New high-resolution estimates of the permafrost thermal state and hydrothermal conditions over the Northern Hemisphere

Youhua Ran[1,3] [*], Xin Li[2,4], Guodong Cheng[1,5], Jingxin Che[6], Juha Aalto[7,8], Olli Karjalainen[9], Jan Hjort[9], Miska Luoto[7], Huijun Jin[1,10], Jaroslav Obu[11], Masahiro Hori[12], Qihao Yu[1,3], Xiaoli Chang[13]

[1]Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China.
[2]National Tibetan Plateau Data Center, Key Laboratory of Tibetan Environmental Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China.
[3]University of Chinese Academy of Sciences, Beijing 100049, China.
[4]CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing 100101, China.
[5]Institute of Urban Study, Shanghai Normal University, Shanghai 200234, China.
[6]School of Science, Nanchang Institute of Technology, Nanchang 330099, China.
[7]Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, Gustaf Hällströmin katu 2a, 00014 Helsinki, Finland.
[8]Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland.
[9]Geography Research Unit, University of Oulu, P.O. Box 8000, FI-90014, Oulu, Finland.
[10]Institute of Cold Regions Science and Engineering and School of Civil Engineering, Northeast Forestry University, Harbin 150040, China
[11]Department of Geosciences, University of Oslo, Postboks 1047 Blindern, 0316 Oslo, Norway.
[12]Earth Observation Research Center, Japan Aerospace Exploration Agency, 2-1-1, Sengen, Tsukuba, Ibaraki 305-8505, Japan.
[13]School of Resource & Environment and Safety Engineering, Hunan University of Science and Technology, Xiangtan 411201, China.

[*]*Correspondence to*: Youhua Ran (ranyh@lzb.ac.cn)

Abstract. Monitoring the thermal state of permafrost is important in environmental science and engineering applications. However, such data are generally unavailable mainly due to the lack of ground observations and the uncertainty of traditional physical models. This study produces novel permafrost datasets for the Northern Hemisphere (NH), including predictions of the mean annual ground temperature (MAGT) at zero annual amplitude depth and active layer thickness (ALT) with a 1-km resolution for the period of 2000–2016, as well as estimates of the probability of permafrost occurrence and permafrost zonation based on hydrothermal conditions. These datasets integrate unprecedentedly large amounts of field data (1,002 boreholes for MAGT and 452 sites for ALT) and multisource geospatial data, especially remote sensing data, using statistical learning modelling with an ensemble strategy. Thus, these data are more accurate than those of previous circumpolar maps (bias=0.02±0.16 ℃, RMSE=1.32±0.13 ℃ for MAGT; bias=2.71±16.46 cm, RMSE=86.93±19.61 cm for ALT). The datasets suggest that the areal extent of permafrost (MAGT≤0 °C) in the NH, excluding glaciers and lakes, is approximately 15.03 (13.84–19.29) $\times 10^6$ km^2, and the areal extent of permafrost regions (permafrost probability>0) is approximately 20.14×10^6 km^2. The areal fractions of humid, semiarid/subhumid, and arid permafrost regions are 51.84%, 44.83%, and 3.33%, respectively. The areal fractions of cold (≤-3.0 ℃), cool (-3.0 ℃ to -1.5 ℃), and warm (>-1.5 ℃) permafrost regions are

37.93%, 14.35%, and 47.72%, respectively. These new datasets based on the most comprehensive field data to date contribute to an updated understanding of the thermal state and zonation of permafrost in the NH. They are potentially useful for various fields, such as in climatology, hydrology, ecology, agriculture, public health, and engineering planning. All of the datasets are

40   published through the National Tibetan Plateau Data Center (TPDC), and the link is https://data.tpdc.ac.cn/en/data/5093d9ff-a5fc-4f10-a53f-c01e7b781368 or https://doi.org/10.11888/Geocry.tpdc.271190 (Ran et al., 2021b).

1 Introduction

Permafrost is defined as ground that remains at or below 0 ℃ for at least two consecutive years (Van Everdingen, 1998). As a key component of the cryosphere in the Northern Hemisphere (NH), permafrost is sensitive to disturbances from climate

45   change, environmental disturbances and human activities (Shur and Jorgenson, 2007). Changes in the thermal state of permafrost (TSP) have multiple important impacts on infrastructure safety, hydrological cycles, ecosystems, and climate systems (Yoshikawa and Hinzman, 2003; Cheng, 2005; Cheng and Wu, 2007; Schuur et al., 2009; Tarnocai et al., 2009; Cheng and Jin, 2013; Gao et al., 2016; Hjort et al., 2018; Ran et al., 2018; Xie et al., 2019; Li et al., 2020a, 2020b).

Over the past half century, hundreds of permafrost maps have been compiled at local to global scales (e.g., Heginbottom et al.,

50   2002; Ran et al., 2012; Cao et al., 2019; Zhelezniak et al., 2021). On a global scale, the first permafrost map, the Circum-Arctic Map of Permafrost and Ground Ice Conditions, was compiled by the International Permafrost Association (IPA) using a manual delineation method by integrating all (readily) available data and regional permafrost maps in the 1990s (Heginbottom et al., 1993; Brown et al., 1997). The IPA permafrost map indicates that the permafrost region underlies an area spanning approximately 22.79×10^6 km^2 in the NH, while the estimated areal extent actually underlain by permafrost ranges

55   from 12.21×10^6 to 16.98×10^6 km^2 (Zhang et al., 2000, 2008). The IPA permafrost map represents the best understanding of the permafrost distribution in the NH in the 1990s and has been frequently used for model evaluation or validation and even to perform derivative permafrost simulations (e.g., Chadburn et al., 2017). However, its quality and consistency vary because it is derived from the experience and knowledge of experts and depends on very limited observations at varied scales and accuracies and limited integration methods of mapping.

60   Since the release of the IPA permafrost map, great advances have been made in monitoring and modelling permafrost. In terms of data, the amounts of both ground-based observations and remote sensing data have remarkably increased in the past 30 years. Global and regional observation networks have been gradually established worldwide, strengthening permafrost monitoring in many regions, including Russia (e.g., Dvornikov et al., 2016), North America (e.g., Romanovsky and Osterkamp, 2001; Smith et al., 2005, 2010), Central Asia and China (e.g., Zhao et al., 2010a, 2021), and Europe (e.g., Harris et al., 2001;

65   Mair et al., 2011; Kellerer-Pirklbauer, 2016). As a result of these regional networks, the Global Terrestrial Network for Permafrost (GTN-P) was established by the IPA in 1999 (Brown et al., 2000, 2008). The GTN-P monitors the TSP and active layer thickness (ALT) through the TSP and the Circumpolar Active Layer Monitoring (CALM) programs (Brown et al., 2008; Romanovsky et al., 2010; Biskaborn et al., 2015, 2019). At present, at a global scale, the measured ground temperature data at

approximately 700 boreholes and ALT data at more than 200 sites can be freely downloaded from the GTN-P website (http://gtnp.arcticportal.org). Data at some monitoring sites have accumulated for decades (e.g., Biskaborn et al., 2015, 2019; Luo et al., 2016). Remote sensing observations are unprecedentedly abundant, and some variables related to permafrost, such as land surface temperature, vegetation cover, and snow cover, can be retrieved from remote sensing sensors with high accuracy and spatial-temporal resolution (e.g., Justice et al., 2002; Zhao et al., 2013). The most significant methodological advances in mapping permafrost mainly include the expansion of permafrost models to applications of various spatial domains (Riseborough et al., 2008; Obu et al., 2019) and the extensive applications of machine learning to infer the occurrence and TSP (Aalto et al., 2018; Ran et al., 2021a).

With data accumulation and technical advances, several new hemispheric-scale permafrost maps have been compiled. Gruber (2012) proposed a simple semiempirical function relationship between the mean annual ground temperature (MAGT) and mean annual air temperature (MAAT) for estimating the global permafrost zonation at a 1-km scale using downscaled MAAT data. The map indicates an areal extent of regions actually underlain by permafrost in the NH at approximately 12.9×10^6 to 17.7×10^6 km^2. Aalto et al. (2018) produced a distribution map of the circum-Arctic MAGT at a depth of zero annual amplitude (ZAA) using statistical forecasting models, with an estimated areal extent of regions actually underlain by approximately $(15.1 \pm 2.8) \times 10^6$ km^2, excluding the areal extent of permafrost to the south of 30°N in the NH. The performance of statistical models depends heavily on the *in situ* MAGT data used as a training set and the adopted predictors. The *in situ* data used in Aalto et al. (2018) were limited to those from the Qinghai-Tibet Plateau and Northeast China, while the predictor variables were mainly derived from the downscaled monthly average of climate data, thus probably limiting the prediction accuracy from the perspective of permafrost mapping. Obu et al. (2019) employed an equilibrium state model for the temperature at the top of the permafrost (TTOP) for the period of 2000–2016, driven by remotely sensed land surface temperatures, downscaled ERA-Interim climate reanalysis data, and a land cover map. The error of the modelled TTOP is ±2 °C in comparison with the data obtained from permafrost boreholes. This discrepancy is probably due to the inherent differences in the TTOP and the MAGT (at the depth of ZAA) and the shortcomings of the model structure, which lead to large uncertainties, especially in regions of warm (>−1 °C) permafrost, such as the Qinghai-Tibet Plateau (e.g., Wu et al., 2002; Riseborough, 2007; Zhao et al., 2017). The modelled areal extent of near-surface permafrost (TTOP<0 °C) covers an area of 13.9×10^6 km^2 (Obu et al., 2019). The abovementioned mapping efforts provide new generation permafrost maps. However, the accuracy of these maps is still limited, especially at middle and low latitudes and at high elevations, due to the limitations of the data and models. These limitations may be overcome by integrating more ground observations and remote sensing data using statistical or machine learning models that can improve upon physical-based models by automatically identifying better solutions (Bergen et al., 2019).

On the other hand, the maps of the IPA and Obu et al. (2019) both present permafrost zonation based on areal continuity. Areal continuity-based map systems effectively reflect the permafrost distribution characteristics in high-latitude areas, but such systems are not suitable for describing high-altitude permafrost because the areal continuity of permafrost distribution is relative and scale-dependent (Nelson, 1989; Ran et al., 2012, 2021a). With a sufficiently high spatial resolution, all permafrost

can be considered continuous or absent. Therefore, how to define the areal continuity of permafrost and the spatial resolution *per se* remain controversial and practically daunting. Additionally, continuity-based systems ignore the vertical distribution

105 and longitudinal zonation of permafrost and thus cannot effectively reflect the hydrothermal conditions of permafrost, which are important for comprehensively understanding the characteristics and vulnerability of permafrost, especially at the regional scale (Cheng, 1984; Jin et al., 2014; Ran et al., 2021a). From a thermal stability perspective, for a given thermal condition and temperature increase in the air, permafrost temperature often responds more quickly in arid regions than in humid regions because of the much greater thermal inertia of wetter soils (Abu-Hamdeh, 2003). Dry soil can reduce evaporation heat

110 consumption and increase the incident radiation on the soil (Pan and Mahrt, 1987). In addition, under different soil hydrological conditions, the response of permafrost to precipitation of various types may differ substantially (Trenberth and Shea, 2005). In humid regions, increased precipitation may heat the air and subsequently the active layer and permafrost, in contrast to that in arid regions. The difference in hydrothermal conditions is also reflected in the responses of ecosystems to precipitation because precipitation may alter or modify the effects of temperature on ecosystems (Zhao et al., 2018). In the meantime,

115 ecosystems are very important for the thermal stability of permafrost under a changing climate (Shur and Jorgenson, 2007), despite the complicated processes involved in the interactions among the climate, permafrost and ecosystem.

In short, the current statement of permafrost maps requires addressing the need for more comprehensive and integrated efforts to map the thermal state and hydrothermal zonation of permafrost. Thus, the objectives of this paper are as follows: (1) to develop and release the MAGT and ALT datasets as a baseline at a 1-km scale in the NH; (2) to provide revised zonal statistics

120 of the thermal state and distribution of permafrost in the NH; and (3) to present a permafrost zonation map with richer features of the hydrothermal conditions of permafrost.

2 Materials and methods

We first compiled a ground measurement database for MAGT and ALT in the NH (see the workflow in Figure 1). Then, the MAGT and ALT predictions with a 1-km resolution in the NH were produced using ensemble statistical forecasting by

125 integrating the remotely sensed freezing degree-days (FDD) and thawing degree-days (TDD) (i.e., the total annual degree-days below and above 0 ℃, respectively), leaf area index (LAI), snow cover duration (SCD), precipitation (mm), solar radiation (kJ m^{-2} day^{-1}), soil organic content (g kg^{-1}), soil bulk density (kg m^{-3}), coarse fragment content (volumetric %), and the compiled ground measurement data. Correspondingly, a dataset for the probability of permafrost occurrence was also produced. Finally, the hydrothermal conditions of permafrost were produced by combining the MAGT and aridity index using

130 a rule-based decision-making process.

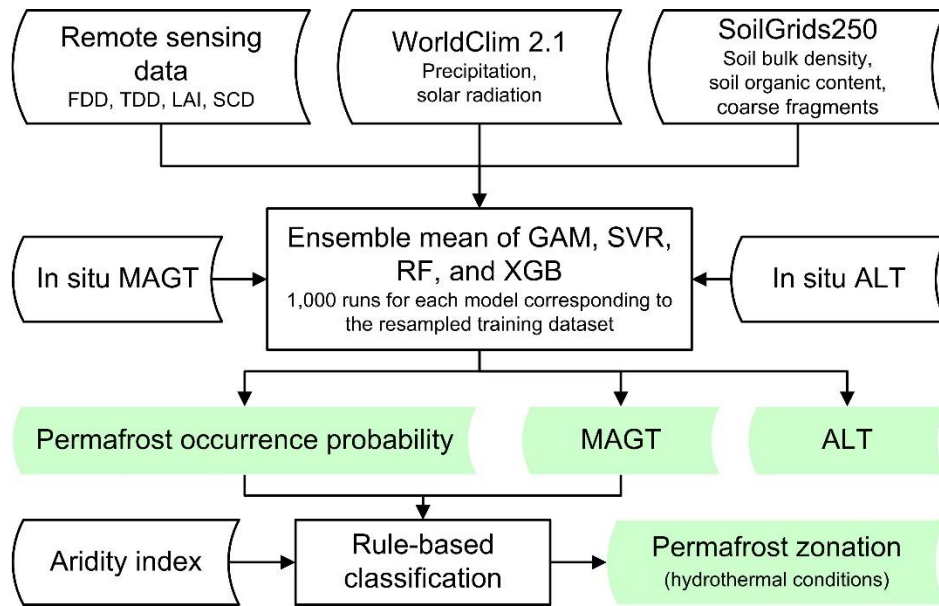



Figure 1: The data processing workflow used to compile the permafrost datasets (FDD: freezing degree-days, TDD: thawing degree-days, LAI: leaf area index, SCD: snow cover duration, MAGT: mean annual ground temperature, ALT: active layer thickness, GAM: generalized additive model, SVR: support vector regression, RF: random forest, and XGB: extreme gradient boosting).

2.1 Field measurement data

The standardized measurements of MAGT at the depth of ZAA from 1,002 boreholes and those of ALT from 452 sites were compiled mainly based on the ground measurement data used in Aalto et al. (2018), but the density of data points on the Qinghai-Tibet Plateau, in the Tianshan Mountains, and in Northeast China was greatly increased (Figure 2). Additional MAGT measurements from 253 boreholes are used mainly from the sources in Ran et al. (2021a), which were compiled mostly from existing literature on the Qinghai-Tibet Plateau (Wu et al., 2007; Yu et al., 2008; Sheng et al., 2010; Li et al., 2011, 2016; Zhang et al., 2011; Sun et al., 2013; Wang et al., 2013; Liu et al., 2015; Qiao et al., 2015; Qin et al., 2017; Cao et al., 2017; Luo et al., 2018a; Wani et al., 2020; Zhao et al., 2021) and the Tianshan Mountains (Yu et al., 2013). Other MAGT measurements from 19 boreholes in Northeast China are from Li et al. (2019) and Chang (2011). For ALT measurements, the data from 157 sites were mainly compiled from the existing literature of permafrost studies on the Qinghai-Tibet Plateau and the Tianshan Mountains (Zhao et al., 2010b; Luo et al., 2012, 2018b; Yu et al., 2013; Wu et al., 2015; Cao et al., 2017, 2018; Ali et al., 2018; Wani et al., 2020), and the data from 19 sites were compiled from the existing literature on Northeast China (Chang, 2011; He et al., 2018; Li et al., 2019, 2020a). Only the measurement data from the undisturbed (natural) sites were used. Incomplete or inaccurate location information in some studies was corrected and checked carefully via communications with the authors. The elevations of these boreholes range from 0 m to 5,428 m above sea level (asl), and 99% of these measurements were made during 2000–2016.

To reduce the potential overfitting due to residual autocorrelation, we resampled the training data 1,000 times by excluding the sampling points within a distance of less than 3 km following Ran et al. (2021a). This resulted in using an average of 776

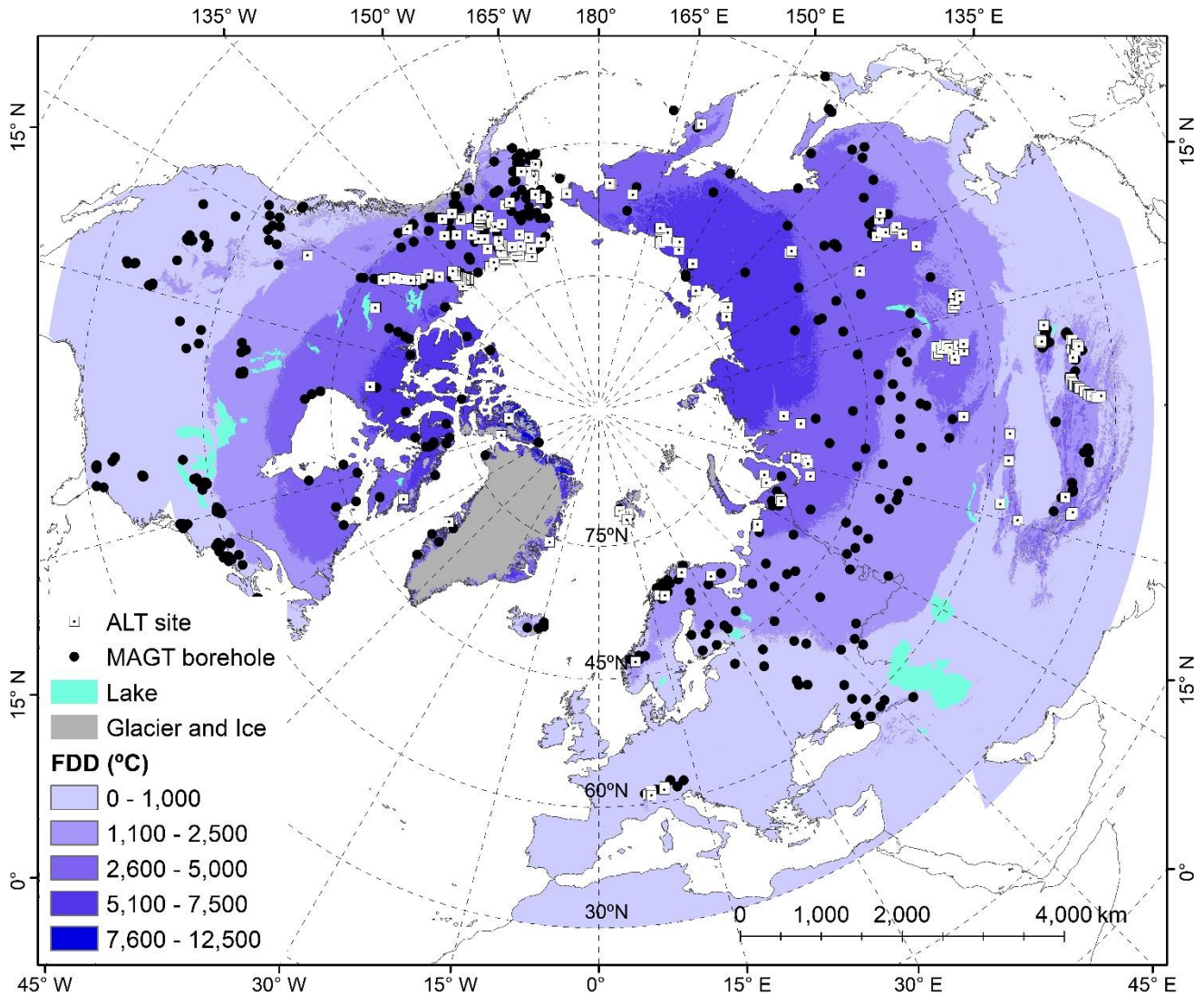
MAGT measurements and 276 ALT measurements for model training, and an average of 76 MAGT measurements and 30 ALT measurements were used for model evaluation per cross-validation run.

155

Figure 2: The distribution of the boreholes (1,002) for monitoring the mean annual ground temperature (MAGT) and the sites (452) for monitoring the active layer thickness (ALT) used in this study. FDD refers to the freezing degree-days (℃) derived from gap-filled daily MODIS land surface temperature.

2.2 Environmental and climate data

160   ### 2.2.1 FDD and TDD

The multiannual average FDD and TDD with a spatial resolution of 1 km^2 were derived from four times daily Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) (MOD11A1 and MYD11A1 version 6) from

2000 to 2016. Due to cloud contamination, gap-filling processing was achieved in the MODIS LST time series using the downscaled ERA-Interim near-surface air temperature data (Dee et al., 2011; Obu et al., 2019) based on the assumption of the

165 similarity of air and surface temperatures under cloudy skies (Gallo et al., 2011). More details about the gap-filling processing can be found in Obu et al. (2019).

2.2.2 Snow cover duration

The annual SCD data were estimated from satellite-derived bimonthly global snow cover extent (SCE) products obtained from the Japan Aerospace Exploration Agency (JAXA) Satellite Monitoring for Environmental Studies website. The SCE products

170 with 0.05 °resolution were derived from polar-orbiting satellite-borne optical sensors (AVHRR and MODIS). Snow cover was identified using radiance data at five spectral bands from visible to thermal infrared wavelength regions, the same as those of AVHRR. The overall accuracy of snow/nonsnow cover classification was estimated to be 0.82–0.99. The detailed analysis method and accuracies of the SCE product can be found in Hori et al. (2017). The SCD is the annual sum of the temporal fraction of snow cover within each bimonthly period. Then, the multiannual average values of SCD from 2000 to 2016 were

175 used in this study.

2.2.3 Leaf area index

The multiannual average values of LAI from 2000 to 2016 were derived from the Global Land Surface Satellite (GLASS), an eight-day, 1-km resolution LAI product. The GLASS LAI product was proposed based on the integration of MODIS and CYCLOPES LAI products, and the remaining cloud contamination values were removed using general regression neural

180 networks (Xiao et al., 2014). The validation results show a higher accuracy for the GLASS LAI product in comparison with the MODIS and CYCLOPES LAI products (Xiang et al., 2014).

2.2.4 Soil data

The soil organic content, bulk density, and coarse fragment content source from SoilGrids250 (https://soilgrids.org) were used in this study. With a spatial resolution of 250 m, the SoilGrids250 is a global gridded soil map developed by the International

185 Soil Reference and Information Center (ISRIC) using machine learning algorithms by integrating 230,000 soil profiles and remote sensing environmental variables (Hengl et al., 2017). The soil organic content, bulk density, and coarse fragment content were derived for seven standard depths (0, 5, 15, 30, 60, 100 and 200 cm) with a depth interval weighted average process and then aggregated to 30 arc sec resolution using the nearest resampling technique.

2.2.5 Downscaled climate data

190 The WorldClim v2.1 climate variables with a 1-km resolution (http://worldclim.org), including solar radiation (kJ m^{-2} day^{-1}) and precipitation (mm) data (Fick and Hijmans, 2017), were used in this study. Solar radiation data were directly used, but the

precipitation data for 1970–2000 were temporally adjusted to those for 2000–2016 based on WorldClim historical monthly weather data from 2000 to 2016 by using their locally smoothed (3×3 pixels) difference following Aalto et al. (2018).

2.3 Statistical learning model

195 The four statistical learning modelling techniques used in this study include the generalized additive model (GAM) (Hastie and Tibshirani, 1990), support vector regression (SVR) (Vapnik, 1995), random forest (RF) (Breiman, 2001), and eXtreme gradient boosting (XGB) (Friedman, 2001; Chen and Guestrin, 2016). The techniques were implemented based on the R packages mgcv (Wood, 2011) for GAM, randomForest (Liaw and Wiener, 2002) for RF, e1071 (Karatzoglou et al., 2006) for SVR, and xgboost (Chen and Guestrin, 2016) for XGB. The GAM is a semiparametric extension of the generalized linear
200 model in which a smooth function is specified to fit the nonlinear function of multiple predictors to the response variable at the same time (Aalto et al., 2018). In this study, the maximum smoothing function was set to three, and the thin plate regression spline was used as the smoothing function. The RF is an ensemble learning algorithm for building and aggregateing multiple decision trees. In this study, the number of trees is set to 400. Each tree is built using a randomly selected training dataset and three environmental variables to split each tree node. SVR features nonlinear kernel transformation, sparse solution, and
205 maximal margin control (Awad and Khanna, 2015). It assumes that the maximum deviation (ε) (maximal margin) between the predicted and measured values can be tolerated, and thus, an ε-insensitive loss function can be found to minimize the prediction error. The output model of the nonparametric technique depends on kernel functions. In this study, the default radial kernel function was used, and the grid search method based on 10-fold cross-validation was used to select the model parameters. A normalization method was used to avoid overfitting. XGB is an efficient implementation of a gradient boosted regression tree,
210 which is an ensemble learning algorithm, and ensembles are built by merging multiple decision trees (Chen and Guestrin, 2016). The learning corrects the prediction errors from the prior models using an optimization algorithm by sequentially adding a tree. Here, XGB was implemented using the xgboost function through 50 rounds of 10-fold cross-validation, and the max.depth was set as 3.

To account for the uncertainty of a single run and single model, an ensemble average of the 1,000 runs for the four model
215 means was used to represent the distribution of MAGT and ALT in this study.

2.4 The probability of permafrost occurrence

The probability of permafrost occurrence is obtained by calculating the fraction of predicted MAGT≤0 °C based on the multimodel ensemble results with 1,000 runs. The areal extent of the permafrost region is defined as the regions where the probability of permafrost occurrence is greater than 0.

220 ## 2.5 Hydrothermal condition-based permafrost zonation

We propose a classification system based on the hydrothermal condition of permafrost by synthesizing the thermal condition system proposed by Cheng (1984), but it was later modified by Ran et al. (2018) and further by taking into account the aridity

system proposed by Jin et al. (2014) and UNEP (1997). This system divides permafrost into nine categories using a two-level hierarchical index system, i.e., with two criteria consisting of the MAGT and climate aridity index (CAI). At the first level,

225 permafrost is divided into cold (MAGT≤−3.0 ℃), cool (−3.0 ℃<MAGT≤−1.5 ℃), and warm (MAGT>−1.5 ℃) using MAGT as the sole indicator. At the second level, permafrost is divided into humid (CAI>0.65), semiarid/subhumid (0.65≥CAI>0.20), and arid (CAI≤0.20) using the CAI, the ratio of mean annual precipitation to potential evapotranspiration (P/PET). CAI data with a 1-km resolution were used and sourced from the global aridity index database (Trabucco and Zomer, 2019) derived from precipitation data and modelled potential evapotranspiration using the Penman-Monteith equation based on WorldClim

230 2.1 climate data for the 1970–2000 period (http://worldclim.org/version2). Finally, majority statistical processes with rectangular 5×5 neighbourhoods and boundary cleaning tools were used to remove the numerous small inclusions of permafrost zonation.

2.6 Accuracy and uncertainty assessment

Model performance was assessed on the basis of the root-mean-square error (RMSE), bias, and

235 square of the correlation coefficient (R^2) computed by distance-blocked 10-fold cross-validation with 1,000 repetitions. The uncertainty of the permafrost area was quantified using the 97.5th and 2.5th percentiles of the multimodel ensemble-simulated MAGT with 1,000 runs.

3 Results and discussion

The cross-validation indicates that the ensemble average of four statistical techniques (GAM, SVR, RF, and XGB) achieved

240 the highest accuracy for MAGT (RMSE=1.32±0.13 ℃, bias=0.02±0.16 ℃), but the RF model was significantly more accurate (RMSE=85.47±20.39 cm, bias=-1.01±16.49 cm) than the ensemble mean (RMSE=86.93±19.61 cm, bias=2.71±16.46 cm) for ALT prediction (p≤0.001, paired sample t test, n=1,000) (Table 1).

Based on the datasets (named as NIEER, i.e., Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences), we analysed the distributive characteristics of MAGT, ALT, and the permafrost area/region changes along latitude,

245 elevation, and aridity index transects and the hydrothermal conditions in the permafrost regions.

Table 1: The predictive performance (mean±SD) of the mean annual ground temperature (MAGT) and active layer thickness (ALT) for the four statistical learning models and their ensemble means.

Variables	Performance measures	GAM	SVR	RF	XGB	Ensemble average
MAGT	RMSE (℃)	1.46±0.13	1.40±0.15	1.40±0.14	1.44±0.15	1.32±0.13
	Bias (℃)	0.01±0.18	0.04±0.18	0.02±0.17	0.03±0.18	0.02±0.16
	R^2	0.97±0.01	0.97±0.01	0.97±0.01	0.97±0.01	0.97±0.01

ALT	RMSE (cm)	99.40±32.91	89.20±19.94	85.47±20.39	91.76±21.24	86.93±19.61
	Bias (cm)	−0.63±19.35	12.09±16.85	−1.01±16.49	0.41±17.24	2.71±16.46
	R^2	0.72±0.14	0.76±0.08	0.78±0.08	0.75±0.09	0.77±0.08

Notes: GAM=generalized additive model, SVR=support vector regression, RF=random forest, and XGB=eXtreme gradient boosting.

3.1 Mean annual ground temperature and active layer thickness in the NH

Figure 3 shows the distribution of MAGT in the NH displaying an obvious latitudinal gradient from the zone of extremely cold (<−10 ℃) permafrost in the High Arctic to the zone of warm (>−2 ℃) permafrost in alpine and high-plateau regions at low latitudes, such as the Qinghai-Tibet and Mongolian plateaus, as well as the Yablonovy and Stanovoy mountains in southeastern Russia. In the Arctic permafrost region, the MAGT shows a clear distributive pattern of Arctic mountain permafrost in the Eastern Siberian Lowlands, on the Central Siberian Plateau, the Ural Mountains, Scandinavia in Europe, and the upper Yukon River Basin in western Canada. The average MAGT is approximately −1.56±1.06 ℃ in the Third Pole, which is cored by the Qinghai-Tibet Plateau, while that in the Arctic is approximately −4.70±3.13 ℃; the average MAGT is especially cold in the High Arctic, where it is close to −9.5 ℃. The pattern of ALT in the NH is generally similar to that of MAGT, but the details vary markedly (Figure 4). The regional average ALT varies from 76.95±21.69 cm in the High Arctic to 232.40±47.95 cm in the alpine and high-plateau permafrost regions at low latitudes with a narrow transition zone in the Mongolian Plateau and Northeast China.

The three-dimensional ground thermal regimes across the NH are investigated on the basis of the simulated MAGT and ALT. Figure 5a illustrates the latitudinal distributive patterns of MAGT and ALT, which show some comparable and contrasting features between those of MAGT and ALT. The regional average MAGT is generally stable at approximately −1.5 ℃ from 28 ℃N in the Himalayas to 60 ℃N in the Subarctic with a stable standard deviation. From 60 ℃N northwards, with rising latitudes, the MAGT decreases linearly, but its variation increases. In contrast, the regional average ALT is nearly stable at 80 cm between 56 °and 84 ℃N, and there is a narrow transitional zone between 56 ℃N and 46 ℃N, where the ALT enlarges rapidly from approximately 100 to 250 cm. These trends are considered to be predominantly associated with the transition from alpine/high-elevation plateau permafrost to lowland permafrost with generally more protective ecosystem properties (e.g., thicker overburdens). The contrasting latitudinal zonation in MAGT and ALT may indicate their varied geoenvironmental impacts on MAGT and ALT at different scales. The ALT represents the hydrothermal state near the ground surface with more spatiotemporal heterogeneity than the MAGT, which represents the thermal state of the relatively deeper ground. The vulnerability of the near-surface ground to external disturbances associated with the inconsistency of the ALT measurement method may be one of the reasons for the large uncertainty in the prediction of the ALT. The complexity of predictive models of ALT needs further investigation.

The elevational effects on MAGT and ALT are modified by latitudinal effects at the hemispherical scale because mountains are widely distributed in the Northern Hemisphere, more in the lower latitudes. This combined effect of latitude and elevation is different in different elevation ranges. From 0 to 2,500 m asl, contrary to our understanding of the elevational effect at a regional scale, the MAGT and ALT increase with increasing elevation with large variations (Figure 5b). However, the variations show opposite trends within this elevation range: with increasing elevation, the variations in MAGT decline, but those in ALT increase. In this elevation range, the elevational effect may be completely covered by latitudinal effects. The combined effect of latitude and elevation is moderate, between 2,500 and 5,000 m asl. where MAGT is stable at approximately −2 ℃ with minor variations, and ALT is stable at approximately 220 cm. Finally, at elevations >5,000 m asl, the effect of elevation appears. The MAGT and ALT both decline monotonically with rising elevation.

The trend of MAGT corresponds well to that of ALT along the rising aridity index (Figure 5c). For CAI values from 0.2 to 0.5, the MAGT and ALT both rapidly decrease with increasing CAI (moisture). However, in the range of CAI values from 0.5 to 3.6, MAGT and ALT both increase with increasing CAI. Then, above values of 3.6 on the CAI, both MAGT and ALT are nearly stable or slightly decreasing.

The three-dimensional zonal (latitudinal, elevation, and aridity) characteristics of MAGT and ALT discussed above probably indicate different controlling factors for MAGT and ALT and their variations. Multilinear regression analysis shows that the contributions of precipitation and soil bulk density to MAGT are statistically significant ($p<0.01$), but those to ALT are insignificant, while the contribution of the coarse fragment content of soils to MAGT is significant ($p<0.01$), whereas that to ALT is insignificant. The FDD, TDD, SCD, LAI, soil organic content, and solar radiation variables all contribute significantly to both MAGT and ALT ($p<0.05$). These differences may also be affected by the data gaps and uncertainty and analysis scales.

In general, at the hemisphere scale, latitude dominates the distributive patterns and trends of MAGT and ALT. However, elevation and aridity become the main factors at the regional scale. For example, in mountainous areas of sub-Arctic and low-latitude permafrost regions, elevation becomes the controller, especially above 5,000 m asl (Figure 5b). In plateau areas, such as the Qinghai-Tibet Plateau, dominated by semiarid/subhumid and arid hydrologic conditions, aridity becomes the main controller for ALT ($p<0.05$), which has not yet received due attention in existing studies; thus, further research is badly needed in the future. This highlights the necessity of the preliminary hydrothermal zonation of permafrost in Section 3.3.

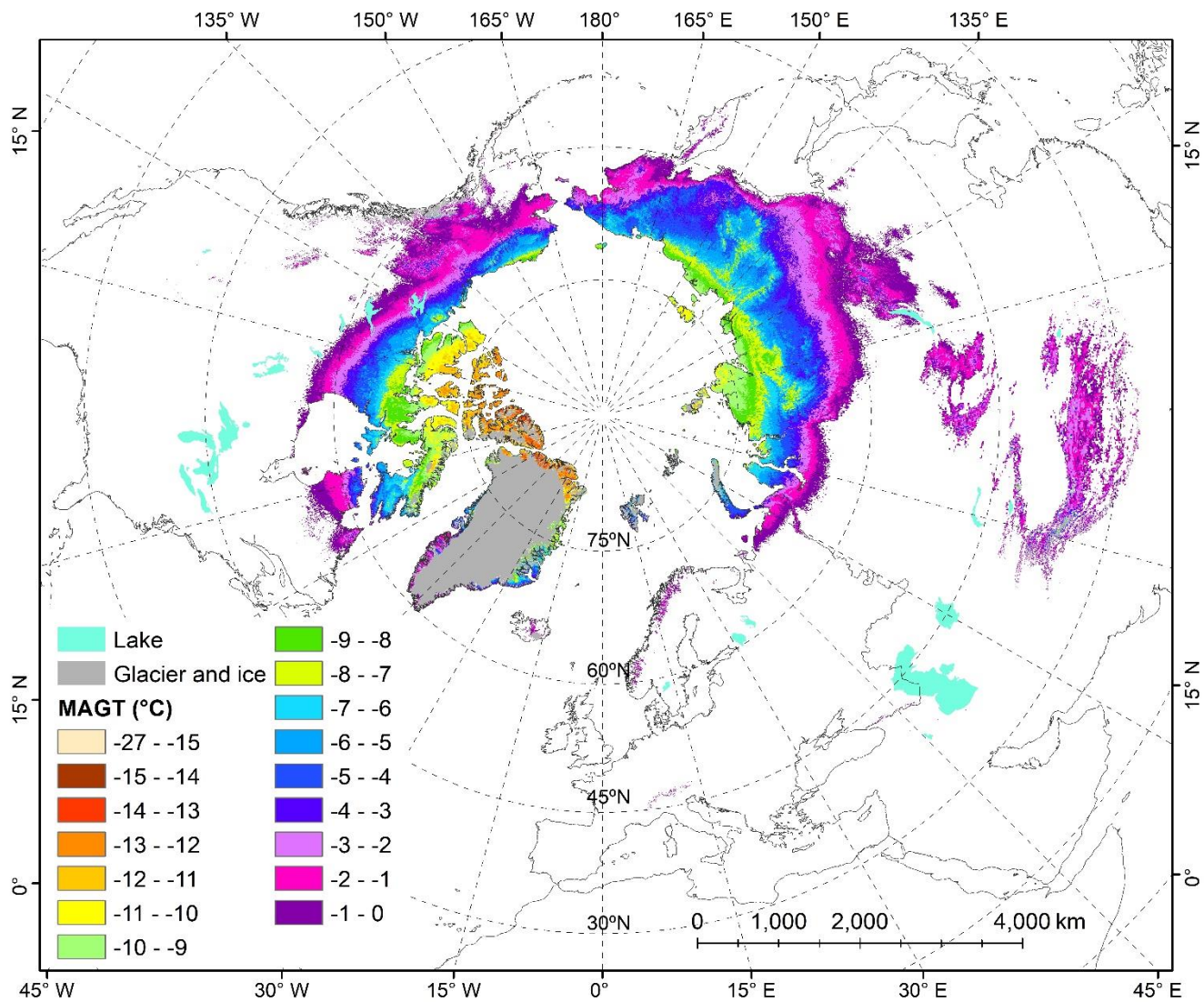
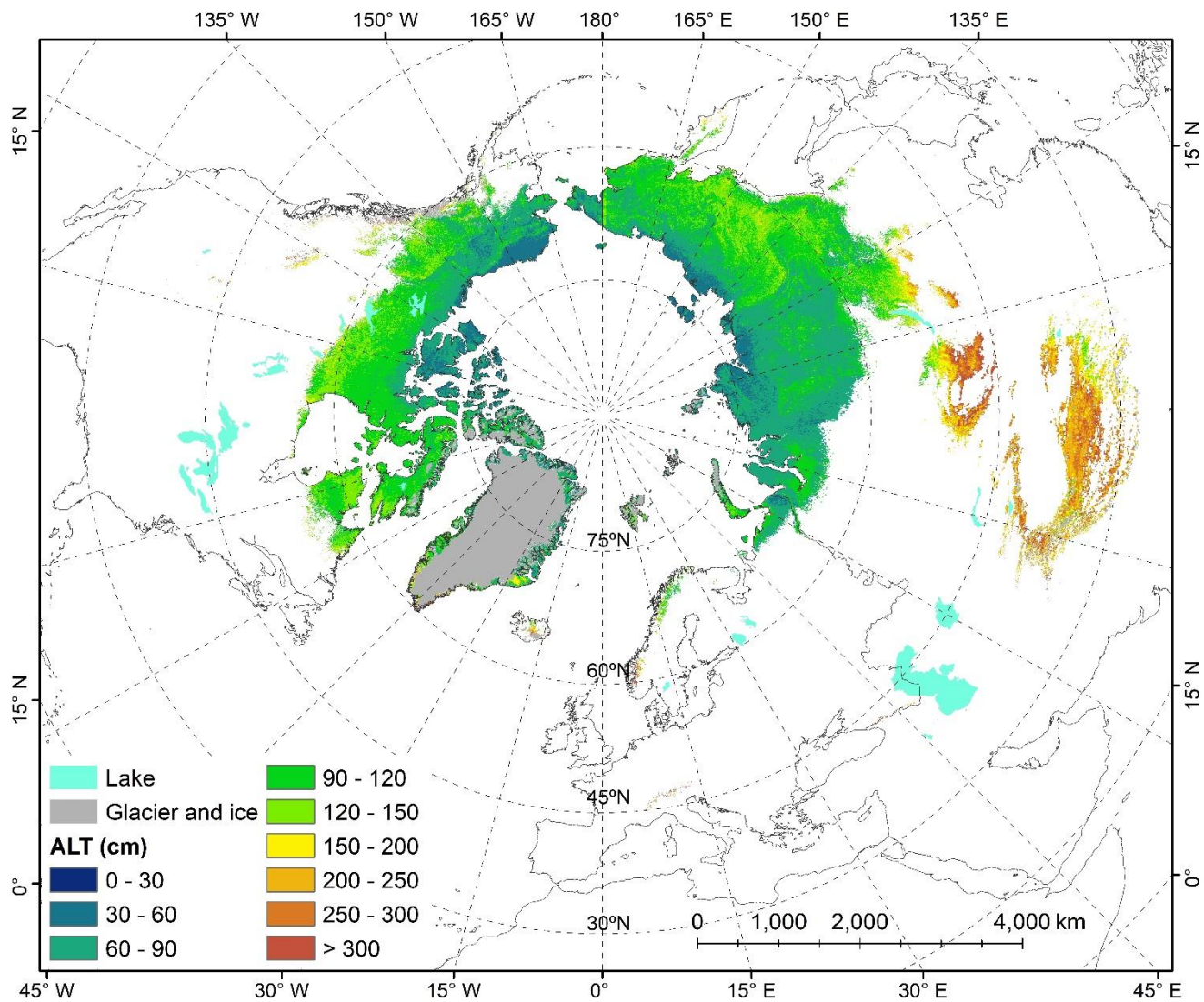



Figure 3: Distribution of the mean annual ground temperature (MAGT) average in the Northern Hemisphere for the period of 2000-2016.

305 **Figure 4: Distribution of the average active layer thickness (ALT) in the Northern Hemisphere for the period of 2000-2016.**

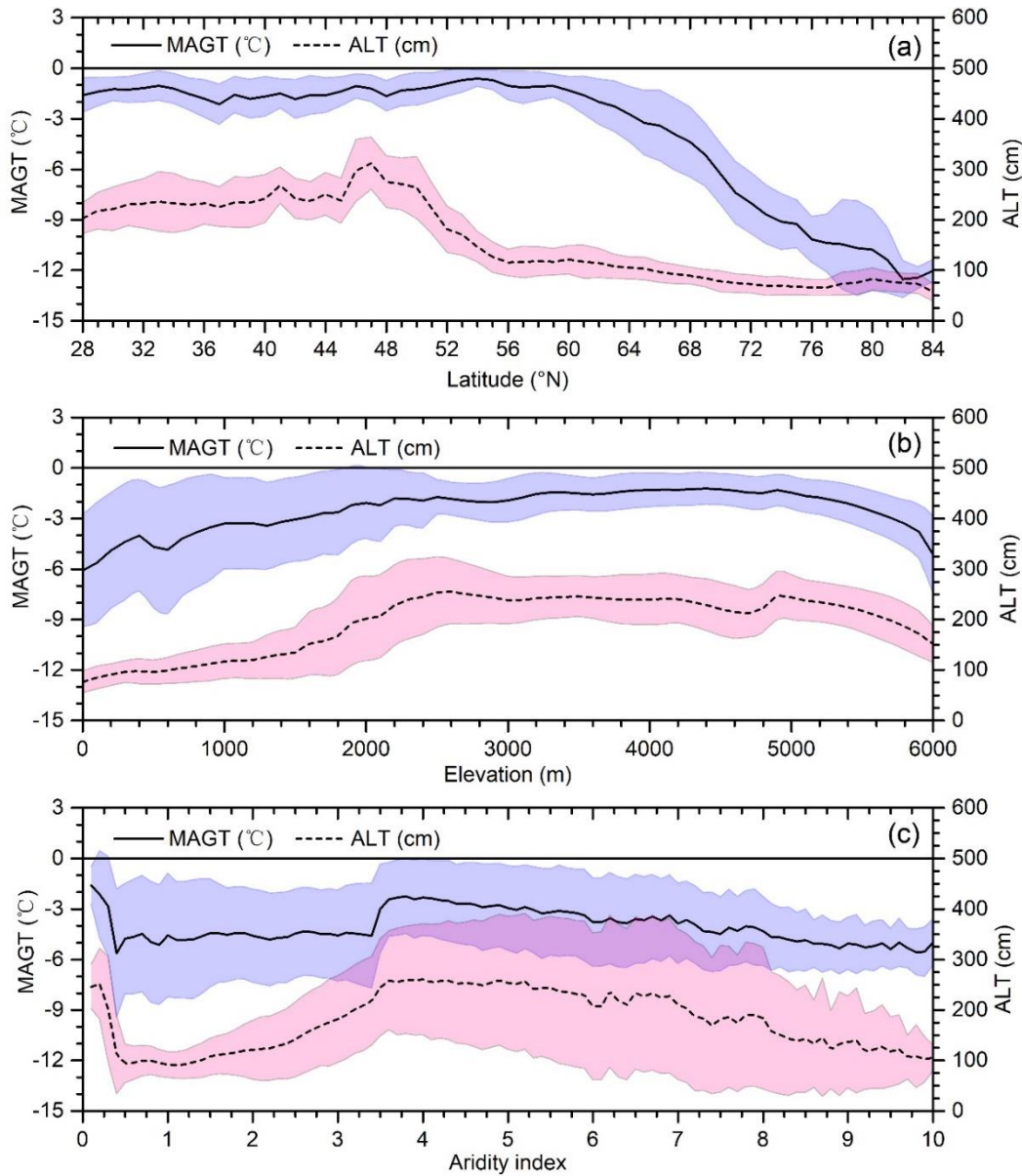



Figure 5: Mean annual ground temperature (MAGT) and active layer thickness (ALT) along latitude (a), elevation (b), and aridity index (P/PET) (c) transects in the Northern Hemisphere (black solid line is mean MAGT, and shaded area with light blue is its standard deviation). The dashed line is the mean ALT, and the shaded area with pink is its standard deviation. The dark pink area in (c) is the overlapped shaded area between ALT and MAGT).

3.2 Permafrost distribution in the NH

For permafrost distribution, two terms, i.e., permafrost area and permafrost region, should be distinguished. In general, the permafrost area is defined as the area with MAGT≤0 °C, while the permafrost region is defined as the area where the permafrost

probability is greater than 0 (Zhang et al., 2000). According to the ensemble average of MAGT, the permafrost area over the NH, excluding glaciers and lakes (small lakes were filtered out by majority statistical processes, and the areas of excluded lakes were generally greater than 50 km^2), was approximately 15.03 (13.84–19.29) $\times 10^6$ km^2. The range in permafrost area was quantified by using the 97.5th and 2.5th percentiles of the multimodel ensemble-simulated MAGT with 1,000 runs. This result is generally consistent with existing studies (Zhang et al., 2008; Gruber, 2012; Chadburn et al., 2017; Aalto et al., 2018; Obu et al., 2019). The new data indicate that the areal extent of the permafrost region is approximately 20.14 $\times 10^6$ km^2 with different occurrence probabilities (Figure 6). This area of the region is slightly less than that estimated in a recent report (Obu et al., 2019) and is approximately 2.65 $\times 10^6$ km^2 less than that estimated in the IPA Circum-Arctic permafrost map. The differences in permafrost regions between the NIEER map proposed in this study and the IPA map are mainly distributed in the southern/lower boundary of permafrost, such as over the Mongolian Plateau and Qinghai-Tibetan Plateau (Figure 7). This difference mainly reflects the difference in cartography technology in the two eras. First, the IPA map was a mechanical compilation of national maps, and each nation had its own mapping standard thereby introducing many errors and uncertainties. Second, the permafrost region of the new map used permafrost probability (>0), based on multimodel ensemble-simulated MAGT (<=0 ℃), as the boundary. This probability mainly reflects the spatial variation of uncertainty of the ensemble model and is different from the definition of permafrost occurrence probability in Gruber (2012) and Cao et al. (2019). In the real world, MAGT (<=0 ℃) is a very strict standard due to the effect of thermal offset within the active layer, and permafrost may exist between the depth of ZAA and the depth of seasonal maximum thaw (ALT).

In general, at a global scale, latitude, elevation and aridity/longitude mainly govern the distribution of permafrost (Cheng, 1984; Noetzli et al., 2017). The latitudinal dependence of the permafrost distribution is based on the latitudinal variation in insolation (incoming solar radiation) and surface energy balance and thus on the subsequent latitudinal zonation of climate, soil, and vegetation. The dependence of permafrost distribution on elevation is due to the dependence of air temperature, soil, and vegetation on elevation and strong lateral water-heat fluxes at different scales from the latitudinal effect (Noetzli et al., 2017). At a global scale, climate aridity affects the distribution of permafrost primarily by climatic continentality. This reflects the vulnerability of permafrost dependence on the annual mean temperature range and variation in the net effect of precipitation and evapotranspiration. Here, we investigate the distribution of the permafrost region and permafrost area along latitude, elevation, and aridity transects in the NH (Figure 8).

The permafrost distribution depends strongly on latitude. In the NH, permafrost occurs from 28 °N in the Himalayas to north of 75 °N in Greenland and the Canadian Arctic Archipelago, and more than 90% is distributed in the regions north of 46 °N, predominantly between 52 °and 74 °N (Figure 8a). Approximately 8% of the existing permafrost regions are distributed in the regions south of 45 °N, mainly in the high-elevation regions of Asia, such as the Qinghai-Tibet Plateau between 28 °and 40 °N. With increasing latitude, the fraction of permafrost regions in the currently exposed land surface increases until approximately 70 °N, where the fraction begins to decrease with the sharp increase in Arctic oceans, glaciers and ice sheets. The profile of the permafrost area is generally consistent with that of the permafrost region. The decrease in the differences in the distributive trends and patterns of the permafrost region and permafrost area with rising northern latitude indicates the latitudinal

dependence of permafrost areal continuity (PEC). The PEC is greater than 90% north of approximately 66 °N, i.e., terrestrial permafrost becomes continuous north of 66 °N.

350   The permafrost distribution also clearly depends on elevation, especially in the mountainous and high-plateau regions at middle and low latitudes. Approximately 80% of permafrost occurs below 1,000 m asl, and less than 10% of permafrost occurs above 3,000 m asl. In the areas below 3,000 m asl, the area of permafrost decreases with increasing elevation, but its fraction of the land surface is approximately stable. In the areas above 3,000 m asl, the area of permafrost and its fraction of the exposed land surface both increase with rising elevation (Figure 8b). This may indicate that latitude is the main controller of permafrost

355   distribution below 3,000 m asl, while elevation is a more significant controller above 3,000 m asl.

For the profile of permafrost distribution along northern longitude (Figure 8c), our study results are generally consistent with those of existing studies (Zhang et al., 2008) but with more spatial details.

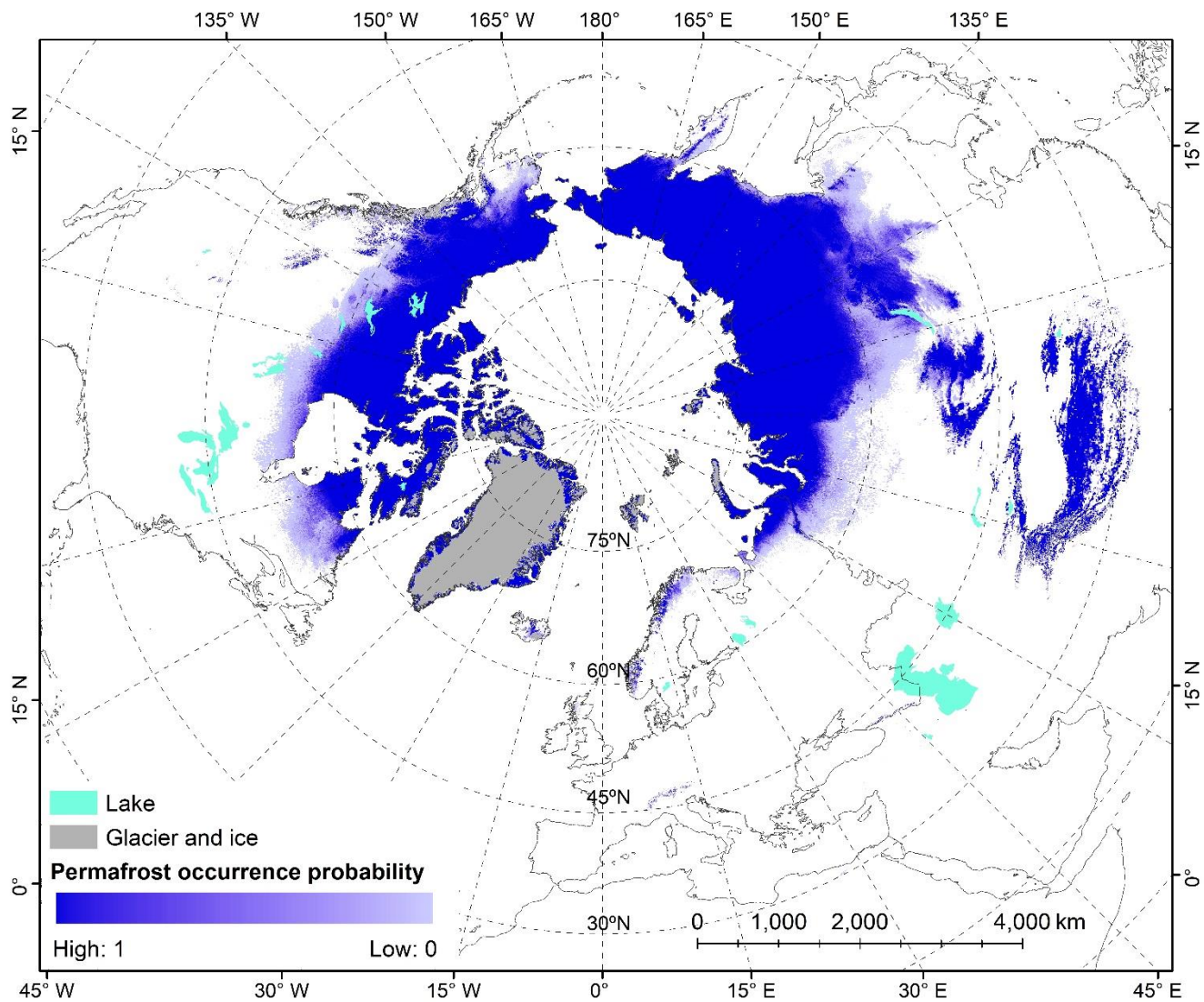



Figure 6: The present-day probability of permafrost occurrence in the Northern Hemisphere for the period of 2000-2016 (the probability is defined by the fraction of predicted MAGT≤0 °C based on the multimodel ensemble results with 1,000 runs).

360

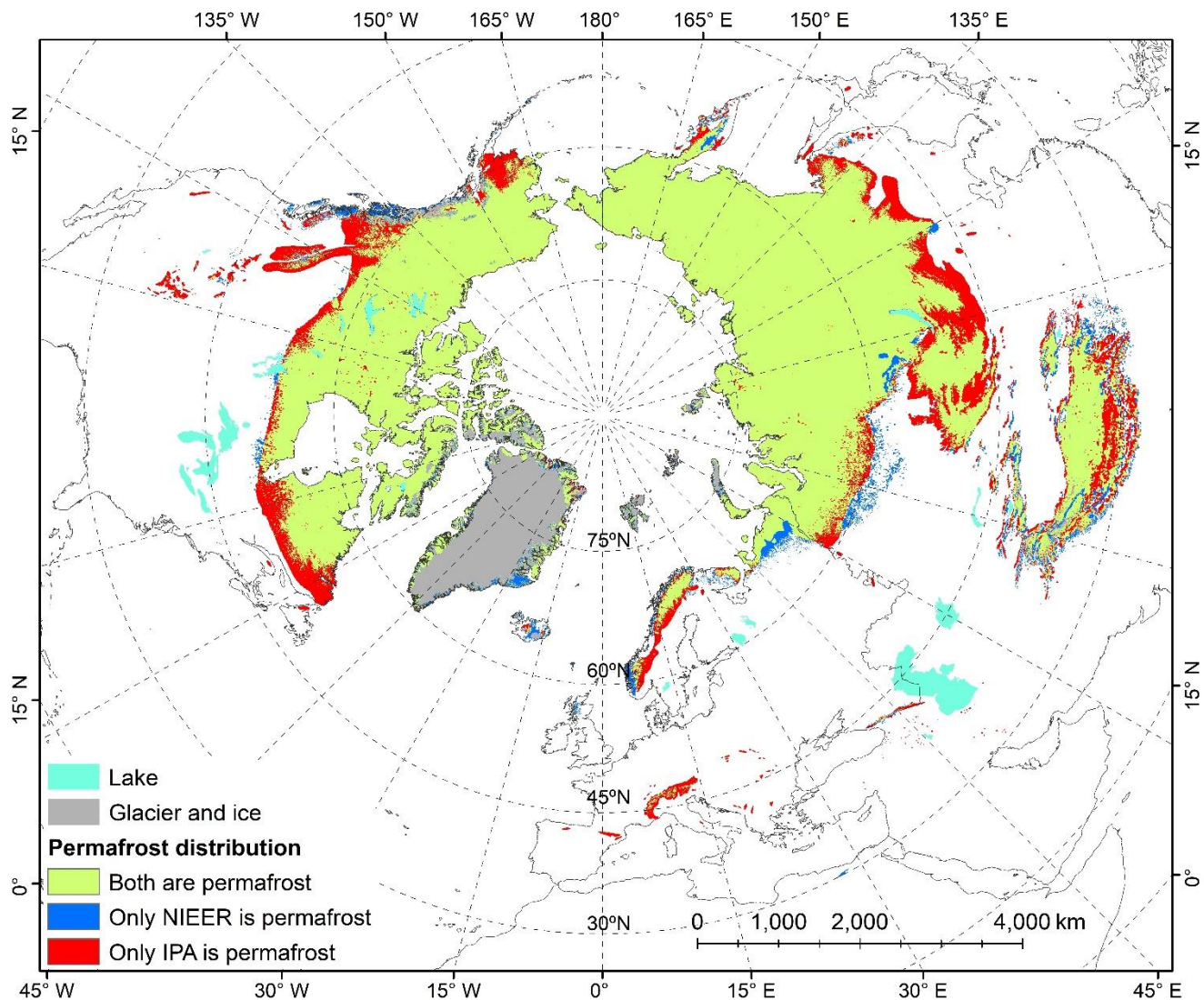



Figure 7: The spatial difference of permafrost region proposed in this study (defined as the regions where the probability of permafrost occurrence is greater than 0) with Circum-Arctic permafrost map from International Permafrost Association (Brown et al., 1997).

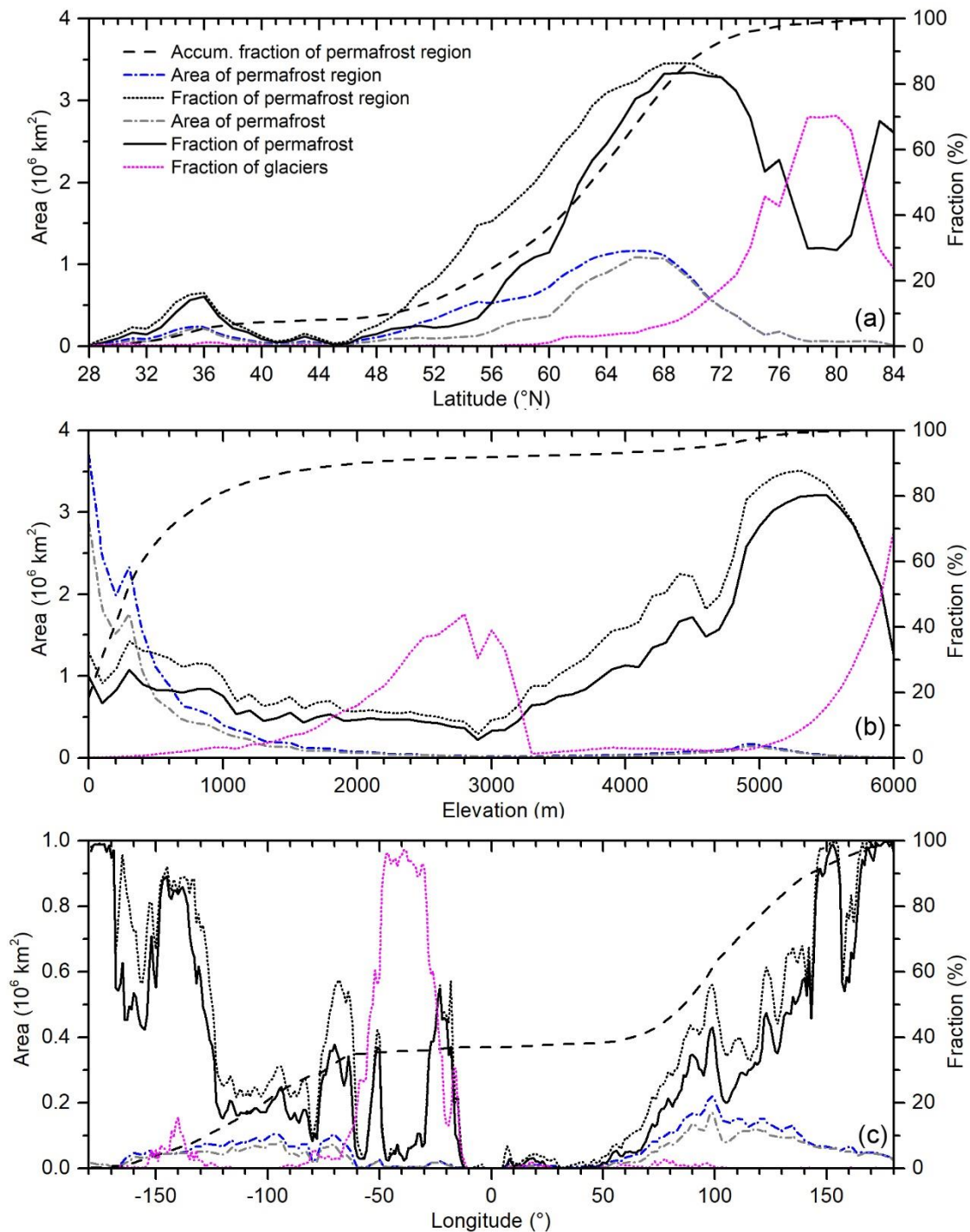



Figure 8: Distributions of the permafrost region (permafrost probability>0) and permafrost area (MAGT≤0 °C) along northern latitude (a), elevation (b), and longitude (c) transects in the Northern Hemisphere (blue short dashed dot: area of permafrost region, grey short dashed dot: area of permafrost area, black short dot: fraction of permafrost region, black dashed: accumulation fraction of permafrost region, black solid: fraction of permafrost area, magenta short dot: fraction of glacier).

365

3.3 Permafrost hydrothermal conditions

The hydrothermal conditions of permafrost differ remarkably over the NH, from the warm-arid type dominating on the Qinghai-Tibet Plateau to the cold-humid type dominating in the High Arctic. Figure 9 shows the distribution of permafrost hydrothermal conditions in the NH. In the Eurasian Arctic, permafrost occurs mainly as the cold-humid type in Western Siberia and as the cold-semiarid/subhumid type in Central and Eastern Siberia. While cold-semiarid/subhumid permafrost prevails in the western Canadian Arctic, the eastern Canadian Arctic contains cold-humid permafrost. The cold-humid type is dominant in Greenland. In the Alaskan Arctic, there is cold-semiarid/subhumid permafrost. In the permafrost regions at middle and low latitudes, such as the Qinghai-Tibet and Mongolian plateaus, permafrost occurs mainly as the warm-arid type. Warm-humid permafrost mainly occurs to the south of the continuous permafrost zone in Western Siberia and eastern Canada.

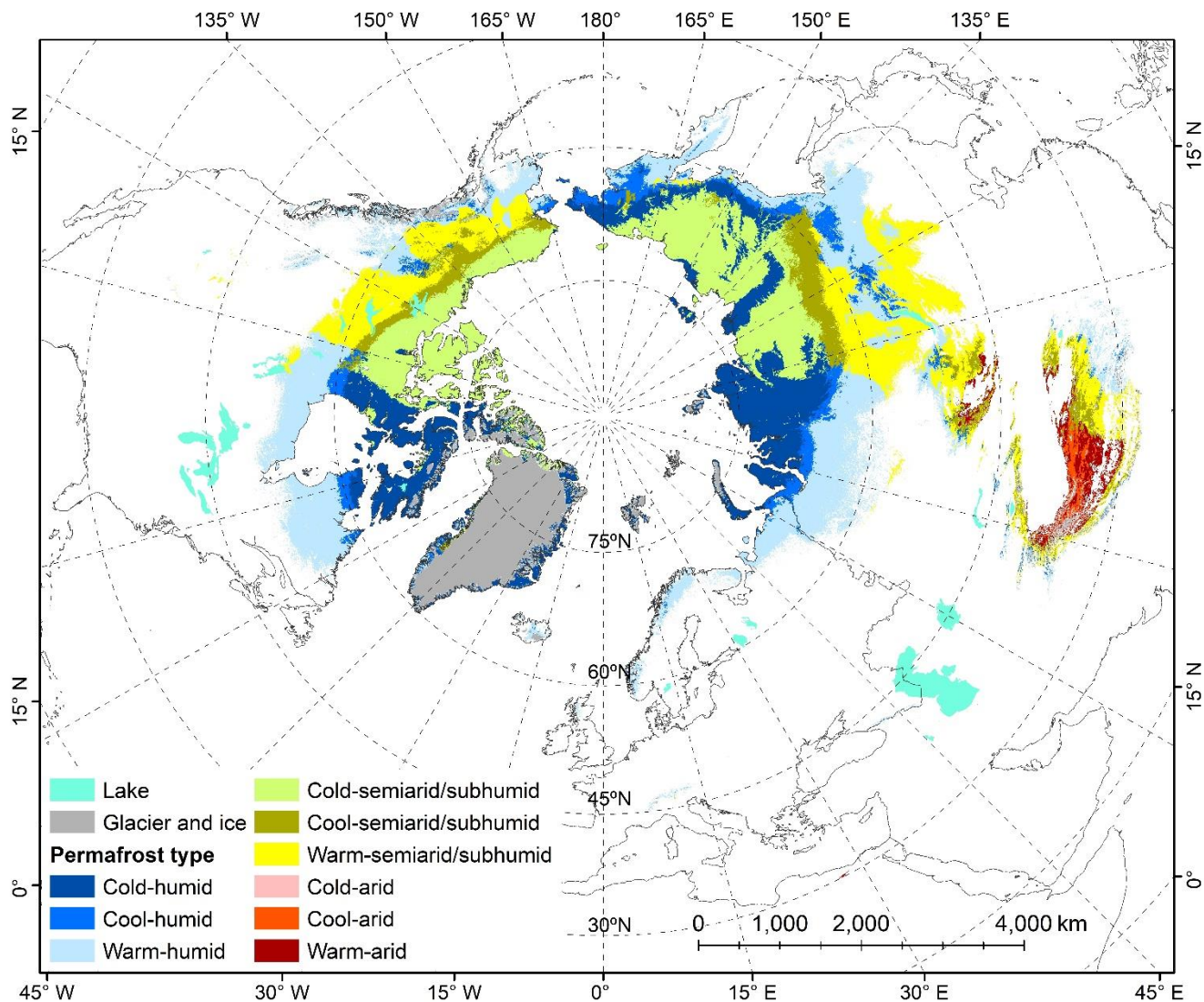
In general, from the perspective of moisture conditions alone, permafrost in the NH is dominated by the humid type, accounting for approximately 51.84% (10.44×10^6 km^2) of the permafrost region. The areas of semiarid/subhumid and arid permafrost regions account for 44.83% (9.03×10^6 km^2) and 3.33% (0.67×10^6 km^2), respectively (Table 2). From the perspective of thermal conditions, the areas of cold, cool, and warm permafrost regions are approximately 7.64×10^6 (37.93%), 2.89×10^6 (14.35%), and 9.61×10^6 km^2 (47.72%), respectively. The hydrothermal conditions are closely related to the ecosystem and climate. Thus, it is helpful to understand the environmental changes in permafrost regions and their interaction with permafrost degradation.

385

Figure 9: Hydrothermal condition-based permafrost zonation in the Northern Hemisphere for the period of 2000-2016.

Table 2: The zonal areas of different permafrost hydrothermal conditions (10^6 km^2) over the currently exposed land surface in the Northern Hemisphere.

Thermal state of permafrost	Humid (AI>0.65)	Semiarid/subhumid (0.65≥AI>0.20)	Arid (AI≤0.20)	Total
Cold (MAGT≤−3 ℃)	3.67	3.93	0.04	7.64
Cool (−3.0 ℃<MAGT≤−1.5 ℃)	1.23	1.46	0.20	2.89
Warm (MAGT>−1.5 ℃)	5.54	3.64	0.43	9.61

Total	10.44	9.03	0.67	20.14

4 Data availability

390   The NIEER datasets generated by this study (Table 3) are publicly available and can be downloaded at the National Tibetan Plateau Data Center (TPDC) (http://data.tpdc.ac.cn/en) via https://data.tpdc.ac.cn/en/data/5093d9ff-a5fc-4f10-a53f-c01e7b781368 or https://doi.org/10.11888/Geocry.tpdc.271190 (Ran et al., 2021b). The MAGT, ALT, and permafrost probability in GeoTiff format and the permafrost hydrothermal zonation map in ESRI shapefile format can be used with GIS software.

395   **Table 3: The list of permafrost datasets for the Northern Hemisphere produced in this study (MAGT: mean annual ground temperature, ALT: active layer thickness, GAM: generalized additive model, SVR: support vector regression, RF: random forest, and XGB: eXtreme gradient boosting).**

Name	Unit	Description
MAGT	℃	An ensemble average of four statistical techniques (GAM, SVR, RF, and XGB) with 1,000 runs
ALT	cm	An ensemble average of four statistical learning techniques (GAM, SVR, RF, and XGB) with 1,000 runs
Permafrost probability	-	The fraction of MAGT≤0 ℃ based on the 1,000 multi-model ensemble of MAGT predictions
Permafrost zonation (hydrothermal conditions)	-	The hydrothermal condition-based permafrost zonation after the processing described in Section 2.5

5 Conclusions

This study produced mean annual ground temperature (MAGT) at the depth of zero annual amplitude (ZAA) and active layer
400   thickness (ALT) datasets with a 1-km resolution for the period of 2000–2016 in the Northern Hemisphere (NH) (named the NIEER map). The datasets integrated unprecedentedly large amounts of ground/field measurement data (1,002 boreholes for MAGT and 452 sites for ALT) and multisource spatial data, including remotely sensed freezing degree-days (FDD) and thawing degree-days (TDD), leaf area index, snow cover duration, precipitation, solar radiation, soil organic content, soil bulk density, and coarse fragment content, using a multi-statistical/machine learning model with 1,000 runs. The cross-validation
405   shows that the accuracy of MAGT (bias=0.02±0.16 ℃, RMSE=1.32±0.13 ℃) and ALT (bias=2.71±16.46 cm, RMSE=86.93±19.61 cm) data are likely more accurate than those in previous studies. The NIEER map suggests that the areal extent of permafrost (MAGT≤0 °C) in the NH, excluding glaciers and lakes, is approximately 15.03 (13.84–19.29) $\times 10^6 \, km^2$, which is generally consistent with recent reports. The area of permafrost regions (permafrost probability>0) (approximately $20.14 \times 10^6 \, km^2$) is less than that estimated in the IPA Circum-Arctic permafrost map. This is more likely to reflect the difference

410 in cartography technology in the two eras. Based on the new permafrost classification system proposed in this study, a hydrothermal condition-based permafrost zonation map was produced by combining the permafrost probability map with MAGT and the aridity index to describe the hydrothermal characteristics of permafrost in the NH. The results show that the areal fractions of humid, semiarid/subhumid, and arid permafrost regions are 51.84%, 44.83%, and 3.33%, respectively. The areal fractions of cold (≤-3.0 ℃), cool (-3.0 ℃ to -1.5 ℃), and warm (>-1.5 ℃) permafrost regions are 37.93%, 14.35%, and

415 47.72%, respectively. Based on high-quality datasets, this study has provided a more comprehensive and accurate understanding of the permafrost distribution in the NH through analyses of changes in the MAGT, ALT, and permafrost area/region along latitude, elevation, and aridity index transects. The model-predicted MAGT and ALT, as well as the corresponding permafrost probability map and hydrothermal zonation map, are potentially valuable for cold regions and Arctic studies in a variety of fields, such as climatology, hydrology, ecology, agriculture, public health, and engineering planning in

420 the NH. Additionally, as baselines, these datasets are also important for predicting and rebuilding changes in permafrost features over the NH in the future and in the past.

Author contributions:

Y.R., X.L., and G.C. designed the study. Y.R. prepared the datasets, wrote the manuscript, plotted the figures and performed the analysis. Y.R. designed the methodology, and J.C. implemented the statistical learning. J.A., O.K., J.H., M.L., Q.Y., and

425 C.X. contributed parts of the field data. J.O. provided the TDD and FDD data. X.L., J.A., O.K., J.H., M.L., and H.J. improved the writing and structure of the paper.

Competing interests:

The authors declare that they have no conflicts of interest.

Acknowledgements:

430 This study was jointly supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19070204) and the National Natural Science Foundation of China (Grant No. 42071421). O.K. and J.H. acknowledge support from the Academy of Finland (Grant No. 315519). FDD and TDD data were generated by the European Space Agency GlobPermafrsot project (Grant No. 4000116196/15/I-NB).

References

435 Aalto, J., Karjalainen, O., Hjort, J., and Luoto, M.: Statistical forecasting of current and future Circum-Arctic ground temperatures and active layer thickness, Geophys. Res. Lett., 45(10), 4889-4898, 2018.

Abu-Hamdeh, N. H.: Thermal properties of soils as affected by density and water content, Biosyst Eng., 86(1), 97-102, 2003.

Ali, S. N., Quamar, M. F., Phartiyal, B., and Sharma, A.: Need for permafrost researches in Indian Himalaya, J. Clim. Chang., 4(1), 33-36, 2018.

440   Awad M., Khanna R. Support Vector Regression. In: Efficient Learning Machines. Apress, Berkeley, CA. doi:10.1007/978-1-4302-5990-9_4, 2015.

Bergen, K. J., Johnson, P. A., Maarten, V., and Beroza, G. C.: Machine learning for data-driven discovery in solid Earth geoscience. Science, 363(6433), eaau0323, 2019.

Biskaborn, B. K., Lanckman, J. P., Lantuit, H., Elger, K., Dmitry, S., William, C., and Vladimir, R.: The new database of the

445   Global Terrestrial Network for Permafrost (GTN-P), Earth. Syst. Sci. Data, 7, 245-259, 2015.

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A. and Allard, M.: Permafrost is warming at a global scale, Nat. Commun., 10(1), 264, 2019.

Brown, J., Ferrians Jr, O. J., Heginbottom, J. A., and Melnikov, E. S.: Circum-Arctic map of permafrost and ground-ice conditions, Reston, VA, U. S. Geol. Surv., 1997.

450   Brown, J., Hinkel, K. M., and Nelson, F. E.: The circumpolar active layer monitoring (CALM) program: research designs and initial results, Polar Geogr., 24(3), 166-258, 2000.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5-32, 2001.

Brown, J., Smith, S. L., Romanovsky, V. E., Christiansen, H. H., Clow, G., and Nelson, F. E.: Global terrestrial network for permafrost [GTN-P], Terr. Essent. Clim. Var. Assess., Mitigation Adapt., 24-25, 2008.

455   Cao, B., Gruber, S., Zhang, T., Li, L., Peng, X., Wang, K., Zheng, L., Shao, W. and Guo, H.: Spatial variability of active layer thickness detected by ground-penetrating radar in the Qilian Mountains, Western China, J. Geophys. Res. Earth. Surf., 122(3), 574-591, 2017.

Cao, B., Zhang, T., Peng, X., Mu, C., Wang, Q., Zheng, L. and Zhong, X.: Thermal characteristics and recent changes of permafrost in the upper reaches of the Heihe River basin, Western China, J. Geophys. Res. Atmos., 123(15), 7935-7949, 2018.

460   Cao, B., Zhang, T., Wu, Q., Sheng, Y., Zhao, L., and Zou, D. Permafrost zonation index map and statistics over the Qinghai-Tibet Plateau based on field evidence, Permafrost periglac. process., 30(3), 178–194, 2019.

Chadburn, S.E., Burke, E.J., Cox, P.M., Friedlingstein, P., Hugelius, G., and Westermann, S.: An observation-based constraint on permafrost loss as a function of global warming, Nat. Clim. Change, 7, 340-344, 2017.

Chang, XL.: Thermal effect of vegetation and snow cover on the underlying permafrost and soils in the active layer in the

465   northern Da Xiang'anling Mountains, Northeastern China, PhD Thesis. Univ. Chinese Acad. Sci., 1-138, 2011.

Chen T Q., and Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, 785–794, doi: 10.1145/2939672.2939785, 2016.

Cheng, G.: Problems on zonation of high-altitude permafrost, Acta. Geogr. Sin., 39: 185-193, 1984. (In Chinese)

470 Cheng, G.: A roadbed cooling approach for the construction of Qinghai–Tibet Railway, Cold Reg. Sci. Technol., 42(2), 169-176, 2005.

Cheng, G., and Wu, T.: Responses of permafrost to climate change and their environmental significance, Qinghai-Tibet Plateau, J. Geophys. Res. Earth. Surf., 112(F02S03), 93-104, 2007.

Cheng, G. and Jin, H.J.: Permafrost and groundwater on the Qinghai-Tibet Plateau and in northeast China, Hydrogeol. J., 21(1), 5–23, 2013.

475 Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., Berg, L. van de, Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., Rosnay, P. de, Tavolato, C., Thépaut, J.-N., Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137(656), 553-597, 2011.

480 Dvornikov, Y., Leibmann, M., Heim, B., Bartsch, A., Haas, A., Khomutov, A., Gubarkov, A., Mikhaylova, M., Mullanurov, D., Widhalm, B., Skorospekhova, T., and Fedorova, I.: Geodatabase and WebGIS project for long-term permafrost monitoring at the Vaskiny Dachi research station, Yamal, Russia, Polarforschung, 85(2), 107-115, 2016.

485 Fick, S.E. and R.J. Hijmans.: WorldClim 2: new 1km spatial resolution climate surfaces for global land areas, Int. J. Climatol., 37 (12), 4302-4315, 2017.

Friedman, J. H.: Greedy function approximation: a gradient boosting machine, Ann. Stat., 29(5), 1189-1232, 2001.

Gao, T., Zhang, T., Cao, L., Kang, S., and Sillanpää, M.: Reduced winter runoff in a mountainous permafrost region in the northern Tibetan Plateau, Cold Reg. Sci. Technol., 126, 36-43, 2016.

490 Gallo, K., Hale, R., Tarpley, D., and Yu, Y.: Evaluation of the relationship between air and land surface temperature under clear-and cloudy-sky conditions, J. Appl. Meteorol. Climatol., 50(3), 767-775, 2011.

Gruber, S.: Derivation and analysis of a high-resolution estimate of global permafrost zonation, Cryosphere, 6(1), 221, 2012.

Harris, C., Haeberli, W., Vonder Mühll, D., and King, L.: Permafrost monitoring in the high mountains of Europe: the PACE project in its global context, Permafrost periglac. process., 12(1), 3-11, 2001.

495 Hastie, T. J., and Tibshirani, R. J.: Generalized additive models (Vol. 43), London: CRC press, 1990.

He R., Jin, H., Chang, X., Wang, Y., and Wang, L.: Freeze-thaw processes of active-layer soils in the Nanweng'he River National Natural Reserve in the Da Xing'anling Mountains, northern Northeast China, Sci. Cold. Arid. Reg., 10(2), 104-113, 2018.

Heginbottom, J. A.: Permafrost mapping: a review, Prog. Phys. Geogr., 26(4), 623-642, 2002.

500 Heginbottom, J. A., Brown, J., Melnikov, E. S., and Ferrians Jr, O. J.: Circumarctic map of permafrost and ground ice conditions, Proc. Int. Conf. Permafrost. Natl. Snow Ice Data Cent. /World Data Cent. Glaciol. Boulder, CO, 2, 1132-1136, 1993.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G., and

505   Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS One, 12(2), 2017.

Hjort, J., Karjalainen, O., Aalto, J., Westermann, S., Romanovsky, V.E., Nelson, F.E., Etzelmüller, B., and Luoto, M.: Degrading permafrost puts Arctic infrastructure at risk by mid-century, Nat. Commun., 9, 5147, 2018.

Hori, M., Sugiura, K., Kobayashi, K., Aoki, T., Tanikawa, T., Kuchiki, K., Niwano, M. and Enomoto, H.: A 38-year (1978–

510   2015) Northern Hemisphere daily snow cover extent product derived using consistent objective criteria from satellite-borne optical sensors, Remote Sens. Environ., 191, 402-418, 2017.

Jin, H., Lü, L., and He, R.: A new aridity-based classification of permafrost zones on the Tibetan Plateau, J. Glaciol. Geocryol., 36(5), 1049-1057, 2014. (In Chinese, English abstract)

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D.P. and Morisette, J. T.:

515   An overview of MODIS Land data processing and product status, Remote Sens. Environ., 83(1-2), 3-15, 2002.

Karatzoglou, A., Meyer, D., and Hornik, K.: Support vector machines in R, J. Stat. Softw., 15(9), 1-28, 2006.

Kellerer-Pirklbauer, A., Bartsch, A., Gitschthaler, C., Reisenhofer, S., Weyss, G., Riedl, C., and Avian, M.: A national strategy for a long-term monitoring of permafrost and periglacial processes and their relationship to natural hazard prevention in Austria, EGUGA, 18, EPSC2016-15245, 2016.

520   Li, J, Sheng, Y., Chen, J., Zhang, B., Wu, J, and Zang, X.: Characteristics of ground temperatures and influencing factors of permafrost development and distribution in the source region of Datong River, Prog. Geogr., 30(7), 827-836, 2011. (In Chinese, English abstract)

Li, J., Sheng, Y., and Wu, J.: Mapping frozen soil distribution and modeling permafrost stability in the Source Area of the Yellow River, Sci. Geogr. Sin., 36(4), 588-596, 2016. (In Chinese, English abstract)

525   Li, X., Jin, H., He, R., Huang, Y., Wang, H., Luo, D., Jin, X., Lv, L., Wang, L., Li, W., Wei, C., Chang, X., Yang, S. and Yu, S.: Effects of forest fires on the permafrost environment in the northern Da Xing'anling (Hinggan) mountains, Northeast China, Permafrost Periglac. Process., 30(3), 163-177, 2019.

Li, X., Jin, H., Wang, H., Wu, X., Huang, Y., He, R., Luo, D. and Jin, X.: Distributive features of soil carbon and nutrients in permafrost regions affected by forest fires in northern Da Xing'anling (Hinggan) Mountains, NE China, Catena, 185, 104304,

530   2020a.

Li, X., Che, T., Li, X., Wang, L., Duan, A., Shangguan, D., Pan, X., Fang, M., and Bao, Q.: CASEarth Poles: Big data for the Three Poles, Bull. Am. Meteorol. Soc., 101: E1475–E1491, 2020b.

Liaw, A., and Wiener, M.: Classification and regression by randomForest, R News, 2(3), 18-22, 2002.

Liu, G. Y., Wang. W., Zhao, L., Chen, J., Pang, Q. Q., Wang, Z. W., Du, E. J.: Using transient electromagnetic method to

535   sound permafrost depth in the West Kunlun Mountains, J. Glaciol. Geocryol., 37(1), 38-48, 2015. (In Chinese, English abstract)

Luo D., Jin H., Lin L., He R., Yang S., and Chang X.: New progress on permafrost temperature and thickness in the Source Area of the Huanghe River, Sci. Geogr. Sin., 32(7), 898-904, 2012.

Luo, D., Wu, Q., Jin, H., Marchenko, S. S., Lü, L., and Gao, S.: Recent changes in the active layer thickness across the northern hemisphere, Environ. Earth Sci., 75(7), 555, 2016.

540 Luo, D., Jin, H., Jin, X., He, R., Li, X., Muskett, R. R. and Romanovsky, V. E.: Elevation-dependent thermal regime and dynamics of frozen ground in the Bayan Har Mountains, northeastern Qinghai-Tibet Plateau, southwest China, Permafrost Periglac. Process., 29(4), 257-270, 2018.

Luo, D., Jin, H., Wu, Q., Bense, V. F., He, R., Ma, Q. and Lü, L.: Thermal regime of warm-dry permafrost in relation to ground surface temperature in the Source Areas of the Yangtze and Yellow rivers on the Qinghai-Tibet Plateau, SW China, Sci. Total
545 Environ., 618, 1033-1045, 2018.

Mair, V., Zischg, A., Lang, K., Tonidandel, D., Krainer, K., Kellerer-Pirklbauer, A., Deline, P., Schoeneich, P., Cremonese, E., Pogliotti, P., Gruber, S. and Böckli, L.: PermaNET-Permafrost Long-term Monitoring Network: Synthesis report, Int. Res. Soc. INTERPRAEVENT, 2011.

Nelson, F.E. Permafrost zonation in eastern Canada: a review of published maps, Phys. Geogr., 10(3), 231-246, 1989.

550 Noetzli, J., S. Gruber, T. Kohl, N. Salzmann, and W. Haeberli: Three-dimensional distribution and evolution of permafrost temperatures in idealized high-mountain topography, J. Geophys. Res., 112, F02S13, doi:10.1029/2006JF000545, 2017.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., Khomutov, A., Kääb, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T., Yamkhin, J. and Zou, D.: Northern Hemisphere permafrost map based on TTOP
555 modelling for 2000–2016 at 1 km^2 scale, Earth-Sci. Rev., 193, 299-316, 2019.

Pan, H. L., and Mahrt, L.: Interaction between soil hydrology and boundary-layer development, Bound. Lay. Meteorol., 38(1), 185-202, 1987.

Qiao, Y., Zhao, L., Pang, Q., Chen, J., Zou, D., and Gao, Z.: Characteristics of permafrost in Gerze county on the Tibetan Plateau, J. Glaciol. Geocryol., 37(06), 1453-1460, 2015. (In Chinese, English abstract)

560 Qin, Y., Wu, T., Zhao, L., Wu, X., Li, R., Xie, C., and Liu, G.: Numerical modeling of the active layer thickness and permafrost thermal state across Qinghai-Tibetan Plateau, J. Geophys. Res.: Atmos., 122(21), 11-604, 2017.

Ran, Y., Li, X., Cheng, G., Zhang, T., Wu, Q., Jin, H., and Jin, R.: Distribution of permafrost in China: an overview of existing permafrost maps, Permafrost Periglac. Process., 23(4), 322-333, 2012.

Ran, Y., Li, X., and Cheng, G.: Climate warming over the past half century has led to thermal degradation of permafrost on
565 the Qinghai–Tibet Plateau, Cryosphere, 12(2), 595-608, 2018.

Ran, Y., Li, X., Cheng, G., Nan, Z., Che, J., Sheng, Y., Wu, Q., Jin, H., Luo, D., Tang, Z. and Wu, X.: Mapping the permafrost stability on the Tibetan Plateau for 2005–2015, Sci. China: Earth Sci., 64(1), 62–79, 2021a.

Ran, Y., Li, X., Cheng, G., Che, J., Aalto, J., Karjalainen, O., Hjort, J., Luoto, M., Jin, H., Obu, J., Hori, M., Yu, Q., Chang, X. High-resolution datasets of permafrost thermal state and hydrothermal zonation in the Northern Hemisphere. National

570   Tibetan Plateau Data Center, 2021. doi: 10.11888/Geocry.tpdc.271190, 2021b.

Riseborough, D.: The effect of transient conditions on an equilibrium permafrost-climate model, Permafrost Periglac. Process., 18(1), 21-32, 2007.

Riseborough, D., Shiklomanov, N., Etzelmüller, B., Gruber, S., and Marchenko, S.: Recent advances in permafrost modelling, Permafrost Periglac. Process., 19(2), 137-156, 2008.

575   Romanovsky, V. E. and Osterkamp, T. E.: Permafrost monitoring system in Alaska: structure and results, Kriosfera Zemli, 5, 59-68, 2001. (in Russian)

Romanovsky, V. E., Smith, S. L., and Christiansen, H. H.: Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–2009: a synthesis, Permafrost Periglac. Process., 21(2), 106-116, 2010.

Schuur, E. A., Vogel, J. G., Crummer, K. G., Lee, H., Sickman, J. O., and Osterkamp, T. E.: The effect of permafrost thaw on

580   old carbon release and net carbon exchange from tundra. Nature, 459(7246), 556-559, 2009.

Sheng, Y., Li, J., Wu, J. C., Ye, B. S., and Wang, J.: Distribution patterns of permafrost in the upper area of Shule River with the application of GIS technique, J. China Univ. Min. Technol., 39(1), 32-39, 2010.

Shur, Y. L., and Jorgenson, M. T.: Patterns of permafrost formation and degradation in relation to climate and ecosystems, Permafrost Periglac. Process., 18(1), 7-19, 2007.

585   Smith, S. L., Burgess, M. M., Riseborough, D., and Mark Nixon, F.: Recent trends from Canadian permafrost thermal monitoring network sites, Permafrost Periglac. Process., 16(1), 19-30, 2005.

Smith, S. L., Romanovsky, V. E., Lewkowicz, A. G., Burn, C. R., Allard, M., Clow, G. D., Yoshikawa, K. and Throop, J.: Thermal state of permafrost in North America: a contribution to the international polar year, Permafrost Periglac. Process., 21(2), 117-135, 2010.

590   Sun, Z. Z., Ma, W., Dang, H. M., Yun, H. B., and Wu, G. L.: Characteristics and causes of embankment deformation for Qinghai-Tibet Railway in permafrost regions, Rock Soil Mech., 34(9), 2667-2671, 2013. (In Chinese, English abstract)

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Global Biogeochem. Cy., 23(2), 2009.

Trabucco, A., and Zomer, R.: Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. figshare.

595   Fileset. https://doi.org/10.6084/m9.figshare.7504448.v3, 2019.

Trenberth, K. E., and Shea, D. J.: Relationships between precipitation and surface temperature, Geophys. Res. Lett., 32(14), 2005.

UNEP (United Nations Environment Programme). World atlas of desertification 2ED. Arnold, London, p 182, 1997.

Vapnik, V.: The Nature of Statistical Learning Theory, N. Y.: Springer, 1995.

600   Van Everdingen, R. O. Multi-Language Glossary of Permafrost and Related Ground-Ice Terms. International Permafrost Association, 1998.

Wang, Q., Zhang, T., Wu, J., Peng, X., Zhong, X., Mu, C. and Cheng, G.: Investigation on permafrost distribution over the upper reaches of the Heihe River in the Qilian Mountains, J. Glaciol. Geocryol., 35(1), 19-25, 2013. (In Chinese, English abstract)

605 Wani, J. M., Thayyen, R. J., Gruber, S., Ojha, C. S. P., and Stumm, D.: Single-year thermal regime and inferred permafrost occurrence in the upper Ganglass catchment of the cold-arid Himalaya, Ladakh, India, Sci. Total Environ., 703, 134631, 2020.

Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, J. R. Stat. Soc. Ser. B-Stat. Methodol., 73(1): 3-36, 2011.

Wu, Q., Zhu, Y., and Liu, Y.: Application of the permafrost table temperature and thermal offset forecast model in the Tibetan

610 Plateau, J. Glaciol. Geocryol., 24(5), 614-617, 2002. (Chinese with English abstract)

Wu, J., Yu, S., Yu, H., and Li, J.: Permafrost in the middleeast section of Qilian Mountains (II): Characters of permafrost, J. Glaciol. Geocryol., 29: 426-432, 2007. (in Chinese with English abstract)

Wu, Q., Hou, Y., Yun, H., and Liu. Y.: Changes in active-layer thickness and near-surface permafrost between 2002 and 2012 in alpine ecosystems, Qinghai–Xizang (Tibet) Plateau, China, Glob. Planet. Change, 124: 149-155, 2015.

615 Xiang, Y., Xiao, Z. Q., Ling, S. L., Wang, J. D., and Song, J. L.: Validation of Global LAnd Surface Satellite (GLASS) leaf area index product, Int. J. Remote Sens., 18(3), 573-584, 2014.

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., and Song, J.: Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance, IEEE Trans. Geosci. Remote Sens., 52(1), 209-223, 2014.

620 Xie, Y.L., Yu, Q.H., You, Y.H., Zhang, Z. Q., Gou, T.T. The changing process and trend of ground temperature around tower foundations of Qinghai-Tibet Power Transmission line, Sciences in Cold and Arid Regions, 11(1), 0013–0020, 2019.

Yoshikawa, K., and Hinzman, L. D.: Shrinking thermokarst ponds and groundwater dynamics in discontinuous permafrost near Council, Alaska, Permafrost Periglac. Process., 14(2), 151-160, 2003.

Yu, H., Wu, Q. B., and Liu, Y. Z.: The long-term monitoring system on permafrost regions along the Qinghai-Tibet Railway,

625 J. Glaciol. Geocryol., 30(3), 475-481, 2008. (In Chinese, English abstract)

Yu, Q., You, Y., Yan, H., Liu X.: Distribution and characteristics of permafrost in Nalati Mountain, western Tianshan Mountains in China, J. Glaciol. Geocryol., 35(1), 10-18, 2013. (In Chinese, English abstract)

Zhang, T., Heginbottom, J. A., Barry, R. G., and Brown, J.: Further statistics on the distribution of permafrost and ground ice in the Northern Hemisphere, Polar Geogr., 24(2), 126-131, 2000.

630 Zhang, T., Barry, R. G., Knowles, K., Heginbottom, J. A., and Brown, J.: Statistics and characteristics of permafrost and ground-ice distribution in the Northern Hemisphere, Polar Geogr., 31(1-2), 47-68, 2008.

Zhang, X. M., Nan, Z. T., Wu, J. C., Du, E. J., Wang, T., and You, Y. H.: Modeling permafrost distribution in Wenquan Area over Qinghai-Tibet Plateau by using multivariate adaptive regression splines, J. Glaciol. Geocryol., 33, 1088-1097, 2011. (In Chinese, English abstract)

635 Zhao, L., Wu, Q., Marchenko, S. S., and Sharkhuu, N.: Thermal state of permafrost and active layer in Central Asia during the International Polar Year, Permafrost Periglac. Process., 21(2), 198-207, 2010a.

Zhao, L., Liu, G. Y., Jiao, K. Q., Li, R., Qiao, Y. P., and Ping, C. L.: Variation of the permafrost in the Headwaters of the Urumqi River in the Tianshan Mountains since 1991, J. Glaciol. Geocryol., 32(2), 223-230, 2010b.

Zhao, X., Liang, S., Liu, S., Yuan, W., Xiao, Z., Liu, Q., Cheng, J., Zhang, X., Tang, H., Zhang, X., Liu, Q., Zhou, G., Xu, S.,
640 and Yu, K.: The Global Land Surface Satellite (GLASS) remote sensing data processing system and products, Remote Sens., 5(5), 2436-2450, 2013.

Zhao, S. P., Nan, Z. T., Huang, Y. B., and Zhao, L.: The application and evaluation of simple permafrost distribution models on the Qinghai–Tibet Plateau, Permafrost Periglac. Process., 28(2), 391-404, 2017.

Zhao, J., Luo, T., Li, R., Wei, H., Li, X., Du, M., and Tang, Y.: Precipitation alters temperature effects on ecosystem respiration
645 in Tibetan alpine meadows, Agric. For. Meteorol., 252, 121-129, 2018.

Zhao, L., Zou, D., Hu, G., Wu, T., Du, E., Liu, G., Xiao, Y., Li, R., Pang, Q., Qiao, Y., Wu, X., Sun, Z., Xing, Z., Sheng, Y., Zhao, Y., Shi, J., Xie, C., Wang, L., Wang, C., and Cheng, G.: A synthesis dataset of permafrost thermal state for the Qinghai-Xizang (Tibet) Plateau, China, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2021-1, in review, 2021.

Zhelezniak, M., Wu, Q. B., Kirillin, A., Wen, Z., Zhirkov, A., and Zhizhin, W. Permafrost distribution and temperature in the
650 Elkon Horst, Russia. Sciences in Cold and Arid Regions, 13(2), 107−122, 2021.