Received: 19 December 2018 | Revised: 19 April 2019 | Accepted: 27 April 2019

DOI: 10.1002/ppp.2006

RESEARCH ARTICLE

WILEY

Permafrost zonation index map and statistics over the Qinghai–Tibet Plateau based on field evidence

Bin Cao[1,2] | Tingjun Zhang[1] | Qingbai Wu[3] | Yu Sheng[3] | Lin Zhao[4] | Defu Zou[5]

[1] Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, China

[2] National Tibetan Plateau Data Center, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

[3] State Key Laboratory of Frozen Soil Engineering, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China

[4] School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing, China

[5] Cryosphere Research Station on the Qinghai–Tibet Plateau, State Key Laboratory of Cryospheric Science, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China

Correspondence
Tingjun Zhang, Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China.
Email: tjzhang@lzu.edu.cn

Funding information
Strategic Priority Research Program of Chinese Academy of Sciences, Grant/Award Numbers: XDA20100313 and XDA20100103; National Natural Science Foundation of China, Grant/Award Numbers: 41801028 and 41871050

Permafrost is prevalent over the Qinghai–Tibet Plateau (QTP), but mapping its distribution is challenging due to the limited availability of ground-truth data sets and strong spatial heterogeneity in the region. Based on a recently developed inventory of permafrost presence or absence from 1475 *in situ* observations, we developed and trained a statistical model and used it to compile a high-resolution (30 arc-seconds) permafrost zonation index (PZI) map. The PZI model captures the high spatial variability of permafrost distribution over the QTP because it considers multiple controlling variables, including near-surface air temperature downscaled from reanalysis, snow cover days and vegetation cover derived from remote sensing. Our results showed the new PZI map achieved the best performance compared to available existing PZI and traditional categorical maps. Based on more than 1000 *in situ* measurements, the Cohen's kappa coefficient and overall classification accuracy were 0.62 and 82.5%, respectively. Excluding glaciers and lakes, the area of permafrost regions over the QTP is approximately 1.54 (1.35–1.66) $\times 10^6$ km^2, or 60.7 (54.5–65.2)% of the exposed land, while area underlain by permafrost is about 1.17 (0.95–1.35) $\times 10^6$ km^2, or 46 (37.3–53.0)%.

KEYWORDS

mountians, permafrost distribution, permafrost zonation index, Qinghai-Tibet Plateau

1 | INTRODUCTION

The Qinghai–Tibet Plateau (QTP) is the largest and highest plateau in the world with a mean elevation above 4000 m a.s.l. The QTP is also known as the Asian water tower, because it is the source of several large rivers (e.g. the Yellow, the Yangtze and the Brahmaputra Rivers) that supply water for about 1.4 billion people downstream.[1] The QTP has the largest extent of permafrost in the low–middle latitudes. Permafrost over the QTP was reported to be sensitive to climate change due, in part, to its high temperatures (normally > −2 °C) and fragile landscape.[2,3] Its distribution has strong influences on the environment, such as the surface energy balance (e.g.[4]), hydrological processes (e.g.[5]), biogeochemical processes (e.g.[6]) and human systems (e.g.[7,8]).

Permafrost, a product of cold climates, typically develops beneath the ground surface at depths dependent on site-specific characteristics such as climate, subsurface water content and the presence of

insulating surface materials (e.g. peat). Permafrost depths exceeding 5 m have been reported in the QTP for the gravel soil near the boundaries of permafrost with high thermal diffusivity.[9] Because of this, permafrost presence or absence is often indiscernible from the surface and its characteristics, such as thermal state or distribution, are spatially heterogeneous. Aside from site-scale observations, permafrost distribution is mainly estimated using models.

In recent decades, many maps have been produced to estimate permafrost distribution at different scales over the QTP using statistical and (semi-)physical methods (e.g.[10-13]). These maps were usually produced at very small scales (from 1:600,000 to 1:10,000,000) and used a categorical classification (such as permafrost, seasonally frozen ground and unfrozen ground) to delineate permafrost regions.[14] More recently, due to the increased availability of high-resolution meteorological data sets, several higher resolution (~ 1–10 km) maps were developed which represent permafrost distribution as a binary presence or absence classification.[10,11,15,16] Based on the TTOP model and improved MODIS land surface temperature, Zou et al[10] presented a new permafrost map over the QTP with a spatial resolution of 1 km. The overall accuracy of this map was reported to be about 82%.[17] Applying numerical models to permafrost mapping in mountainous regions is challenging due to the need for high-resolution meteorological data to drive the model, which is difficult to obtain.[11]

Empirical-statistical models are a simple yet effective approach to estimating permafrost distribution, especially in mountains where *in situ* measurements are sparse.[18] Generalized linear models (GLMs) are a class of empirical-statistical model which have been applied to derive permafrost distribution at both global and regional scales (e.g.[14,19-22]). Compared with traditional categorical maps, GLMs can estimate permafrost distribution more precisely by introducing continuous permafrost probability indices which range from 0 to 1. They also take advantage of high-resolution topoclimatic explanatory variables such as mean annual air temperature (MAAT). For example, Gruber[19] derived a global permafrost zonation index (PZI) map (denoted here as PZI_{global}) through a heuristic-empirical relationship between permafrost and MAAT. Model parameters were established largely based on the boundaries of continuous and isolated permafrost in the International Permafrost Association (IPA) map and do not rely on field observations.[23] Studies have suggested that the PZI_{global} map is suitable for describing the overall spatial pattern of permafrost distribution over the QTP.[17,24] However, because the PZI_{global} map uses MAAT as its single predictor variable, it may underestimate the strong spatial heterogeneity of permafrost distribution over the QTP caused by local factors including snow cover and vegetation. Furthermore, these local factors normally could provide some hints on techniques and measures we can use to artificially simulate their similar effects and hence could further improve mapping accuracy.[14] The PZI_{global} map represents an improvement over the IPA map by adding detail and accurately representing major topographic features lacking in the IPA map through the use of high-resolution MAAT and topography data sets. Unfortunately, its model parameters calibrated using 1990s-era measurements of the IPA map, and up-to-date ground-truth information is absent from this calibration. Based on recent evaluations and inter-study comparisons, the accuracy of any QTP permafrost map is heavily reliant on the use of *in situ* measurements in the calibration of associated models. Incorporating more of the available field measurements for permafrost map update is recommended to improve the applicability of the permafrost maps over the QTP.[16]

Besides the empirically based GLMs, statistically based models are expected to be a valuable alternative for establishing regional PZI maps in mountains.[14,20-22] A statistically based PZI map could describe characteristics of permafrost distribution at a regional scale better by considering multiple controlling factors (e.g. MAAT, precipitation and radiation) and by using a model that has been calibrated based on a large number of *in situ* measurements. Although rock glaciers have been used as indicators of current or former permafrost conditions to calibrate and test permafrost maps,[24] they are absent in much of the QTP due to low precipitation.[25] As a result, a regional PZI map derived based on *in situ* measurements over the QTP is not yet available. There is a growing awareness that global warming will result in permafrost degradation. This, coupled with an increasing amount of infrastructure built on permafrost in the QTP, means that simulations of permafrost distribution will become increasingly important in the context of both environmental change and engineering design.

Recently, a new inventory of permafrost presence or absence was developed over the QTP.[17] At the same time, the increased availability of accurate digital elevation models (DEMs), and improved climate data products (e.g. gridded re-analysis and remote sensing data sets) have allowed for the development of new models for simulating high-resolution permafrost distribution over the QTP. In this study, we develop and describe a statistical PZI model to predict the probability of permafrost over the QTP based on multiple predictor variables and a large number of *in situ* measurements. We evaluate the new PZI model and map based on an inventory of permafrost presence or absence evidence from various methods. We then compare the new PZI map (denoted as PZI_{QTP} map) with the PZI_{global} and traditional categorical maps. Specially, we aim to

(1) compile a new calibrated PZI map on the QTP that is based on *in situ* evidence and multiple climatic predictor variables;
(2) provide summary statistics describing the permafrost distribution on the QTP; and
(3) discuss the bias and uncertainties of the new PZI map by comparing it to existing maps.

2 | DATA

2.1 | Inventory of permafrost presence or absence evidence

In this study, we use a recently developed inventory of permafrost presence or absence evidence to derive and evaluate the PZI model

over the QTP.[17] In general, this inventory provides a representative training data set for site-specific permafrost presence or absence across a wide range of topographic conditions (e.g. elevation and slope aspect) and climatic conditions (e.g. near-surface air temperature and snow cover). There are a total of 1475 *in situ* sites derived from various methods. Of these, 1141 are acquired from borehole temperature measurements, 6 from soil pits, 184 from ground surface temperature measurements, and 144 from ground-penetrating radar

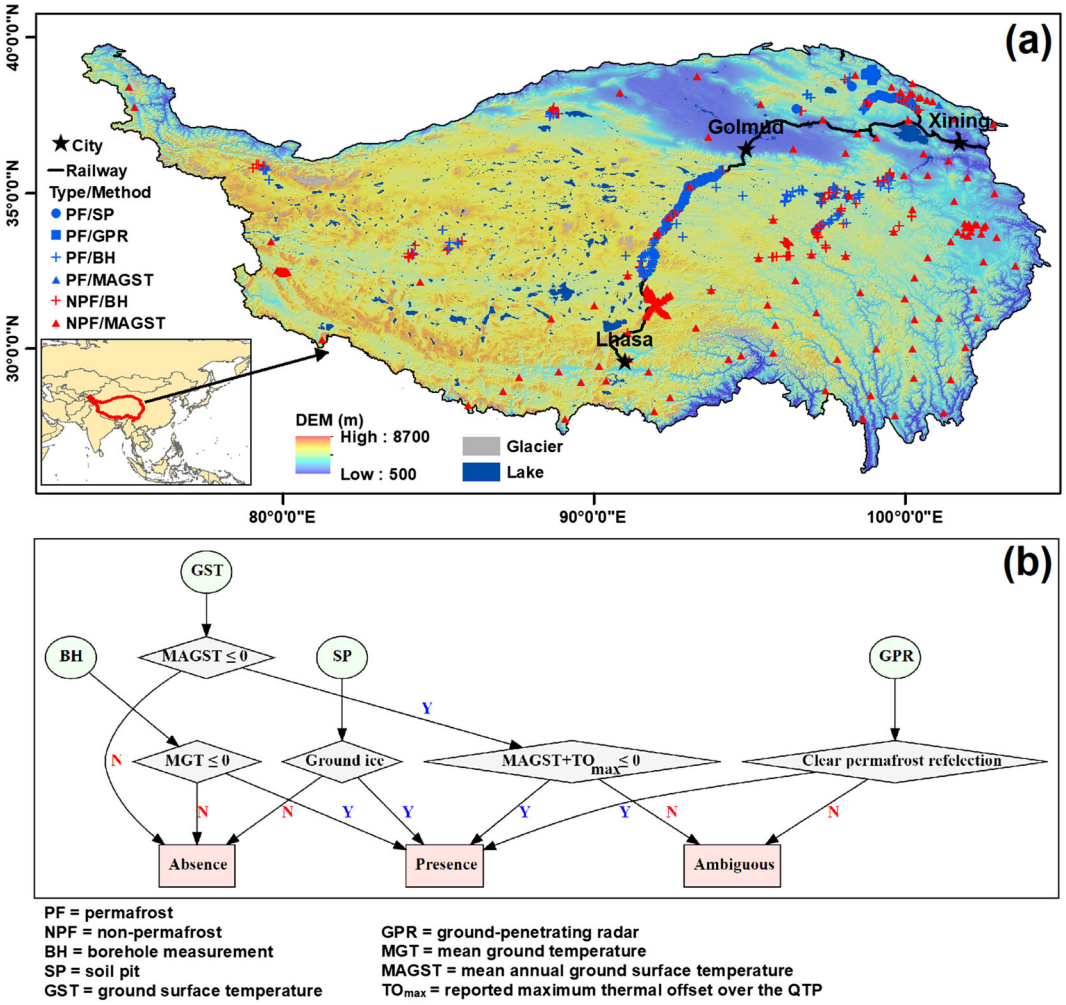

FIGURE 1 (a) Distribution of permafrost presence or absence evidence over the QTP. (b) Flow chart used to determine if evidence acquired from various methods supported or refuted permafrost presence. A red capital N or blue capital Y indicates answers of no or yes, respectively, at each decision point. TO_{max} indicates thermal offset boundary of 0.79 ° C obtained from literature. Ambiguous indicates the data are not sufficient to assess permafrost presence or absence; these sites are not included in the inventory. The flow chart was developed following Cao et al.[17]

TABLE 1 Summary of inventory of the permafrost presence or absence evidence over the Qinghai–Tibet Plateau

Method	Original evidence		Aggregated evidence	
	No. of samples	Elevation (m a.s.l.)	No. of samples	Elevation (m a.s.l.)
BH	1141 (PF: 860; NPF: 281)	3284–5232	796 (PF: 567; NPF: 229)	3295–5232
SP	6 (PF: 6; NPF: 0)	3885–5100	6 (PF: 6; NPF: 0)	3885–5100
GST	184 (PF: 2; NPF: 182)	1583–4835	176 (PF: 2; NPF: 174)	1583–4835
GPR	144 (PF: 144; NPF: 0)	3681–4166	62 (PF: 62; NPF: 0)	3687–4139
Total	1475 (PF: 1012 NPF: 463)	1583–5232	1040 (PF: 637 NPF: 403)	1583–5232

Notes. Original means all the *in situ* measurements are based on *in situ* measurements (details could be found from Cao et al.[17]), while Aggregated are the aggregated inventory by merging multi-study evidence in the same grid of unprojected SRTM30. BH = borehole temperature, SP = soil pit, GST = ground surface temperature, GPR = ground-penetrating radar, PF = permafrost, NPF = non-permafrost.

surveys (Figure 1a, Table 1). Figure 1b shows the flow chart that was used to assess permafrost presence or absence from each of the various methods. A more detailed description of this flowchart is provided by Cao et al.[17] In order to avoid problems with spatial autocorrelation and over-weighting of dense observations, point-scale permafrost presence and absence measurements were aggregated to a regular grid co-registered with the unprojected SRTM30 grid (30 arc-seconds). Within each grid cell, the presence or absence of permafrost was determined by counting the number of sites in each category. The grid cell was assigned a value based on which category (presence or absence) was in the majority. For grid cells with the same number of permafrost presence and absence sites, the value of the nearest site from the grid center was chosen to represent the cell. As a consequence, there are 1040 aggregated sites left for further analysis (Table 1).

2.2 | Meteorological observations

In order to evaluate the downscaled MAAT and derive snowpack information, the mean daily near-surface air temperature and snow depth for 87 stations from the China Meteorological Administration (CMA) were used (Figure A1). The data were collected each day by trained professional technicians then evaluated using automated quality control scripts, and hence are more reliable than raw data.[26]

2.3 | Explanatory variables

The spatial distribution and thermal state of permafrost result from the long-term interaction of climate, surface conditions and subsurface conditions. The MAAT, mean annual snow cover days (MASCD) and Normalized Difference Vegetation Index (NDVI) are selected here as potential explanatory variables representative of climate and surface conditions. It has been widely investigated that the presence of a peat layer may reduce the ground temperature (e.g[27,28]). However, we do not consider the subsurface soil conditions here because (a) measurements of subsurface soil conditions, even information for presence or absence of a peat layer, are not available at most sites in the inventory; (b) the available geological maps over the QTP have an unsuitably coarse resolution; and (c) dividing the 1475 or 1040 aggregated measurements into different sub-groups based on subsurface conditions or soil type would result in too few samples per group for robust statistical analysis. While radiation measurements, corresponding to the influences of topography, are widely used as an explanatory variable for deriving PZI in mountainous areas (e.g.[13,14,20-22]), there are too few sites in steep areas in the inventory to justify including radiative variables. For example, 52.3% and 83.3% of sites have slopes smaller than 5° and 10°, respectively.[17]

2.3.1 | Mean annual air temperature

The MAAT during the reference period 1979–1999 was used here as a predictor variable. It is well known that the total permafrost simulation chain may be significantly affected by the choice of input data sets.[29] ERA-Interim, produced by the European Center for Medium-range Weather Forecast (ECMWF), was chosen for this study as it was found to describe the complex environments over the QTP better than other re-analyses,[30] and its downscaled results have achieved promising results (e.g.[31-33]).

To derive the high-resolution PZI map, a statistical downscaling method was employed to yield the near-surface air temperature. It has been shown that the downscaled air temperature is sensitive to the downscaling technique used[31] and, for this reason, we downscaled air temperature using two methods which differ in how they treat the surface effects, and combined the two results in a smooth way. The method developed by Fiddes[32] (denoted as downscaling method 1, DSM1) was used to derive MAAT for areas with hypsometric position smaller than 0.1, corresponding to areas dominated by the upper-air temperature, such as mountain tops (see Appendix A1 for details on the simulation of hypsometric position). The method developed by Gao et al.[33] (denoted as downscaling method 2, DSM2) was used for areas with hypsometric position larger than 0.3, corresponding to areas dominated by surface effects. A smooth combination was used for the other areas (from 100% surface effects at hyposmetric position of 0.3 to no surface effects at 0.1), and so the MAAT can be written as:

$$MAAT = (1-\alpha)MAAT_1 + \alpha MAAT_2 \quad (1)$$

where $MAAT_1$ is the MAAT derived from DSM1 and $MAAT_2$ is from DSM2; α is the fractional contribution of $MAAT_2$ to MAAT, and can be expressed as:

$$\alpha = \begin{cases} 0, & H \leqslant H_u \\ \dfrac{H - H_u}{H_l - H_u}, & H_u < H < H_l \\ 1 & H \geqslant H_l \end{cases} \quad (2)$$

where H is the hyposmetric position estimated from the SRTM30, H_l is the lower boundary of hyposmetric position, and H_u is the upper boundary. H_l and H_u were 0.3 and 0.1, respectively. Detailed evaluation of downscaled MAAT can be found in Appendix A2 and Figure A1.

2.3.2 | Snow cover duration

Snow cover plays an important role in permafrost distribution by influencing energy exchange between the atmosphere and the surface.[34] The scaled MASCD (SMASCD) was obtained as an indicator of snow cover because no reliable snow water equivalent products were available for the QTP (Appendix B, Figure B1). The annual snow cover days were derived from the daily snow cover product developed by Wang et al.[35] based on MODIS products (MOD10A1 and MYD10A1). The snow-specific and overall classification accuracies over the QTP were reported as 85% and 98%, respectively. To make the snow cover days more comparable, they were normalized to the interval 0–1 by dividing by the total number of days in each year, and the resulting SMASCD during 2003–2010 was used as a predictor. The original SMASCD product with a spatial resolution of 500 m was bi-linearly interpolated to the unprojected SRTM30 grid.

2.3.3 | Vegetation coverage

To represent the influence of vegetation on permafrost distribution, the 16-day NDVI from MODIS/Terra (MOD13Q1, v006) was used. This is a level-3 MODIS product and is available on a global grid at a spatial resolution of 250 m. The annual maximum NDVI was computed for each year during 2001–2017 to approximately represent the amount of vegetation, and then a median value of the annual maximum for each pixel was calculated over the entire period to avoid sensitivity to extreme values. The median of annual maximum NDVI ($NDVI_{max}$) was used as a predictor variable within the PZI model. The original $NDVI_{max}$ product with a spatial resolution of 250 m was bi-linearly interpolated to the unprojected SRTM30 grid.

2.4 | Digital elevation model and topography

A 3 arc second (~ 90 m) DEM was created by spatially averaging the Global Digital Elevation Model version 2 (GDEM2), which has an original resolution of 1 arc second.[36,37] This was done to reduce the level of noise in the original data set.[31] Subsequently, topographic indices including slope and aspect were calculated for all sites in the inventory using RSAGA.[38] The algorithm used was a third-order finite difference weighted by the reciprocal of squared distance.[39] The outline of the QTP is taken from Zhang et al.,[40] glacier outlines are from Liu et al.[41] representing conditions in 2010, and lake data are provided by the Third Pole Environment Database. The unprojected grid of SRTM30 DEM with a resolution of 30 arc seconds (~1 km) was used to align predictor variable data sets as well as the final output PZI map in this study.

3 | METHODS

3.1 | Permafrost zonation index model

A GLM is introduced here as the statistical model of permafrost distribution. GLMs can be applied to model the continuous occurrence probabilities of binary categorical response variables such as the presence (Y=1) or absence (Y=0) of permafrost based on one or multiple independent variables.[22] In contrast to the linear regression model, the GLM places fewer restrictions on the types and statistical distributions of independent predictor variables. The independent variables may be continuous or categorical, and need not be normally distributed. MAAT, SMASCD and $NDVI_{max}$ were selected to be predictor variables (Figure 2). Vegetation and snow cover duration were expected to be influenced by near-surface air temperature, and therefore corresponding interaction terms were added to the PZI model. Fundamentally, the statistical model simulates the probability of permafrost occurrence based on known values of the predictor variables, and can be written as:

$$f(P) = \alpha_0 + \beta_1 T_a + \beta_2 Snow + \beta_3 Veg + \beta_4 T_a{\cdot}Snow + \beta_5 T_a{\cdot}Veg \quad (3)$$

where P is the probability of permafrost zonation, and is referred to as the PZI representing permafrost distribution. T_a, *Snow* and *Veg* are the near-surface air temperature, snow cover and vegetation cover, respectively, and are represented by MAAT downscaled from ERA-Interim, SMASCD and $NDVI_{max}$ derived from MODIS (see Section 2.3). α_0 is the regression constant, and β_1, β_2, β_3, β_4 and β_5 are regression coefficients. $f(P)$ is the link function in a logistic format

$$f(P) = \ln\left(\frac{P}{1-P}\right) \quad (4)$$

The PZI could therefore also be written as:

$$P = \frac{e^{(\alpha_0+\beta_1 T_a+\beta_2 Snow+\beta_3 Veg+\beta_4 T_a{\cdot}Snow+\beta_5 T_a{\cdot}Veg)}}{1+e^{(\alpha_0+\beta_1 T_a+\beta_2 Snow+\beta_3 Veg+\beta_4 T_a{\cdot}Snow+\beta_5 T_a{\cdot}Veg)}} \quad (5)$$

All model fitting and statistical calculations were conducted using the R programming language (version 3.4.3).

3.2 | Permafrost region and area calculations and statistics

In order to estimate and evaluate permafrost distribution as determined by the PZI model, it is important to highlight the difference between the extent of permafrost regions and permafrost area. A permafrost region is the exposed land surface below which the probability of permafrost occurrence is above some threshold value. The choice of threshold is arbitrary, but 10% permafrost coverage is often used.[19,42] Importantly, there may not be permafrost underlying the entirety of a permafrost region. On the other hand, permafrost area must necessarily be underlain by permafrost. The PZI developed by Gruber[19] was designed so that the numerical PZI value represented the proportion of a pixel underlain by permafrost. In such a case, specified thresholds are required for both the extent of permafrost region and permafrost area simulation based on the PZI as model output. Following Gruber,[19] we selected only the areas with PZI $\geq$ 0.01 for further analysis and area can therefore be calculated by multiplying PZI by pixel area for each pixel within the threshold. Permafrost regions were defined as the areas with PZI $\geq$ 0.1.

The pixel area of the unprojected SRTM30 grid used in this study can be calculated for any arbitrary pixel as Δy (longitude direction) × Δx (latitude direction), where Δy is defined as:

$$\Delta y = Res\frac{\pi R}{180} \quad (6)$$

where *Res* is the resolution of SRTM30 in degrees and R is the mean radius of the Earth with a constant value of 6371 km. Δx can be found as follows:

$$\Delta x = \cos(lat)\Delta y \quad (7)$$

where *lat* is the latitude of the pixel within the SRTM30 grid.

3.3 | Evaluation of PZI model and map

Preliminary investigations of the PZI model (preliminary model) were conducted using all *in situ* observations and predictor variables, but

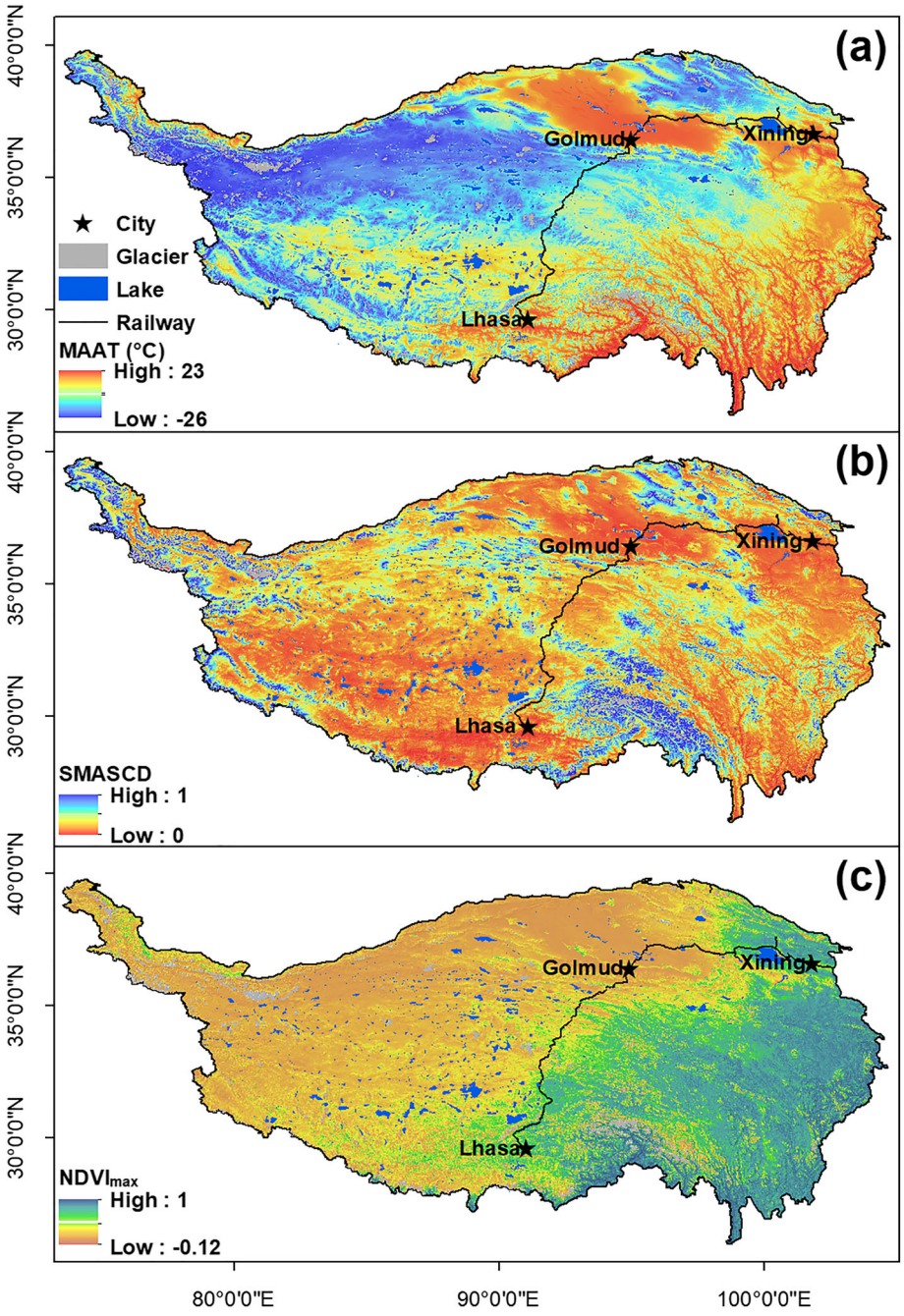

FIGURE 2 The predictor variables of (a) MAAT (~ 1 km) during 1979–1999 downscaled from ERA-Interim, (b) SMASCD (~ 500 m) during 2003–2010 and (c) $NDVI_{max}$ (~ 250 m) during 2001–2017 from MODIS, used in the PZI model(s)

only the significant variables were selected for final PZI simulations. To assess the performance of the PZI model, we conducted a 10-fold cross-validation. This is a commonly used method to compare and select a model for a given predictive modeling problem and it generally has a lower bias than other methods such as a simple split of training and test data (e.g.[20,31,43]). In this method, the inventory is first divided into two subsets based on permafrost presence or absence. Each time, ~90% of the permafrost presence and absence sites are randomly selected from the subsets for deriving model parameters, and the remaining 10% are used for evaluation. This process is repeated 10 times with different samples each time.

The model was evaluated using classification metrics for categorical variables[35,44]:

$$\begin{aligned} PCC_{PF} &= \frac{PF_T}{PF_T + PF_F} \times 100\% \\ PCC_{NPF} &= \frac{NPF_T}{NPF_T + NPF_F} \times 100\% \\ PCC_{tol} &= \frac{PF_T + NPF_T}{PF_T + PF_F + NPF_T + NPF_F} \times 100\% \end{aligned} \quad (8)$$

where *PCC* is percentage of sites correctly classified, and the variables *PF* and *NPF* represent counts of model predictions of permafrost or

non-permafrost, respectively. The subscripts *T* (True, correctly classified) and *F* (False, incorrectly classified) indicate whether the prediction was correct or incorrect. Put another way, PF_T is a measure of the true positives and PF_N a measure of false positives. The threshold of PZI = 0.5 was used to separate predictions of permafrost and non-permafrost.[21,22] Because there are more sites with permafrost (61.3%) than sites without permafrost (38.7%) in the inventory, Cohen's kappa coefficient (κ) was also introduced here for map evaluation. Parameter κ measures inter-rater agreement for categorical items,[45] and is expressed as:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{11}$$

where p_e and p_o are the probability of random disagreement and agreement, respectively, calculated as

$$p_e = \frac{(PF_T + PF_F) \times (PF_F + NPF_T) + (PF_F + NPF_T) \times (NPF_F + NPF_T)}{(PF_T + PF_F + NPF_F + NPF_T)^2}$$
$$p_o = \frac{PF_T + NPF_T}{PF_T + PF_F + NPF_F + NPF_T} \tag{12}$$

Two other permafrost maps for the QTP were also evaluated using the *in situ* data as benchmarks against which the PZI_{QTP} map can be compared. The first was the map developed by Zou et al.[10] (referred to as QTP_{TTOP}), which is the best calibrated categorical map of the area. The second was the PZI_{global} map, which is the only PZI map available that covers the entire QTP. The model parameters of the PZI_{global} map are estimated using the boundary conditions for continuous and island permafrost, that is PZI = 0.9 when the MAAT is equal to −8.0 °C and PZI = 0.1 when the MAAT is equal to -1.5 °C (referred to as normal case or PZI_{norm}). Furthermore, two variants representing either cold (conservative) or warm (non-conservative) conditions are introduced into the PZI_{global} map to allow the propagation of uncertainty caused by input data sets and model suitability. The two variants, denoted as the PZI_{cold} and PZI_{warm} maps, differ in the parameters used. A more detailed description of the model parameters is provided by Gruber.[19] There is an obvious temporal mismatch when using the recently collected and developed inventory to evaluate the PZI_{global} map, which was estimated based on the IPA map released in the 1990s.[19,23] However, this problem is somewhat unavoidable because *in situ* measurements before the 1990s are lacking. We argue that although the evaluation and inter-study comparisons with the PZI_{global} map presented here may underestimate its performance, it is necessary in order to assess the performance of the new PZI_{QTP} map. To compare and evaluate the modeled PZI to the measured proportion of permafrost presence, which describes the fraction of sites that show evidence for permafrost, the coefficient of determination (R^2), mean bias (BIAS) and root mean squared error (RMSE) were computed.

$$BIAS = \frac{1}{N}\sum_{i=1}^{N}(MOD_i - OBS_i)$$
$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(MOD_i - OBS_i)^2}{N}} \tag{14}$$

where *MOD* is the modeled PZI, *OBS* is the observed proportion of permafrost presence, and *N* is the total number of measurements.

4 | RESULTS AND INTERPRETATION

4.1 | Exploration of inventory of permafrost presence or absence

Summary statistics for the predictor variables at the 1040 measured sites are presented in Table 2. The range of each variable suggests that the sites represent a wide variety of field conditions and that the evidence here is sufficiently representative to derive a robust, statistically based PZI map over the QTP. Values ranged from −9.6 to 16.6 °C for MAAT, 0 to 0.57 for SMASCD and 0.01 to 0.88 for $NDVI_{max}$. Our results show clear differences in predictor variables between sites with and without permafrost (Figure 3). The mean MAAT for permafrost sites was about 3.8 °C lower than for sites lacking permafrost. The SMASCD was generally low at all measured sites with the mean of 0.11 ± 0.06, and was 0.13 ± 0.07 for sites with permafrost and 0.09 ± 0.05 for permafrost-free sites. Finally, permafrost sites had lower $NDVI_{max}$ values (0.40 ± 0.21) than permafrost-free sites (0.52 ± 0.22).

In general, the proportion of permafrost presence was strongly affected by changes in MAAT, SMASCD and $NDVI_{max}$ (Figure 4). At the measured sites, permafrost was more prevalent in areas with low MAAT; few permafrost-affected sites (proportion < 0.06) were found in areas with MAAT higher than −2 °C. The proportion of permafrost sites was positively correlated with the SMASCD, and was about 0.19 higher for sites with SMASCD of 0.20–0.25 than for sites with SMASCD below 0.05. For the $NDVI_{max}$, the proportion of permafrost was about 0.83 for $NDVI_{max}$ between 0.2 and 0.3, but was about 0.41 lower for sites with $NDVI_{max}$ between 0.7 and 0.8. The low proportion of permafrost presence for sites with $NDVI_{max}$ < 0.1 is thought to reflect the abundance of non-permafrost sites in desert and in areas around lakes and river channels. As a consequence, the *in situ* sites showed greater presence of permafrost at lower MAAT and $NDVI_{max}$ and higher SMASCD. These results indicated that predicting PZI from MAAT, SMASCD and $NDVI_{max}$ from a large amount of field evidence appears to be promising.

TABLE 2 Summary statistics of explanatory variables for the aggregated inventory of permafrost presence or absence

Variables	Range	Mean	Median	SD	Q1	Q3
MAAT (°C)	−9.6 to 16.6	−3.5	−4.1	3.0	−5.2	−2.9
SMASCD	0.00–0.58	0.11	0.10	0.06	0.07	0.14
$NDVI_{max}$	0.01–0.88	0.45	0.48	0.22	0.24	0.65

Notes. SD = standard deviation, Q1 = 25th percentile, Q3 = 75th percentile.

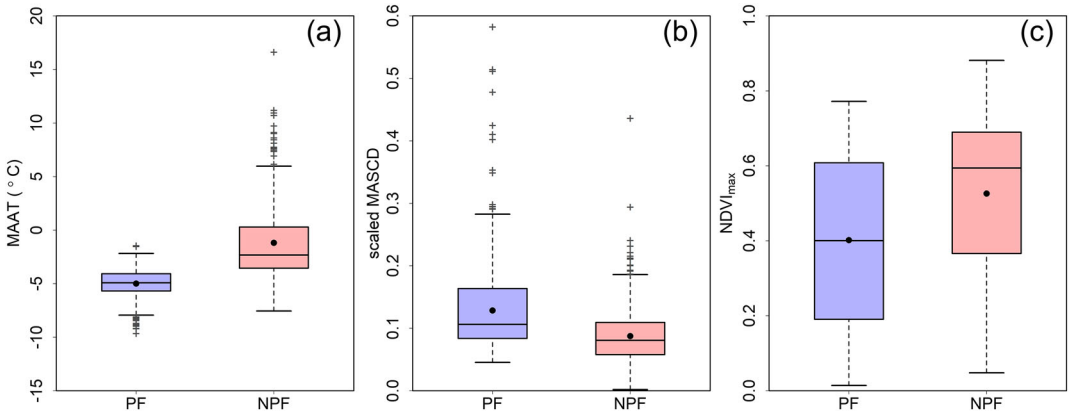

FIGURE 3 Comparisons of (a) MAAT, (b) SMASCD and (c) $NDVI_{max}$ for permafrost presence (PF) and absence (NPF) sites. Black dots are mean values

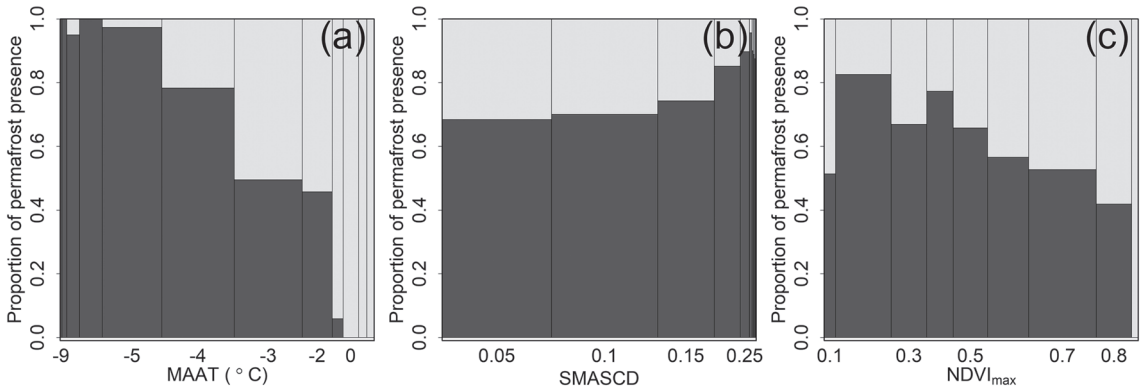

FIGURE 4 Spinogram depicting the proportion of permafrost presence at the measured sites for different binned values of (a) MAAT, (b) SMASCD and (c) $NDVI_{max}$. Bar widths are proportional to the number of observations within any given interval of a predictor variable

4.2 | Evaluation and interpolation of PZI model

The preliminary model results showed that MAAT, SMASCD, $NDVI_{max}$, and the interaction term of MAAT · $NDVI_{max}$ were all statistically significant with $p < 0.01$ (Table 3). The lack of significance for the interaction term between MAAT and SMASCD is probably due to the fact that snow distribution is controlled by precipitation rather than air temperature in the (semi-)arid QTP. In the final model, this non-significant interaction term was removed. It is not surprising that PZI is negatively correlated with MAAT. More interestingly, PZI is positively correlated with the SMASCD. This is thought to be because there is little snow in most areas of the QTP; the maximum mean daily snow depth is 1 cm based on ground observations (Figure 5a). As a result, snow is more effective at cooling the ground by consuming latent heat as it melts and by minimizing the effects of incoming solar radiation rather than acting as an effective insulator of the underlying soil.[34,46] The timing of snow cover is also relevant here because the ground may not be snow-free until the end of May (Figure 5b). The $NDVI_{max}$ predictor variable exhibits a clear spatial pattern (Figure 2c). Values for $NDVI_{max}$ decrease from the southeastern QTP (low and warm) to the northeastern QTP (high and cold). In other words, higher $NDVI_{max}$ values are found in warm areas with less permafrost. This is reflected in the model as a negative coefficient on the $NDVI_{max}$ term (Table 3).

TABLE 3 Parameters of the permafrost zonation index model over the Qinghai–Tibet Plateau

Predictor	Preliminary Model	Final Model
Intercept	2.227**	−1.881 ±0.348**
MAAT	−0.801**	−0.707 ±0.097**
SMASCD	10.459*	5.144 ±0.789*
$NDVI_{max}$	−6.261**	−5.811 ±0.792**
MAAT · SMASCD	not significant	–
MAAT · $NDVI_{max}$	−1.190**	−1.093 ±0.184**

Notes. The model parameters are derived using all the 1040 aggregated field evidence. Significance of Wald test: **<0.01, *<0.05. Non-significant means the predictor variable is involved in the model fitting, but the result is not significant, while a dash means not involved in. The values after ± are standard deviations derived from 10-fold cross-validation.

Based on the estimated model coefficients (Table 3), the sensitivities of PZI to each of the explanatory variables are strongly dependent on the absolute values of those variables due to non-linear processes (Figure 6a,b,c). For example, the PZI is relatively unaffected by changes

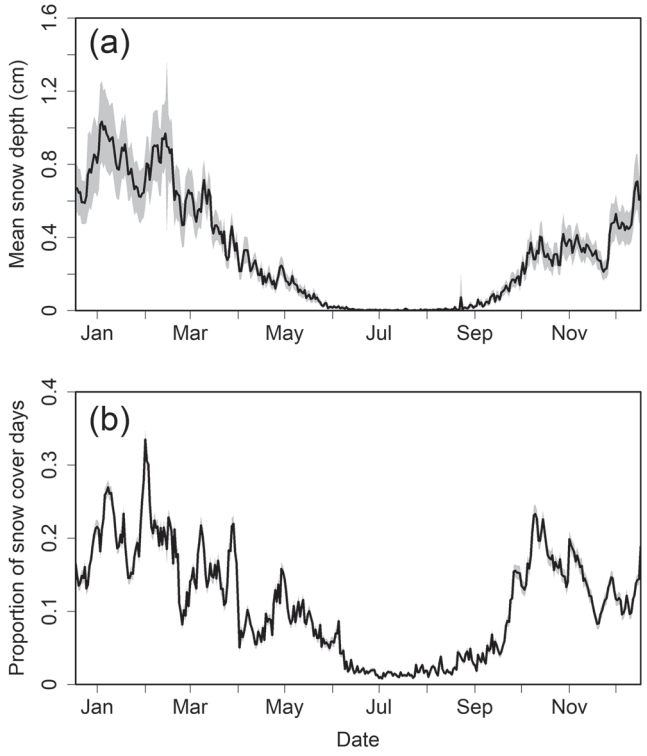

FIGURE 5 (a) Mean daily snow depth at 87 stations of China Meteorological Administration (CMA) during 1981–2010 and (b) mean daily snow cover proportion at 1040 permafrost presence or absence evidence sites during 2003–2010. The distribution of CMA sites is shown in Figure B1a. The gray shadow represents a 95% confidence interval about the mean

in MAAT for extremely cold areas (MAAT < −8 °C) or warm areas (MAAT > 0 °C), but it is highly sensitive to changes in MAAT when MAAT is between −4 and −2 °C. As a result, the PZI only changes by 0.003 when MAAT increases from −10 °C to −8 °C, but the same 2 °C increase from −4 °C to −2 °C corresponds to a decrease of 0.536 in the PZI assuming the $NDVI_{max}$ and SMASCD are equal to their mean values across all measured sites (Figure 6d, Table 2). Similarly, an increase in SMASCD of 0.2 (from 0.01 to 0.21) corresponds to an increase of 0.247 in PZI (assuming MAAT and $NDVI_{max}$ are equal to their mean values), and a change in $NDVI_{max}$ of 0.2 (from 0.35 to 0.55) corresponds to a difference of 0.097 in PZI (assuming MAAT and SMASCD are equal to their mean values). In other words, a change in $NDVI_{max}$ from 0.15 to 0.55 has the same effect as an increase in MAAT from −4.0 °C to −3.34 °C or a decrease in SMASCD from 0.20 to 0.05 assuming the other variables are equal to their mean values across all measured sites.

The mean PZI at sites with evidence for permafrost was 0.81. Most of these sites (89.6%) were located in areas with PZI ⩾ 0.5 (Table 4, Figure 7). Among the 406 permafrost absence sites, there were 120 sites for which PZI ⩽ 0.01 or without PZI in the map, which are predicted to be regions of permafrost absence. There were an additional 167 sites with evidence for no permafrost where PZI < 0.5 with mean of 0.23 ± 0.18. This results in 71.2% of the permafrost absence sites being located in areas without PZI or with PZI < 0.5. In order to compare the predicted PZI values to field data, all field sites were first partitioned into 10 groups based on the PZI derived from 10-fold cross-validation. The measured proportion of permafrost presence was calculated in each partition as the number of sites with evidence for permafrost divided

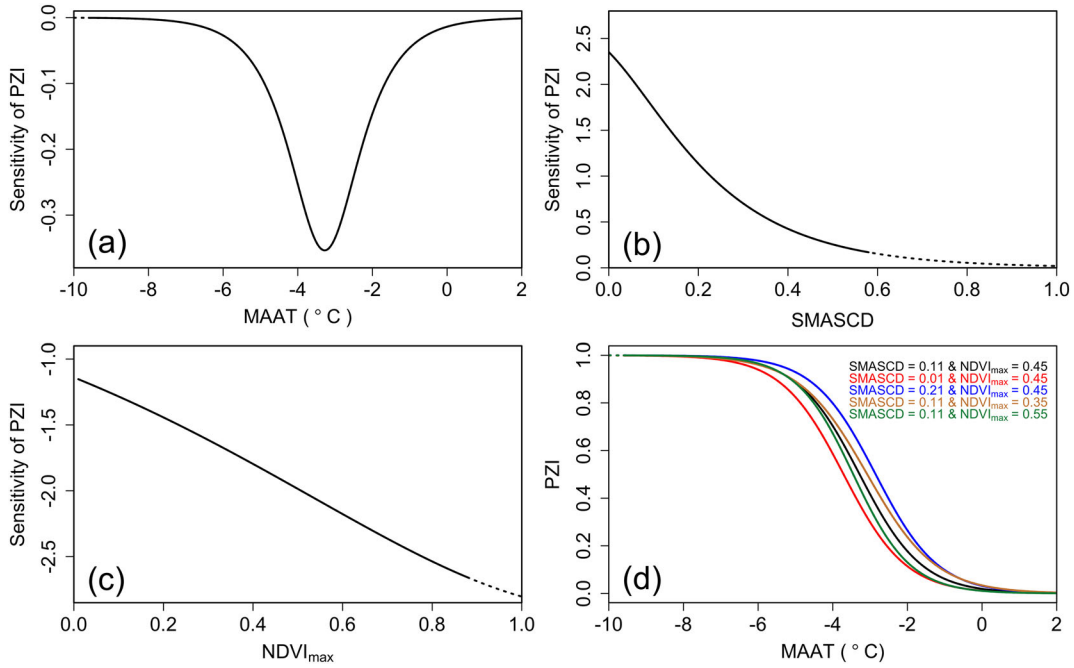

FIGURE 6 Sensitivities of modeled PZI to the explanatory variables of (a) mean annual air temperature (MAAT), (b) scaled mean annual snow cover days (SMASCD) and (c) maximum normalized difference of vegetation index ($NDVI_{max}$). Changes of (d) modeled PZI with the explanatory variables. The dashed lines are extrapolations beyond the measured ranges of explanatory variables (Table 2)

TABLE 4 Summary and evaluation of the PZI_{QTP} and existing permafrost maps over the Qinghai–Tibet Plateau

Name	PZI_{QTP}	PZI_{global}			QTP_{TTOP}
		normal	warm	cold	
Method	statistical	heuristic-empirical			semi-physical
PCC_{PF} (%)	89.6	74.6	34.1	94.0	93.6
PCC_{NPF} (%)	71.2	85.4	99.0	59.3	63.6
PCC_{tol} (%)	82.5	78.8	59.2	80.6	82.0
κ	0.62	0.57	0.28	0.57	0.60
Permafrost region (10^6 km^2)	1.54 (1.35–1.66)	1.68	1.42	1.84	–
Permafrost area (10^6 km^2)	1.17 (0.95–1.35)	1.00	0.76	1.25	1.06 ± 0.09

Notes. The results are slightly different from those in Cao et al.[17] as the evaluation present here is derived based on 1040 aggregated *in situ* measurements.

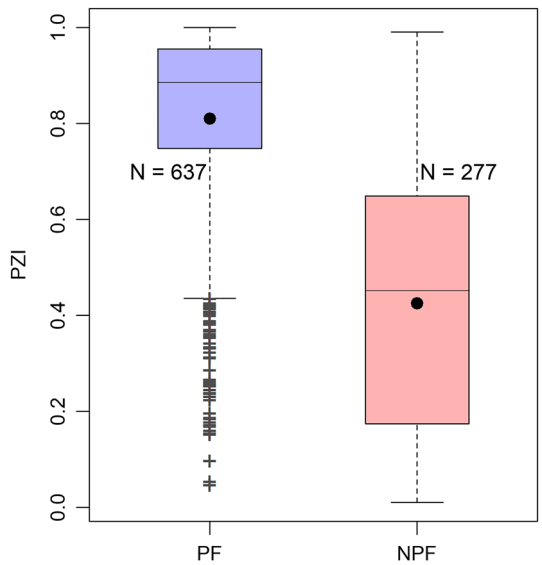

FIGURE 7 Summary of predicted PZI for sites with evidence for permafrost presence (PF) and permafrost absence (NPF) based on 10-fold cross-validation. Data correspond to PZI values in the range [0.01–1] as present in the PZI_{QTP} map

by the total number of permafrost presence sites in the partition. A comparison of measured proportion and predicted proportion (PZI) of permafrost occurrence is presented in Figure 8a. In general, the modeled PZI shows good agreement with the measured proportion with an R^2 of 0.94, an RMSE of 0.08 and a BIAS of 0.01.

4.3 | Computation and statistics of PZI map

The final PZI_{QTP} map, with a spatial resolution of 30 arc seconds (~ 1 km) over the entire QTP, is shown in Figure 9. Because permafrost distribution is expressed as the continuous variable PZI from 0.01 to 1, the PZI_{QTP} map provides more information compared to the previous binary classification maps which show only presence or absence. The PZI_{QTP} map shows a clear spatial pattern of permafrost distribution over the QTP with higher PZI in colder, more regularly snow-covered, and less vegetated areas, as predicted by the statistically based model. A higher PZI value indicates the tendency for permafrost to be (a) more widespread (or more likely to be present at one particular location), (b) colder, (c) thicker and (d) more persistent at depth in response to warming.

Excluding glaciers and lakes, the permafrost region predicted by the PZI_{QTP} map is approximately 1.54 (1.35–1.66) ×10^6 km^2 or 60.7 (54.24–65.2)% of the QTP, while areas underlain by permafrost are predicted to be about 1.17 (0.95–1.35) ×10^6 km^2 or 46.0 (37.3–53.0)% of the QTP (Table 5, Figure 10). The permafrost areas are mainly distributed between elevations of 3600 and 6000 m (98.0–98.8%), and 22.1–23.4% of the entire permafrost area occurs in the elevation range between 4900 and 5100 m. The permafrost area was found to be about 1.16 (0.93–1.34) ×10^6 km^2 for PZI ⩾ 0.1, 1.06 (0.81–1.26) ×10^6 km^2 for PZI ⩾ 0.5 and 0.76 (0.50–1.00) ×10^6 km^2 for PZI ⩾ 0.9, excluding glaciers and lakes.

5 | DISCUSSION

5.1 | Comparison with existing maps

We conducted a cross-validation to assess the PZI_{QTP} map performance, and the other two reference maps are evaluated by directly using the aggregated 1040 ground-truth measurements. In general, the PZI_{QTP} map achieved the best performance for permafrost distribution over the QTP. Compared to the three cases of the PZI_{global} map, PCC_{tol} was 1.9–23.3% higher and κ was 0.05–0.34 higher for the PZI_{QTP} map. Results were similar for the QTP_{TTOP} map (Table 4). While the summary statistics indicated the best agreement of the PZI_{QTP} map with observations, Figure 8 showed the improvements are comprehensive by comparing the modeled PZI against *in situ* measurements for different PZI ranges. The PZI simulated in this study has the lowest RMSE (reduced by 0.06–0.22) and BIAS (reduced by 0.07–0.25) compared to the PZI_{global} map. Additionally, the PZI_{QTP} map in general shows good agreement with all the ten groups, while obvious bias was found for all the three cases of the PZI_{global} map. The worst performance of the warm case in the PZI_{global} map was

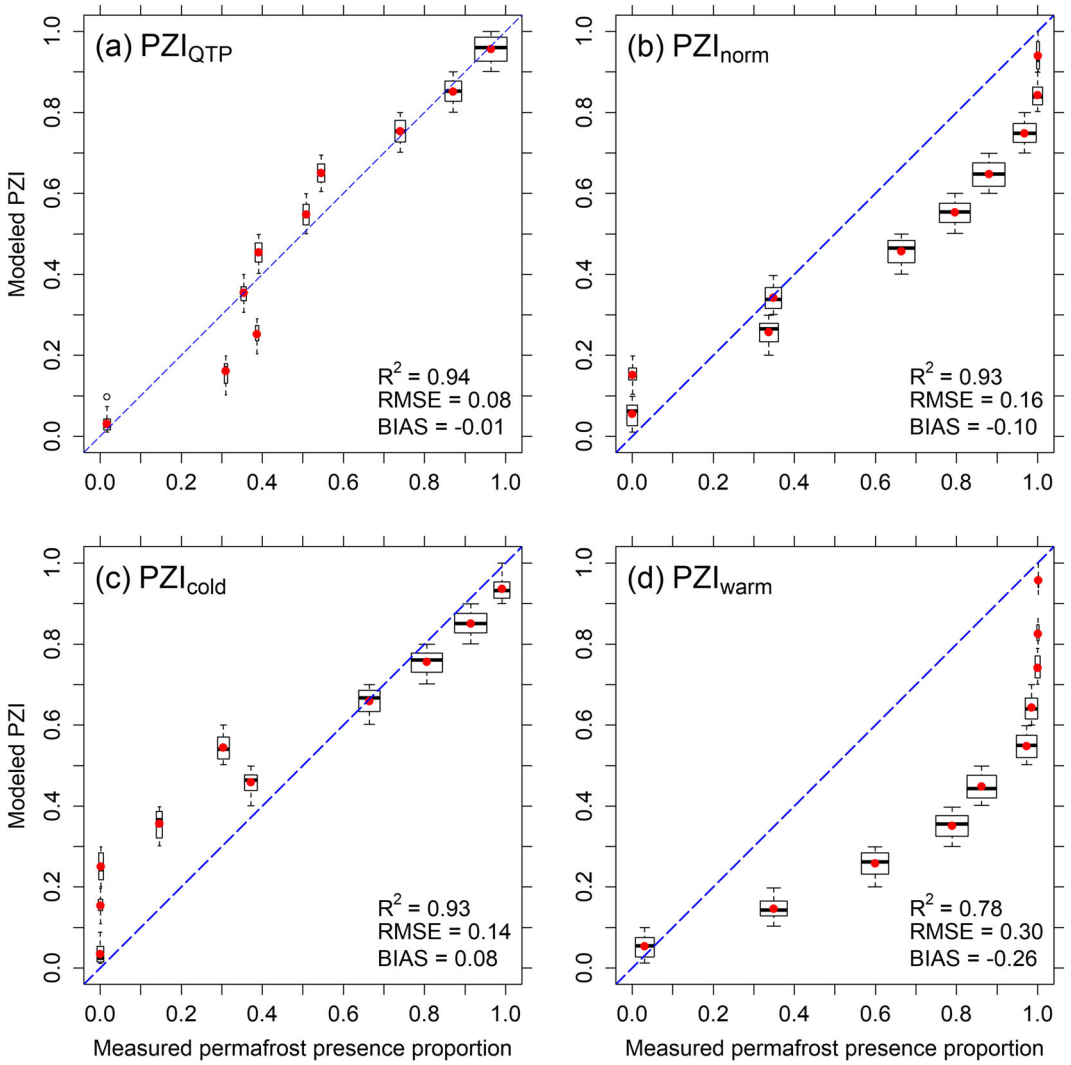

FIGURE 8 Measured permafrost presence proportion against predicted PZI for the (a) PZI_{QTP} map based on 10-fold cross-validation, (b) PZI_{norm}, (c) PZI_{cold}, and (d) PZI_{warm} maps. Norm, cold and warm represent the three cases of the PZI_{global} map. The evidence is grouped by predicted PZI (y-axis) with an interval of 0.1. Each group may contain multiple evidence, and the amount of evidence is represented by the box width. Corresponding to the PZI range of the map, only sites with PZI > 0.01 are present. R^2, RMSE and BIAS statistics are derived using the permafrost presence proportion and mean PZI (the red dot) of all evidence in each PZI group

demonstrated by the significantly warm bias compared with the measured proportion of permafrost presence for all groups and by the remarkable negative BIAS of −0.26. The normal case underestimated the PZI for the groups with PZI > 0.5 (BIAS = −0.10), while the cold case overestimated the PZI for the groups with PZI < 0.5 (BIAS = 0.08).

The improved performance of the PZI_{QTP} map can be explained by the more suitable input data and by the model with calibrated parameters and multiple explanatory variables used. Based on the evaluation results, 195 (132) of the 1040 aggregated sites are classified differently between the PZI_{QTP} map and the PZI_{norm} (PZI_{cold}) map (Figure 11) and 85% (81%) of the PZI difference could be traced back to the MAAT difference (Figure 12a). Our results showed that the MAAT used in the PZI_{QTP} map has better performance than that in the PZI_{global} map, and the RMSE was reduced by 0.57 °C (Figure A1). The PZI_{QTP} map consequently benefits from the improved MAAT. For sites with conflicting classifications but with similar MAAT values (differing by less than 0.2 °C), the PZI differences are caused by the snow cover days and vegetation. The differences in PZI values between the PZI_{QTP} map and PZI_{norm} map are greater for sites with larger SMASCD (Figure 12b), which is consistent with observed and modeled cooling effects of shallow snow.[27] The lack of a significant correction between PZI difference of the PZI_{QTP} map and the PZI_{cold} map and SMASCD is probably due to the narrow span of SMASCD for sites with conflicting classification. Similarly, strong correlations between $NDVI_{max}$ and PZI difference are found for those sites with similar MAAT with the R^2 of 0.49 and 0.81 for the PZI_{norm} and PZI_{cold} maps, respectively. Consequently, the overall accuracy (PCC_{tol}) of the sites with conflicting classifications between the PZI_{QTP} map and PZI_{norm} (PZI_{cold}) map is 60.0% (57.6%) for the PZI_{QTP} map, and 40.0% (42.4%) for the PZI_{norm} (PZI_{cold}) map. Given the locations of these sites (Figure 11), our results highlight how the PZI_{QTP} map could improve the classification accuracy of areas near the permafrost

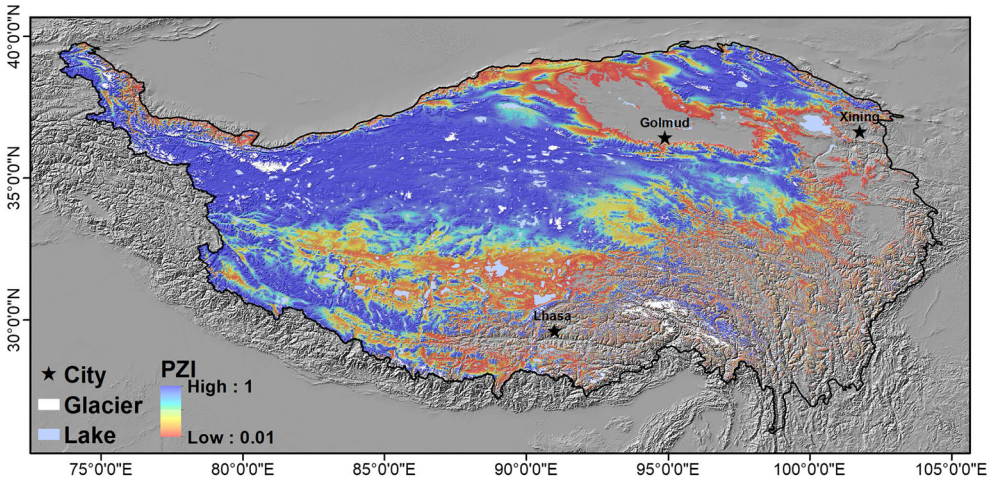

FIGURE 9 Distribution of statistically based PZI over the QTP with a spatial resolution of 30 arc seconds, superimposed on a DEM hillshade

TABLE 5 Summary statistics for binned permafrost zonation index ranges in the Qinghai–Tibet Plateau

PZI range	Total area (10^6 km^2)	Permafrost area (10^6 km^2)	Relative area (%)
0.01–0.1	0.27 (0.22–0.34)	0.01	10.7 (8.8–13.5)
0.1–0.5	0.35 (0.29–0.44)	0.10 (0.08–0.12)	13.9 (11.3–17.2)
0.5–0.9	0.41 (0.36–0.44)	0.30 (0.26–0.32)	16.1 (14.0–17.1)
0.9–1.0	0.78 (0.51–1.02)	0.76 (0.50–1.00)	30.7 (20.1–40.0)
Total	1.82 (1.73–1.88)	1.17 (0.95–1.35)	71.4 (68.0–74.0)

Notes. Total area is the area with simulated PZI occuring in the given range and permafrost area was derived as PZI multiplied by pixel area. Relative area refers to the total area of the Qinghai–Tibet Plateau. The extent range of permafrost region and area are derived based on the parameter strand deviation present in Table 3. The statistics are conducted by excluding glaciers and lakes.

boundary by considering more local factors controlling permafrost distribution. This is especially important on the QTP because much of the bias of permafrost maps in areas near permafrost boundaries is caused by the sensitivity of permafrost to near-surface air temperature and the strong influences of local factors.[17]

The total permafrost area calculated using the QTP_{PZI} map is similar to that of the QTP_{TTOP} map (1.06 ± 0.09 × 10^6 km^2), which has been reported to be the best calibrated categorical map over the QTP. The slightly smaller permafrost area (about 0.11 × 10^6 km^2) for the QTP_{TTOP} map is probably due to the land surface temperature used, which represents the conditions during 2003–2012, and is expected to be higher than the reference period of 1979–1999 used for the PZI_{QTP} map. The QTP_{TTOP} map performs slightly better in permafrost areas than the PZI_{QTP} map, as reflected by the higher PCC_{PF} (+4%). However, this is offset by poorer performance in non-permafrost areas with lower PCC_{NPF} (−7.6%). As a whole, the PZI_{QTP} map showed slightly better performance than the QTP_{TTOP} map.

5.2 | Uncertainties of PZI map

Although the PZI model over the QTP was estimated and evaluated by a large number of *in situ* measurements (1040 points aggregated from

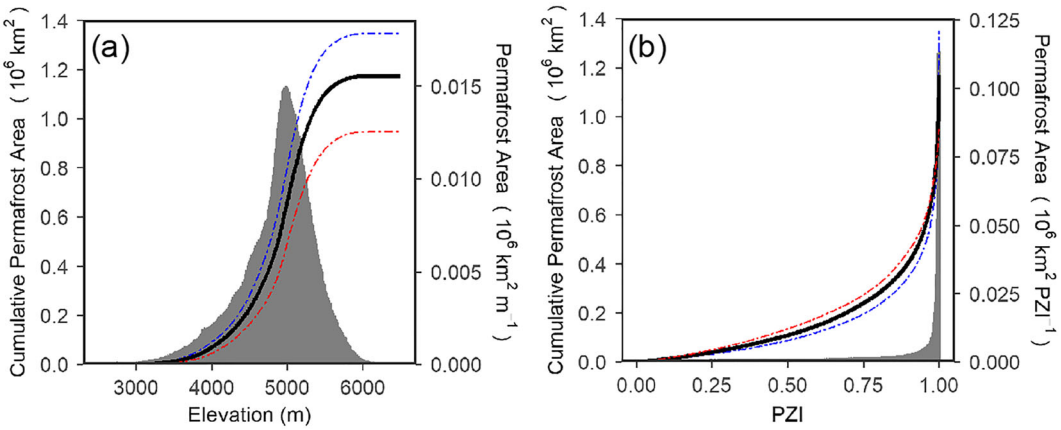

FIGURE 10 Cumulative permafrost region over the range of (a) elevation and (b) PZI. The black lines are the cumulative permafrost area, and the histogram is the permafrost area for different elevations and PZIs. The dashed lines are the maximum (blue) and minimum (red) cumulative permafrost areas, respectively, based on the standard deviation of final model in Table 3

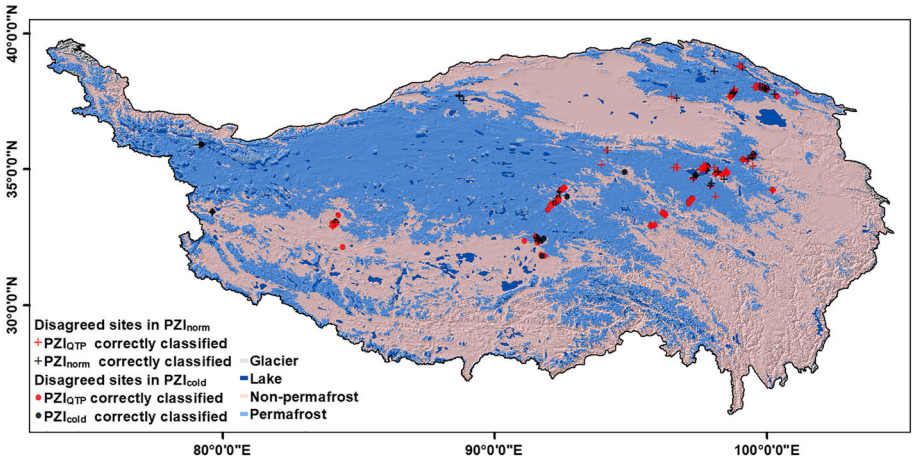

FIGURE 11 A map of sites for which classifications are contradictory between the PZI_{QTP} map and either PZI_{norm} map (gray) or PZI_{cold} (blue) map, superimposed on the background of the categorical permafrost map developed by Zou et al.[10]

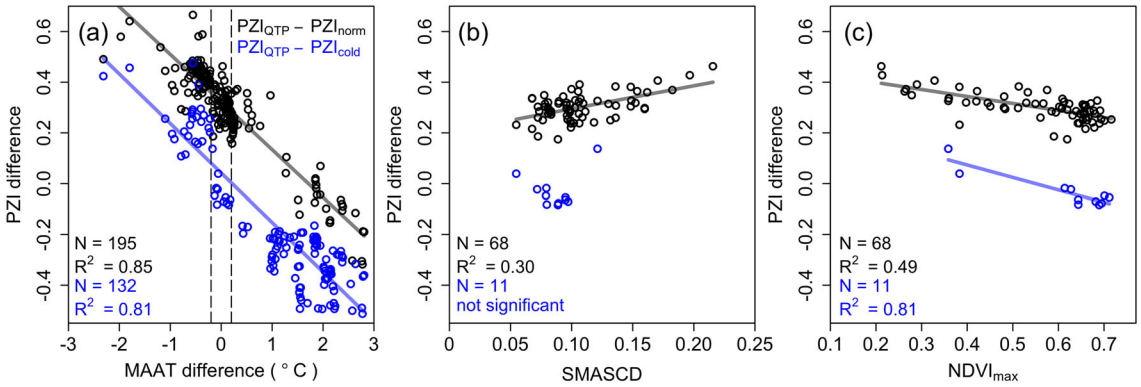

FIGURE 12 A detailed analysis of sites for which classifications are contradictory between the PZI_{QTP} map and either the PZI_{norm} (gray) or PZI_{cold} (blue). The three panels plot the difference in PZI values in relation to the difference in: (a) MAAT, (b) SMASCD and (c) $NDVI_{max}$. (b) and (c) only include points with an MAAT difference of less than 0.2 ° C between the two maps; this interval is marked in (a) with dashed lines. N is the number of *in situ* sites with contradictory classifications. The warm case of the PZI_{global} map is not included here as it is not suitable for use over the QTP

1475 measurements) and our results showed significant relationships between PZI and the selected predictor variables (Table 3), the validity of the relationship outside the sample cannot be evaluated in this study. The PZI_{QTP} map is based on statistical results, and hence its accuracy is influenced by the representativeness of the sample sites. However, the sample sites are biased towards locations along main roads (e.g. Qinghai–Tibet highway and Erga Road in the northeastern QTP), and observations are scarce in northwestern areas due to the harsh climate and lack of access. The PZI_{QTP} map is expected to have larger uncertainty caused by the inadequate sampling, especially at high elevation (> 5200 m) and in steep (slope ⩾ 20°) terrain.

Permafrost presence or absence is site-specific and usually depends not only on the predictor variables used here but also on many additional factors such as radiation,[47,48] precipitation[20] and subsurface soil conditions,[27] especially the presence of a peat layer. For example Cao et al.[28] reported that the permafrost lower limit is about 300 m lower on north-facing slopes than on south-facing slopes due to reduced solar radiation and the presence of a peat layer. Furthermore, lower ground temperatures were reported in alpine swamp meadow areas due to the higher soil moisture and organic matter compared to typical alpine meadows.[49] In this case, the presence of swamp meadow is typically indicative of colder ground and a higher likelihood of permafrost, regardless of the PZI value. While snow is minimal in most parts of the QTP, the snow redistribution phenomenon is significant due to blowing snow.[50] In snowdrift areas, this may negate the snow cooling effect mentioned above, and may even cause snow to act as a strong insulator and warm the ground.[34] A sub-grid interpretation scheme based on expert assessments can therefore be important and useful to further improve the prediction of the PZI_{QTP} map presented here.[21]

Because the PZI is highly sensitive to the predictor variables (Figure 6), another component of uncertainty can be attributed to the input data sets. Although lapse rate is well described in MAAT downscaling, surface effects that can affect fine-scale simulations are simplified.[31] Cold air pooling is one such effect that may overestimate the near-surface air temperature in valleys and consequently underestimate the permafrost presence. Furthermore, although the SMASCD product used here was characterized by high classification accuracies, the algorithm used to derive it, known as the snow-line method (SNOWL),

depends strongly on elevation.[35] Thus, the influences of latitude and topography are ignored. As a consequence, SMASCD was overestimated in southern areas and on south-facing slopes and was underestimated in northern areas and on north-facing slopes. However, it is challenging to quantitatively assess the impact of this on the PZI map because the uncertainty arises from the model development phase. We hope more reliable data sets can be developed over the QTP that help to establish more accurate permafrost maps in the future.

The distribution and thermal state of permafrost are the results of long-term climate. The temporal evolution of permafrost and the associated effects on the present-day spatial distribution (e.g. degradation and relict permafrost) are not reflected by the PZI_{QTP} map, which is based on an assumption of equilibrium conditions.

6 | CONCLUSION

We defined a PZI model based on downscaled re-analysis air temperature, remotely sensed snow cover and vegetation data sets, and the new developed inventory of permafrost presence or absence evidence. This model was used to create a new permafrost map of the QTP (the PZI_{QTP}), which was evaluated alongside two other permafrost maps. The new model and map presented here achieved the best performance with an overall classification accuracy of 82.5% and a κ of 0.62.

Besides providing more reliable predictions of permafrost distribution, the PZI_{QTP} map allows users to further distinguish permafrost presence or absence in the field based on *in situ* conditions such as the presence of peat cover. This is possible because the map is expressed as a continuous index resembling the probability of permafrost distribution for a given area.

Excluding glaciers and lakes, the permafrost region over the QTP was found to be approximately 1.54 (1.35–1.66) $\times 10^6$ km^2, or 60.7 (54.5–65.2)% of the exposed land area of the QTP, while areas underlain by permafrost were found to cover about 1.17 (0.95–1.35) $\times 10^6$ km^2, or 46.0 (37.3–53.0)%.

Spatial sampling bias is expected to cause higher uncertainty in steep terrain and at high elevation. Improved model evaluation will require further investigation using more systematic samples.

ACKNOWLEDGEMENTS

The authors would like to thank two anonymous reviewers and Stephan Gruber for their constructive suggestions. We thank Wanwan Shao for help in model sensitivity analyses. This study was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20100313, XDA20100103), the National Natural Science Foundation of China (41871050, 41801028), and partly by the Fundamental Research Funds for the Central Universities (lzujbky_2016_281, 862863).

The mean annual air temperature of the PZI_{global} map is from http://www.geo.uzh.ch/microsite/cryodata/pf_global/. We thank CMA (http://data.cma.cn/) for providing the near-surface air and ground surface temperatures, and thank ECMWF for the ERA-Interim re-analysis data. The MOD13Q1 NDVI data set was retrieved from USGS/Earth Resources Observation and Science (EROS) Center (https://lpdaac.usgs.gov/data_access/data_pool) and processed based on Google Earth Engine (GEE), the ASTER data set was downloaded from United States Geological Survey (http://gdex.cr.usgs.gov/gdex/), glacier inventory and snow cover product were provided by the Environmental and Ecological Science Data Center for West China and the National Natural Science Foundation of China (http://westdc.westgis.ac.cn/), and lake inventory was from the Third Pole Environment Database (http://www.tpedatabase.cn).

DATA AVAILABILITY

The predictor variables including MAAT, SMASCD and $NDVI_{max}$, and the PZI map are available as ASCII files in the Supporting Information.

COMPETING INTERESTS

The authors declare that no competing interests are present.

AUTHOR CONTRIBUTION

BC carried out this study by organizing the field measurements, compiling the inventory of permafrost presence or absence evidence, analyzing data, developing the model and by structuring as well as writing the paper. TZ guided the research. QW, YS, LZ and DZ contributed to organize the inventory of permafrost presence or absence evidence.

ORCID

Bin Cao https://orcid.org/0000-0003-2473-2276
Qingbai Wu https://orcid.org/0000-0002-7965-0975

REFERENCES

1. Immerzeel WW, van Beek LPH, Bierkens MFP. Climate change will affect the asian water towers. *Science*. 2010;328:1382-1385.
2. Wu Q, Zhang T. Recent permafrost warming on the Qinghai-Tibetan Plateau. *J Geophys Res-Atmos*. 2008;113(D13):D13108.
3. Zhao L, Wu Q, Marchenko S, Sharkhuu N. Thermal state of permafrost and active layer in central Asia during the international polar year. *Permafr Periglac Process*. 2010;21(2):198-207.
4. Guo D, Yang M, Wang H. Characteristics of land surface heat and water exchange under different soil freeze/thaw conditions over the central Tibetan Plateau. *Hydrol Process*. 2011;25(16):2531-2541.
5. Bibi S, Wang L, Li X, Zhou J, Chen D, Yao T. Climatic and associated cryospheric, biospheric, and hydrological changes on the Tibetan Plateau: A review. *Int J Climatol*. 2018;38(S1):e1-e17.
6. Mu C, Zhang T, Wu Q, et al. Editorial: Organic carbon pools in permafrost regions on the Qinghai–Xizang (Tibetan) Plateau. *The Cryosphere*. 2015;9(2):479-486.
7. Gruber S, Haeberli W. Permafrost in steep bedrock slopes and its temperature-related destabilization following climate change. *J Geophys Res*. 2007;112:F02S18.
8. Wu Q, Zhang Z, Gao S, Ma W. Thermal impacts of engineering activities and vegetation layer on permafrost in different alpine ecosystems of the Qinghai–Tibet Plateau, China. *The Cryosphere*. 2016;10(4):1695-1706.

9. Cao B, Gruber S, Zhang T, et al. Spatial variability of active layer thickness detected by ground-penetrating radar in the qilian mountains, western china. *J Geophys Res Earth Surf*. 2017;122(3):574-591.
10. Zou D, Zhao L, Sheng Y, et al. A new map of permafrost distribution on the Tibetan Plateau. *The Cryosphere*. 2017;11(6):2527-2542.
11. Wu X, Nan Z, Zhao S, Zhao L, Cheng G. Spatial modeling of permafrost distribution and properties on the Qinghai—Tibet Plateau. *Permafr Periglac Process*. 2018;29(2):86-99.
12. Nan Z, Li S, Liu Y. Mean annual ground temperature distribution on the Tibetan Plateau: permafrost distribution mapping and further application. *J Glaciol Geocryol*. 2002;24(2):142-148.
13. Li J, Sheng Y, Wu J, Chen J, Zhang X. Probability distribution of permafrost along a transportation corridor in the northeastern Qinghai province of China. *Cold Reg Sci Tech*. 2009;59(1):12-18.
14. Niu F, Yin G, Luo J, Lin Z, Liu M. Permafrost distribution along the Qinghai-Tibet Engineering Corridor, China using high-resolution statistical mapping and modeling integrated with remote sensing and GIS. *Remote Sens*. 2018;10(2):215.
15. Nan Z, Huang P, Zhao L. Permafrost distribution modeling and depth estimation in the Western Qinghai-Tibet Plateau (in Chinese with English abstract). *Acta Geograph Sin*. 2013;68(3):318.
16. Zhao S, Nan Z, Huang Y, Zhao L. The application and evaluation of simple permafrost distribution models on the Qinghai–Tibet Plateau. *Permafr Periglac Process*. 2017;28(2):391-404.
17. Cao B, Zhang T, Wu Q, Sheng Y, Zhao L, Zou D. Brief communication: Evaluation and inter-comparisons of qinghai–tibet plateau permafrost maps based on a new inventory of field evidence. *The Cryosphere*. 2019;13(2):511-519.
18. Riseborough D, Shiklomanov N, Etzelmüller B., Gruber S, Marchenko S. Recent advances in permafrost modelling. *Permafr Periglac Process*. 2008;19(2):137-156.
19. Gruber S. Derivation and analysis of a high-resolution estimate of global permafrost zonation. *The Cryosphere*. 2012;6(1):221-233.
20. Boeckli L, Brenning A, Gruber S, Noetzli J. A statistical approach to modelling permafrost distribution in the European Alps or similar mountain ranges. *The Cryosphere*. 2012;6(1):125-140.
21. Boeckli L, Brenning A, Gruber S, Noetzli J. Permafrost distribution in the European Alps: calculation and evaluation of an index map and summary statistics. *The Cryosphere*. 2012;6(4):807-820.
22. Azócar GF, Brenning A, Bodin X. Permafrost distribution modelling in the semi-arid chilean andes. *The Cryosphere*. 2017;11(2):877-890.
23. Brown J, Ferrians O, Heginbottom JA, Melnikov E. *Circum-Arctic Map of Permafrost and Ground-ice Conditions*. Boulder, CO: NSIDC: National Snow and Ice Data Center; 1997.
24. Schmid M-O, Baral P, Gruber S, et al. Assessment of permafrost distribution maps in the Hindu Kush Himalayan region using rock glaciers mapped in Google Earth. *The Cryosphere*. 2015;9(6):2089-2099.
25. Gruber S, Fleiner R, Guegan E, et al. Review article: Inferring permafrost and permafrost thaw in the mountains of the Hindu Kush Himalaya region. *The Cryosphere*. 2017;11(1):81-99.
26. Wang K, Zhang T, Zhong X. Changes in the timing and duration of the near-surface soil freeze/thaw status from 1956 to 2006 across China. *The Cryosphere*. 2015;9(3):1321-1331.
27. Zhou J, Kinzelbach W, Cheng G, Zhang W, He X, Ye B. Monitoring and modeling the influence of snow pack and organic soil on a permafrost active layer, Qinghai–Tibetan Plateau of China. *Cold Reg Sci Tech*. 2013;90-91:38-52.
28. Cao B, Zhang T, Peng X, et al. Thermal characteristics and recent changes of permafrost in the upper reaches of the Heihe river basin, western China. *J Geophys Res-Atmos*. 2018;123(15):7935-7949.
29. Guo D, Wang A, Li D, Hua W. Simulation of changes in the near-surface soil freeze/thaw cycle using clm4.5 with four atmospheric forcing data sets. *J Geophys Res-Atmos*. 2018;123(5):2509-2523.
30. Bao X, Zhang F. Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau. *J Climate*. 2013;26(1):206-214.
31. Cao B, Gruber S, Zhang T. REDCAPP (v1.0): Parameterizing valley inversions in air temperature data downscaled from reanalyses. *Geosci Model Dev*. 2017;10(8):2905-2923.
32. Fiddes J, Gruber S. Toposcale v.1.0: downscaling gridded climate data in complex terrain. *Geosci Model Dev*. 2014;7(1):387-405.
33. Gao L, Bernhardt M, Schulz K. Elevation correction of ERA-Interim temperature data in complex terrain. *Hydrol Earth Syst Sci*. 2012;16(12):4661-4673.
34. Zhang T. Influence of the seasonal snow cover on the ground thermal regime: An overview. *Rev Geophys*. 2005;43(4):RG4002.
35. Wang W, Huang X, Deng J, Xie H, Liang T. Spatio-temporal change of snow cover and its response to climate over the Tibetan Plateau based on an improved daily cloud-free snow cover product. *Remote Sens*. 2015;7(1):169-194.
36. Tachikawa T, Hato M, Kaku M, Iwasaki A. Characteristics of ASTER GDEM version 2. In: 2011 IEEE International Geoscience and Remote Sensing Symposium; 2011; Vancouver, BC, Canada:3657-3660.
37. Meyer D, Tachikawa T, Kaku M, et al. Aster global digital elevation model version 2–summary of validation results. *Japan-US ASTER Science Team*. 2011:1-26.
38. Conrad O, Bechtel B, Bock M, et al. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci Model Dev*. 2015;8(7):1991-2007.
39. Horn BKP. Hill shading and the reflectance map. *Proc IEEE*. 1981;69(1):14-47.
40. Zhang Y, Li B, Zheng D. A discussion on the boundary and area of the Tibetan Plateau in China. *Geogr Res*. 2002;21(1):1.
41. Liu S, Yao X, Guo W, et al. The contemporary glaciers in China based on the Second Chinese Glacier Inventory (in Chinese with English abstract). *Acta Geograph Sin*. 2015;70(1):3.
42. Zhang T, Heginbottom JA, Barry RG, Brown J. Further statistics on the distribution of permafrost and ground ice in the Northern Hemisphere. *Polar Geogr*. 2000;24(2):126-131.
43. Gubler S, Fiddes J, Keller M, Gruber S. Scale-dependent measurement and analysis of ground surface temperature variability in alpine terrain. *The Cryosphere*. 2011;5(2):431-443.
44. Zhang T, Armstrong RL, Smith J. Investigation of the near-surface soil freeze-thaw cycle in the contiguous United States: Algorithm development and validation. *J Geophys Res Atmos*. 2003;108(D22):8860.
45. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
46. Jin R, Li X, Che T. A decision tree algorithm for surface soil freeze/thaw classification over China using SSM/I brightness temperature. *Remote Sens Environ*. 2009;113(12):2651-2660.
47. Chang X, Jin H, Zhang Y, et al. Thermal impacts of boreal forest vegetation on active layer and permafrost soils in northern Da Xing'anling (Hinggan) Mountains, Northeast China. *Arct Antarct Alp Res*. 2015;47:47-47 –13.
48. Zhang YL, Li X, Cheng GD, et al. Influences of topographic shadows on the thermal and hydrological processes in a cold region mountainous watershed in northwest China. *J Adv Model Earth Syst*. 2018;10(7):1439-1457.

49. Li J, Sheng Y, Chen J, Wu J, Wang S. Variations in permafrost temperature and stability of alpine meadows in the source area of the Datong River, Northeastern Qinghai-Tibet Plateau, China. *Permafr Periglac Process*. 2014;25(4):307-319. PPP-13-0034.R3.
50. Li G, Wang ZS, Huang N. A snow distribution model based on snowfall and snow drifting simulations in mountain area. *J Geophys Res-Atmos*. 2018;123(14):7193-7203.
51. Gallant JC, Dowling TI. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour Res*. 2003;39(12):1347.
52. Dee DP, Uppala SM, Simmons AJ, et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q J R Meteorol Soc*. 2011;137(656):553-597.
53. Takala M, Luojus K, Pulliainen J, et al. Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sens Environ*. 2011;115(12):3517-3529.
54. Shi J, Xiong C, Jiang L. Review of snow water equivalent microwave remote sensing. *Sci China Earth Sci*. 2016;59(4):731-745.

How to cite this article: Cao B, Zhang T, Wu Q, Sheng Y, Zhao L, Zou D. Permafrost zonation index map and statistics over the Qinghai-Tibet Plateau based on field evidence. *Permafrost and Periglac Process*. 2019;30:178–194. https://doi.org/10.1002/ppp.2006

APPENDIX A

MAAT DOWNSCALING AND EVALUATION

A.1 | Hypsometric position

Hypsometric position is defined for a given site as the ratio of the number of grid cells with higher elevation than that site to the total number of grid cells in a prescribed neighborhood of analysis.[51] It ranges from 1 (deepest valley) to 0 (highest peak). The prescribed neighborhood of analysis is taken as 30 km × 30 km.[31]

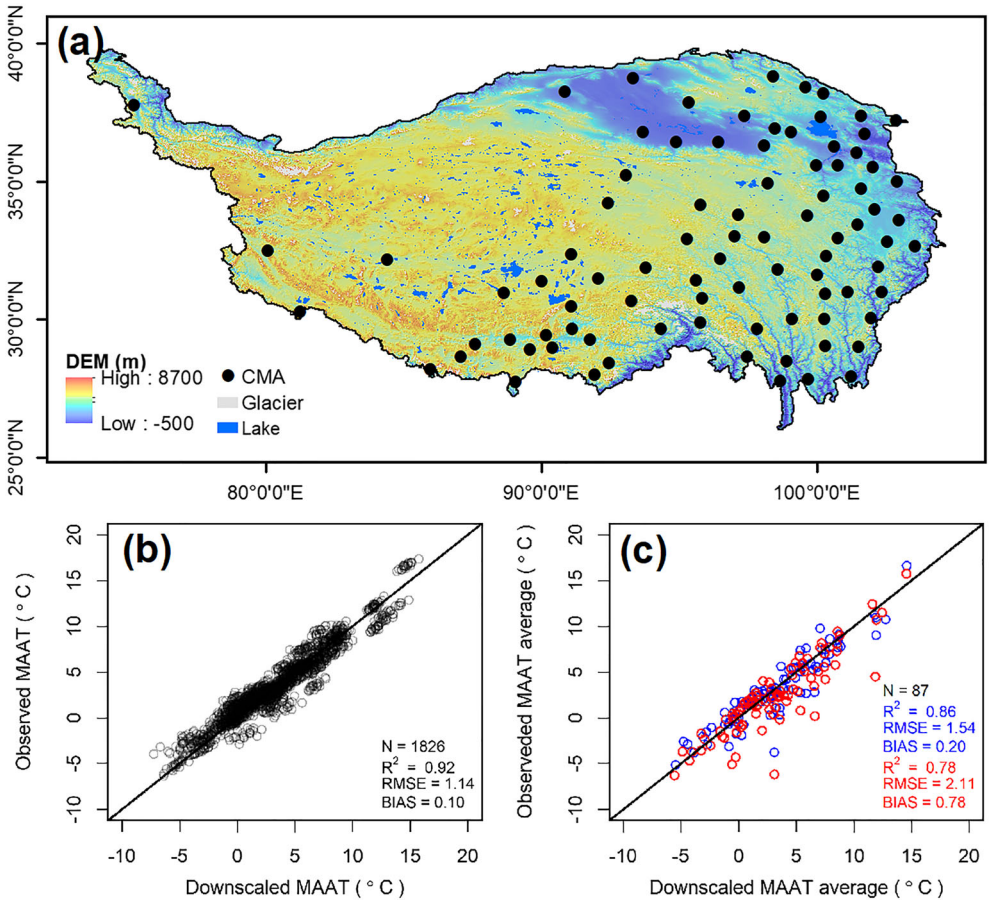

FIGURE A1 (a) The 87 meteorological stations of China Meteorological Administration (CMA) on the Qinghai–Tibet Plateau used for evaluation of downscaled MAAT (1979–1999), (b) downscaled (MOD) MAAT, and the MAAT average used in this study (blue) as well as in the PZI_{global} map (black, mean of 1969–1999). N is the number of observations

A.2 | Downscaled MAAT evaluation

In total, 87 meteorological stations from China Meteorological Administration (CMA) with total of 1826 MAAT observations during 1979–1999 were used for evaluating downscaled MAAT. Figure A1 shows plots of observations against the downscaled MAAT, and demonstrates that the downscaled MAAT is a good approximation of the air temperature over the QTP, with the R^2, RMSE and BIAS of 0.92, 1.14 °C and 0.10 °C, respectively. This line of reasoning may be somewhat circular as some of the observations used here are assimilated in the ERA-Interim,[52] but it nevertheless gives a valuable indication of the quality of the data product after downscaling to about 1 km.

APPENDIX B

SNOW WATER EQUIVALENT EVALUATION

The performance of two most widely used snow water equivalent (SWE) products, GlobSnow (version 2)[53] and ASMR-E,[54] over the Qinghai–Tibet Plateau discussed in this study. As the mountains were masked from GlobSnow due to poor reliability and the SWE was set as 0 mm for the areas below about 35°N, it does not cover the entire QTP (Figure B1a). For AMSR-E, the SWE equivalent was evaluated using the observations of 87 meteorological stations from CMA. The low R^2 (0.02), high RMSE (13.47) and BIAS (−6.55 mm) indicate considerable uncertainty of AMSR-E for the QTP (Figure B1b). As a consequence, no SWE products were used in this study.

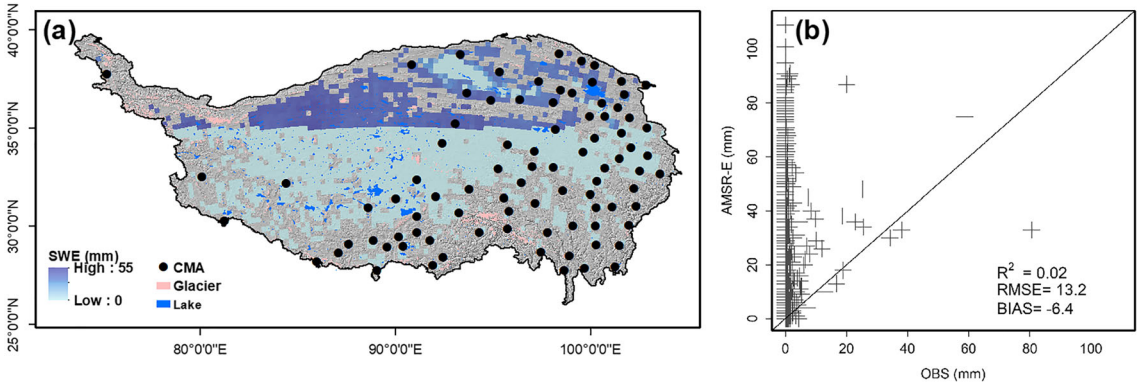

FIGURE B1 (a) Median of annual maximum snow water equivalent (SWE) from GlobSnow2 during 1979–1999 shown on the hillshade and 87 meteorological stations of CMA on the Qinghai–Tibet Plateau used for the evaluation of AMSR-E SWE (2002–2010). (b) Monthly AMSR-E SWE compared to observations (OBS)