemisphere permafrost map based on TTOP modelling for 2000-2016 at 1 km^2 scale. *Earth-Science Reviews*, 2019. https://doi.org/10.1016/j.earscirev.2019.04.023

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., … Zeng, X. (2010). Technical Description of version 4.0 of the Community Land Model (CLM) (No. NCAR/TN-478+STR). University Corporation for Atmospheric Research. https://doi.org/10.5065/D6FB50WZ

Painter, S., Coon, E., Atchley, A., Berndt, M., Garimella, R., Moulton, J., Svyatskiy, D., & Wilson, C. (2016). Integrated surface/subsurface permafrost thermal hydrology: Model formulation and proof−of−concept simulations. *Water Resources Research*, 52(8):6062-6077.

Pang, Q., Cheng, G., Li, S., & Zhang, W. (2009). Active layer thickness calculation over the Qinghai-Tibet Plateau. *Cold Regions Science and Technology*, 57(1), 23-28. https://doi.org/10.1016/j.coldregions.2009.01.005

Pang, Q., Zhao, L., Ding, Y., & Li, S. (2010). Analysis about the influence on the thermal regime in permafrost regions with different underlying surfaces. *Sciences in Cold and Arid Regions*, 2(3), 0203-0211.

Pang, Q., Zhao, L., Li, S., & Ding, Y. (2012). Active layer thickness variations on the Qinghai-Tibet Plateau under the scenarios of climate change. *Environmental earth sciences*, 66(3), 849-857. https://doi.org/10.1007/s12665-011-1296-1

Paquin, J. P., & Sushama, L. (2015). On the Arctic near-surface permafrost and climate sensitivities to soil and snow model formulations in climate models. *Climate Dynamics*, 44(1), 203-228. https://doi.org/10.1007/s00382-014-2185-6

Peng, X., Zhang, T., Frauenfeld, O., Wang, K., Luo, D., Cao, B., ... & Wu, Q. (2018). Spatiotemporal changes in active layer thickness under contemporary and projected climate in the Northern Hemisphere. *Journal of Climate*. 31 (7):251-266.

Ping, C. L., Michaelson, G. J., Jorgenson, M. T., Kimble, J. M., Epstein, H., Romanovsky, V. E., & Walker, D. A. (2008). High stocks of soil organic carbon in the North American Arctic region. *Nature Geoscience*, 1(9), 615. https://doi.org/10.1038/ngeo284

Qin, D., Yao, T., Ding, Y., & Ren, J. (2016). Introduction to cryospheric science. *China Meteorological Press*.

Qin, Y., Wu, T., Zhao, L., Wu, X., Li, R., Xie, C., ... & Liu, G. (2017). Numerical modelling of the active layer thickness and permafrost thermal state across Qinghai-Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, 122(21), 11-604. https://doi.org/10.1002/2017JD026858

Qu, Y., Zhu, Z., Chai, L., Liu, S., Montzka, C., Liu, J., ... & Guo, Z. (2019). Rebuilding a Microwave Soil Moisture Product Using Random Forest Adopting AMSR-E/AMSR2 Brightness Temperature and SMAP over the Qinghai-Tibet Plateau, China. *Remote Sensing,* 11(6), 683. https://doi.org/10.3390/rs11060683

Ran, Y., Li, X., & Cheng, G. (2018). Climate warming over the past half century has led to thermal degradation of permafrost on the Qinghai-Tibet Plateau. *The Cryosphere, 12*(2), 595-608. https://doi.org/10.5194/tc-12-595-2018

Reuter, H. I., Nelson, A., & Jarvis, A. (2007). An evaluation of void-filling interpolation methods for SRTM data. *International Journal of Geographical Information Science*, 21(9), 983-1008. https://doi.org/10.1080/13658810601169899

Riseborough, D., Shiklomanov, N., Etzelmüller, B., Gruber, S., & Marchenko, S. (2008). Recent advances in permafrost modelling. *Permafrost and Periglacial Processes*, 19(2), 137-156. https://doi.org/10.1002/ppp.615

Romanovsky, V., Osterkamp, T. (2015). Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost. *Permafrost & Periglacial Processes*, 11(3):219-239.

Shiklomanov, N. I., & Nelson, F. E. (2002). Active-layer mapping at regional scales: A 13-year spatial time series for the Kuparuk region, north-central Alaska. *Permafrost and Periglacial Processes*, 13(3), 219-230. https://doi.org/10.1002/ppp.425

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485-498. https://doi.org/10.1175/BAMS-D-11-00094.1

Walvoord, M. A., Voss, C. I., Ebel, B. A., Minsley, B. J. (2018). Development of perennial thaw zones in boreal hillslopes enhances potential mobilization of permafrost carbon. *Environmental Research Letters*, 4(1). https://doi.org/10.1088/1748-9326/aaf0cc

Wan, W., Zhao, L., Xie, H., Liu, B., Li, H., Cui, Y., ... & Hong, Y. (2018). Lake Surface Water Temperature Change Over the Tibetan Plateau From 2001 to 2015: A Sensitive Indicator of the Warming Climate. *Geophysical Research Letters*, 45(20), 11-177. https://doi.org/10.1029/2018GL078601

Wang, C., Wang, Z., Kong, Y., Zhang, F., Yang, K., & Zhang, T. (2019b). Most of the Northern Hemisphere Permafrost Remains under Climate Change. *Scientific reports*, 9(1), 3295. https://doi.org/10.1038/s41598-019-39942-4

Wang, J., & Zhang, M. (2016). Change of snowfall/rainfall ratio in the Tibetan Plateau based on a gridded dataset with high resolution during 1961-2013 (in Chinese with English abstract). *Acta Geographica Sinica*, 71(1), 142-152.

Wang, Q., Jin, H., Zhang, T., Cao, B., Peng, X., Wang, K., ... & Li, L. (2017). Hydro-thermal processes and thermal offsets of peat soils in the active layer in an alpine permafrost region, NE Qinghai-Tibet plateau. *Global and Planetary Change, 156*, 1-12. https://doi.org/10.1016/j.gloplacha.2017.07.011

Wang, T., Wu, T., Wang, P., Li, R., Xie, C., & Zou, D. (2019a). Spatial distribution and changes of permafrost on the Qinghai-Tibet Plateau revealed by statistical models during the period of 1980 to 2010. *Science of the Total Environment, 650*, 661-670. https://doi.org/10.1016/j.scitotenv.2018.08.398

Wang, K., Jafarov, E., & Overeem, I. (2020a). Sensitivity evaluation of the Kudryavtsev permafrost model. *Science of the Total Environment*, 720, 137538. https://doi.org/10.1016/j.scitotenv.2020.137538.

Wang, T., Yang, D., Yang, Y., Piao, S., Li, X., Cheng, G., & Fu, B. (2020b). Permafrost thawing puts the frozen carbon at risk over the Tibetan Plateau. *Science Advances.* 6, eaaz3513.

Wang, W., Wu, T., Zhao, L., Li, R., Zhu, X., Wang, W., ... & Hao, J. (2018a). Exploring the ground ice recharge near permafrost table on the central Qinghai-Tibet Plateau using chemical and isotopic data. *Journal of Hydrology*, 560, 220-229. https://doi.org/10.1016/j.jhydrol.2018.03.032

Wang, Y., Spencer, R. G., Podgorski, D. C., Kellerman, A. M., Rashid, H., Zito, P., ... & Xu, Y. (2018b). Spatiotemporal transformation of dissolved organic matter along an alpine stream flow path on the Qinghai-Tibet Plateau: importance of source and permafrost degradation. *Biogeosciences*, 15(21), 6637-6648. https://doi.org/10.5194/bg-15-6637-2018

Westermann, S., Langer, M., Boike, J., Heikenfeld, M., Peter, M., Etzelmüller, B., & Krinner, G. (2016). Simulating the thermal regime and thaw processes of ice-rich permafrost ground with the land-surface model CryoGrid 3. *Geoscientific Model Development*, 9(2), 523-546. https://doi.org/10.5194/gmd-9-523-2016

Westermann, S., Østby, T. I., Gisnås, K., Schuler, T. V., & Etzelmüller, B. (2015). A ground temperature map of the North Atlantic permafrost region based on remote sensing and reanalysis data. *The Cryosphere*, 9(3), 1303-1319. https://doi.org/10.5194/tc-9-1303-2015

Wheeler, D., Shaw, G., & Barr S. (2013). Statistical techniques in geographical analysis. *Routledge*. https://doi.org/10.4324/9780203821503

Wu, Q., Zhang, T., & Liu, Y. (2012a). Thermal state of the active layer and permafrost along the Qinghai-Xizang (Tibet) Railway from 2006 to 2010. *The Cryosphere*, 6(3), 607-612. https://doi.org/10.5194/tc-6-607-2012

Wu, T., Wang, Q., Zhao, L., Batkhishig, O. & Watanabe, M. (2011). Observed trends in surface freezing/thawing index over the period 1987-2005 in Mongolia. *Cold regions science and technology*, 69(1), 105-111.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., ... & Zhang, F. (2019). The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, 12(4), 1573-1600. https://doi.org/10.5194/gmd-12-1573-2019

Wu, X., Fang, H., Zhao, Y., Smoak, J. M., Li, W., Shi, W., ... & Ding, Y. (2017b). A conceptual model of the controlling factors of soil organic carbon and nitrogen densities in a permafrost-affected region on the eastern Qinghai-Tibetan Plateau. *Journal of Geophysical Research: Biogeosciences*, 122(7), 1705-1717. https://doi.org/10.1002/2016JG003641

Wu, X., Nan, Z., Zhao, S., Zhao L., & Cheng G. (2018). Spatial modeling of permafrost distribution and properties on the Qinghai−Tibet Plateau. *Permafrost and Periglacial Processes*, 29(2):86-99.

Wu, X., Xu, H., Liu, G., Ma, X., Mu, C., & Zhao, L. (2017a). Bacterial communities in the upper soil layers in the permafrost regions on the Qinghai-Tibetan plateau. *Applied soil ecology*, 120, 81-88. https://doi.org/10.1016/j.apsoil.2017.08.001

Wu, X., Zhao, L., Chen, M., Fang, H., Yue, G., Chen, J., ... & Ding, Y. (2012b). Soil organic carbon and its relationship to vegetation communities and soil properties in permafrost areas of the central western Qinghai-Tibet plateau, china. *Permafrost and Periglacial Processes*, 23(2), 162-169. https://doi.org/10.1002/ppp.1740

Wu, X., Zhao, L., Fang, H., Zhao, Y., Smoak, J. M., Pang, Q., & Ding, Y. (2016). Environmental controls on soil organic carbon and nitrogen stocks in the high-altitude arid western Qinghai-Tibetan Plateau permafrost region. Journal of Geophysical Research: *Biogeosciences*, 121(1), 176-187. Xin, X., Gao, F., Wei, M., Wu, T., Fang, Y., & Zhang, J. (2018). Decadal prediction skill of BCC-CSM1. 1 climate model in East Asia. *International Journal of Climatology*, 38(2), 584-592. https://doi.org/10.1002/joc.5195

Xu, X., Wu, Q., & Zhang, Z. (2017a). Responses of active layer thickness on the qinghai-tibet plateau to climate change (in Chinese with English abstract). *Journal of Glaciology and Geocryology*, 39(01): 1-8.

Xu, X., Zhang, Z., & Wu, Q. (2017b). Simulation of permafrost changes on the Qinghai-Tibet Plateau, China, over the past three decades. *International journal of digital earth*, 10(5), 522-538. https://doi.org/10.1080/17538947.2016.1237571

Xue, B., Wang, L., Yang, K., Tian, L., Qin, J., Chen, Y., ... & Li, X. (2013). Modeling the land surface water and energy cycles of a mesoscale watershed in the central Tibetan Plateau during summer with a distributed hydrological model. *Journal of Geophysical Research: Atmospheres*, 118(16), 8857-8868. https://doi.org/10.1002/jgrd.50696

Yang, C., Wu, T., Wang, J., Yao, J., Li, R., Zhao, L., ... & Hao, J. (2019). Estimating Surface Soil Heat Flux in Permafrost Regions Using Remote Sensing-Based Models on the Northern Qinghai-Tibetan Plateau under Clear-Sky Conditions. *Remote Sensing*, 11(4), 416. https://doi.org/10.3390/rs11040416

Yang, K., He, J., Tang, W., Qin, J., & Cheng, C. C. (2010b). On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau. *Agricultural and Forest Meteorology,* 150(1), 38-46. https://doi.org/10.1016/j.agrformet.2009.08.004

Yang, M., Nelson, F. E., Shiklomanov, N. I., Guo, D., & Wan, G. (2010a). Permafrost degradation and its environmental effects on the Tibetan Plateau: A review of recent research. *Earth-Science Reviews, 103*(1-2), 31-44. https://doi.org/10.1016/j.earscirev.2010.07.002

Yang, Z., Ou, Y., Xu, X., Zhao, L., Song, M., & Zhou, C. (2010c). Effects of permafrost degradation on ecosystems (in Chinese with English abstract). *Acta Ecologica Sinica*, 30(1), 33-39. https://doi.org/10.1016/j.chnaes.2009.12.006

Zhang, T., Frauenfeld, O.W., Serreze, M.C., Etringer, A., Oelke, C., Mccreight, J., Barry, R.G., Gilichinsky, D., Yang, D., & Ye, H. (2005). Spatial and temporal variability in active layerthickness over the Russian Arctic drainage basin. *Journal of Geophysical Research ： Atmosphere*. 110, 2227–2252.

Zhang, Z., Wu, Q., Zhang, Z., & Hou, Y. (2012a). Analysis of the mean annual ground temperature changes on the Qinghai-Tibet plateau permafrost region under condition of climate warming (in Chinese with English abstract). *Journal of Engineering Geology*, 04, 610-613.

Zhang, W., Wang, G., Zhou, J., Liu, G., & Wang, Y. (2012b). Simulating the Water-Heat Processes in Permafrost Regions in the Tibetan Plateau Based on CoupModel (in Chinese with English abstract). *Journal of Glaciology and Geocryology*, 34(5), 1099-1109.

Zhang, Z., & Wu, Q. (2012a). Predicting changes of active layer thickness on the Qinghai-Tibet Plateau as climate warming (in Chinese with English abstract). *Journal of Glaciology and Geocryology*, 34(3), 505-511.

Zhang, Z., & Wu, Q. (2012b) Thermal hazards zonation and permafrost change over the Qinghai-Tibet Plateau. *Natural Hazards*, 61(2), 403-423. https://doi.org/10.1007/s11069-011-9923-4

Zhao, D., & Wu, S. (2019). Projected Changes in Permafrost Active Layer Thickness Over the Qinghai-Tibet Plateau Under Climate Change. *Water Resources Research*, 55, 7860-775. https://doi.org/10.1029/2019WR024969

Zhao, L., & Sheng, Y. (2019). Permafrost and its changes on qinghai-tibet plateau (in Chinese). *Beijing: Science Press*.

Zhao, T. J., Zhang, L. X., Shi, J. C., & Jiang, L. M. (2011). A physically based statistical methodology for surface soil moisture retrieval in the Tibet Plateau using microwave vegetation indices. *Journal of Geophysical Research: Atmospheres*, 116(D8). https://doi.org/10.1029/2010JD015229

Zheng, G., Yang, Y., Yang, D., Dafflon, B., Lei, H., & Yang, H. (2019). Satellite-based simulation of soil freezing/thawing processes in the northeast Tibetan Plateau. *Remote Sensing of Environment*, 231, 111269. https://doi.org/10.1016/j.rse.2019.111269

Zhu, X., Wu, T., Zhao, L., Yang, C., Zhang, H., Xie, C., ... & Du, Y. (2019). Exploring the contribution of precipitation to water within the active layer during the thawing period in the permafrost regions of central Qinghai-Tibet Plateau by stable isotopic tracing. *Science of The Total Environmen*t, 661, 630-644. https://doi.org/10.1016/j.scitotenv.2019.01.064

Zou, D., Zhao, L., Yu, S., Chen, J., Hu, G., Wu, T., ... & Wang, W. (2017). A new map of permafrost distribution on the Tibetan Plateau. *The Cryosphere*, 11(6), 2527. https://doi.org/10.5194/tc-11-2527-2017

Figure 1. Location of the investigated regions and observation sites. Green dots and red triangles stand for the mean annual ground temperature (MAGT) and active layer thickness (ALT) monitoring sites, respectively. The black polygons depict the five typical regions.

Figure 2. Observed *vs*. simulated mean annual ground temperature (MAGT) for 84 borehole sites based on four statistical techniques (GLM = generalized linear model, GAM = generalized additive model, GBM = generalized boosting method, RF = random forest.) and an ensemble method (the average of the four methods). The red dashed lines are the ±1 °C intervals around the 1:1 line (in black solid line).

Figure 3. Observed *vs*. modeled active layer thickness (ALT) based on four statistical techniques (GLM = generalized linear model, GAM = generalized additive model, GBM = generalized boosting method, RF = random forest.) and an ensemble method (the average of the four methods). The red dashed lines are the ±1 m interval around the 1:1 line (in black solid line).

Figure 4. Spatial distribution of permafrost on the QTP based on the MAGT.

Figure 5. Distribution of the ALT on the permafrost regions of the QTP.

Figure 6. Forecast mean annual ground temperature (MAGT) and active layer thickness (ALT) across the study domains under different RCPs (RCP2.6, RCP4.5 and RCP8.5) for the 2070s (average of 2061–2080).

Figure 7. The uncertainty related to the spatial forecasts of mean annual ground temperature (MAGT) and active layer thickness (ALT) in RCP 2.6(a), RCP 4.5 (b), RCP 8.5 (c) scenarios. The uncertainty is quantified using a repeated (n = 1,000)

bootstrap sampling procedure inside the study domain. The boxplots depict the mean, median, 1st and 3rd quartiles and range of variation over 1000 predictions for modeling techniques.

Figure 8. Projections of the changes in permafrost area on the QTP under RCP2.6, RCP4.5, RCP6.0 and RCP8.5 via 7(a) surface frost index (SFI) and 7(b) Kudryavtsev method (KUD). The graph is derived from Chang *et al.* (2018). Shaded areas show the standard deviations across the CMIP5 models, the black lines show the equivalent present-day area, and the grey dotted line represent the degraded area in 2070 under different RCPs.

Figure 9. Spatial differences between our results (2000–2015) and those of Zou *et al* (2003–2012; TTOP model). P and SFG represent permafrost and seasonally frozen ground, respectively; Result is the permafrost distribution of this study. The permafrost distribution is obtained from Zou *et al.* (2017).

Figure 10. Spatial distribution of the permafrost regions prone to degradation.

Table 1. Model Error statistics of the ALT and MAGT in different typical regions

Region		(WQIR) East	(XKLIR) West	(GZIR) South	(AEJIR) North	(G109IR) Central	(QTP) Entire
MAGT	RMSE (°C)	0.60	0.56	0.61	0.73	0.45	0.53
	Bias (°C)	0.025	0.06	-0.15	-0.14	-0.03	-0.02
ALT	RMSE (m)	0.60	0.62	0.68	0.11	0.76	0.69
	Bias (m)	0.24	0.06	-0.46	0.09	0.18	-0.11

Table 2. Key characteristic metrics of permafrost under different RCPs

	Present	RCP2.6	RCP4.5	RCP8.5
	2000-2015	2061-2080		
MAGT (°C)	-1.35	-0.66	-0.14	0.25
ALT (m)	2.3	2.5	2.5	2.7
Area ($\times 10^6$ km^2)	1.04	0.91	0.62	0.44

Note: the statistics of mean annual ground temperatures (MAGT) in three scenarios (RCP2.6, RCP4.5, RCP8.5) were based on the permafrost range under present status.

Table 3. Discrepancy area of permafrost on QTP

	Area discrepancy ($\times 10^6$ km^2)	Percentage (%)
Both P	0.86	35.41
Result P and Zou SFG	0.18	7.41
Result SFG and Zou P	0.20	8.23
Both SFG	1.19	48.95
Total	2.43	100

Table 4. Compare the statistical errors between different types of models

	Numerical model	Time period	RMSE	R	Source
MAGT (°C)	Equilibrium model	2000-2016	1.85	0.20	Obu et al., 2019
	Transient model	2007-2010	**0.31**	**0.93**	Wu et al., 2018
	Statistical and ML	2000-2015	0.53	0.85	This study
ALT (m)	Equilibrium model	Before 2009	**0.47**	0.46	Pang et al., 2012
	Transient model	2007-2010	0.57	**0.86**	Wu et al., 2018
	Statistical and ML	2000-2015	0.69	0.71	This study

Note: bold data represents the best result for each model.