https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Open Access Earth System Science Data Discussions

An Ensemble Dataset of Permafrost Thaw Conditions for Northern High Latitudes Using Open Satellite Data

Dianfan Guo[1,2], Cuizhen Wang[3*], Shuying Zang[1,2*]

[1]Heilongjiang Province Key Laboratory of Geographical Environment Monitoring and Spatial Information Service in Cold Regions, Harbin Normal University, Harbin, 150025, China
[2] Heilongjiang Province Collaborative Innovation Center of Cold Region Ecological Safety, Harbin, 150025, China
[3] Department of Geography, University of South Carolina. Columbia, SC 29208, USA

Correspondence to: Shuying Zang (zsy6311@hrbnu.edu.cn), Cuizhen Wang (Cwang@mailbox.sc.edu)

Abstract. Permafrost in the northern high latitudes is experiencing accelerated warming, casting important implications for carbon–climate feedback and ecosystem stability. Integrating three open-access permafrost products, this study generates an ensemble permafrost dataset of permafrost percent (PP) and mean annual ground temperature (MAGT), and proposes a Permafrost Thaw Index (PTI) to measure thaw vulnerability in a two-tier ranking system. Sixteen open-access datasets derived primarily from satellite observations are ensembled to represent environmental conditions, including land surface temperature (LST), vegetation dynamics, snow cover, freeze–thaw state, soil properties, and topography. An ensemble machine learning approach, XGBoost, is employed to predict PTI from these datasets with an overall accuracy of 91.8%. Thermal variables, particularly LST, LST trends, and frozen days, contribute most strongly to PTI prediction. The PTI map reveals clear latitudinal gradients and regional variations in alignment with eco-climatic transitions. The total permafrost coverage in the northern high latitudes (> 45°N) is approximately 18.7 million km^2. About 41.4% remains highly stable, while nearly half is subject to high thaw pressure. The most stable conditions persist in the High Arctic tundra of North America, the Arctic East Siberia, and the high-elevation alpine tundra of North Asia. The highest vulnerability occurs along the southern margins dominated with boreal and montane forests. A comparison of the predicted PTI with borehole records from 26 established stations reveals strong agreement (Spearman's r = 0.69). This study demonstrates the value of multi-source open satellite data in permafrost research. The proposed PTI framework provides a scalable approach for monitoring permafrost dynamics and supporting climate impact assessments over the northern high latitudes. The ensemble products of this study, including the PTI map and the ensemble land cover and permafrost distributions (PP and MAGT), can be freely downloaded at https://doi.org/10.5281/zenodo.19148960 (Guo et al., 2026).

1 Introduction

Permafrost is a key element of the terrestrial cryosphere. As estimated by the International Permafrost Association (IPA), permafrost covers almost 25% of the Northern Hemisphere (Brown et al., 2002), and its perennially frozen soil and deposits contain 50% of global underground organic carbon pool (Hugelius et al., 2013). Under ongoing climate warming, ground

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

temperature increases, destabilizing the long-term thermal equilibrium of frozen soils that results in widespread permafrost warming and thawing (Biskaborn et al., 2019; Smith et al., 2022). As permafrost thaws, frozen organic carbon decomposes, and the released greenhouse gases enhance atmospheric warming, creating positive carbon-climate feedback (Natali et al., 2021). Meanwhile, permafrost dynamics are influenced by environmental characteristics such as snow cover, vegetation, hydrology, soil composition, ice content, and topography. Therefore, understanding permafrost thaw in northern latitudes requires a complex, integrated framework coping with climate, ecosystems, and spatial variability of the environments.

Global efforts have been made to advance our understanding of permafrost distributions and climate responses, including the Global Terrestrial network for Permafrost (GTN-P) that operates more than 1,000 boreholes and monitoring sites to collectively measure permafrost temperature and active layer depth (Biskaborn et al., 2015). Permafrost landscapes are vast and difficult to access. These ground-based records, while essential, are spatially limited and cannot fully represent global permafrost conditions. As the earliest mapping efforts, the IPA published the 1:100,000 paper maps of the Circum-Arctic permafrost extent and ground ice content based on the knowledge of regional experts that may trace back to the 1950s in some regions (Brown et al., 1997). The IPA map characterizes four permafrost types based on the continuity in the specified permafrost region: continuous permafrost (>90% of the land is underlaid by permafrost), discontinuous permafrost (50–90%), sporadic permafrost (10–50%) and isolated patches (<10%). Although highly utilized in current literature of permafrost studies, these manually traced maps lack spatial variability and contain high uncertainties at local scales.

Taking advantage of satellite sensors, thermal remote sensing and ground temperature models have been coupled to rebuild global permafrost maps with improved spatial detail. Relying on MODIS thermal imagery and thermal equilibrium models, Obu et al. (2019) developed the first km-scale northern hemisphere permafrost map. Through multiple machine learning and deep learning of a rich set of MODIS imagery, in-situ permafrost records, and environmental variables, the Northwest Institute of Eco-environment and Research (NIEER), China, released a set of 1-km permafrost occurrence and hydrothermal-based zonation maps of the north hemisphere (Ran et al., 2022). Some studies employed numerical permafrost model simulations that integrate remotely sensed data to obtain spatially continuous information of the permafrost state. Among these efforts are the Permafrost_CCI map products released by the European Space Agency (EPA) Climate Change Initiative (CCI), which contain yearly products of permafrost extent fraction, ground temperature and active layer thickness in the Arctic and northern high mountains (Westerman et al., 2024).

Satellite-based Earth observation (EO) has also become commonly applied to explore the causal connections of permafrost degradation in the landscape scale. In the open data era, massive open-access EO datasets have become available. For example, serving as the public interface of NASA's Earth Science Data Systems Program, NASA Earthdata (https://earthdata.nasa.gov) provides open, standardized satellite data products with long-term, comprehensive global coverage. Among these open-access products include various terrestrial variables related to thermal state of permafrost, e.g., land surface temperature, snow cover, vegetation state, and freeze-thaw state (Trofaier et al., 2017). Many of these environmental variables have been identified as the Essential Climate Variables (ECVs) as they are highly relevant to the

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

climate system and are globally available via satellite observations (Bojinski et al., 2014). International programs such as the ESA CCI have dedicated to produce geospatial data products of the ECV variables in a yearly basis and global scale (ESA, 2024).

This study explores the currently available permafrost maps and a large set of open-access EO datasets in the northern high-latitude permafrost region to examine the environmental influences on permafrost thaw. In an ensemble machine learning framework, a permafrost thaw index (PTI) is proposed to assess the stability of northern high-latitude permafrost in a warming climate. The PTI map in this study demonstrates the feasibility of using open Earth observation data for risk-aware, decision-support assessment in permafrost research.

2 Data and Methods

2.1 Available data products

This study explores the Arctic-boreal zone in the northern high latitudes (> 45°N) where approximately 73% of the land is underlain by permafrost (Brown et al., 2002). The mountain permafrost in lower latitudes such as the Tibetan Plateau is not considered since its ecosystems and climate responses differ from those of the high latitudes. Upon the IPA map, permafrost in the study region primarily lies in the territories of Russia, United States (Alaska), Canada, Denmark (including Greenland), and Mongolia, with smaller portions in European countries like Finland, Sweden, Norway and Iceland. Permafrost in the high latitudes falls in cold lands with seasonal characteristics of dry summers and cold winters, as defined by the Köppen-Geiger climate scheme (Beck et al., 2008). The mean annual air temperature is below 0°C in most of the region and can be as low as -25°C in the high Arctic (Fick and Hijmans, 2017). Based on the World Wildlife Fund (WWF) biogeographic regionalization of terrestrial ecoregions (Olson et al., 2001), the dominant biomes are tundra and boreal taiga extending to montane forests and steppes growing in the southern margin. Glaciers and waterbodies are excluded in this study.

Since the IPA maps were published in the 1990s, several global permafrost maps have been developed from satellite remote sensing and ground thermal models. This study also compiled a comprehensive set of environmental data layers to characterize the thaw processes (Table 1). Except for soil data including SoilGrids and Global Peatland Map (GPM v2), all datasets used in this study are satellite-based open data products.

Table 1. Permafrost maps and environmental open data products explored in this study.

Product name	Variable	Grid Size	Source Data	Time Period	Reference
Permafrost maps					
IPA map	Permafrost extent (4 ranks)	1:10,000,000; vector	Manual; expert knowledge	1950s-1990s	NSIDC (Brown et al., 2002)
	Ground ice content				

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Open Access Earth System Science Data Discussions

	(3 ranks)				
IRYP v2 map	Ice-rich Yedoma Permafrost (circum-Arctic)	vector	Field sites; previous maps	1976-2014	PANGAEA (Strauss et al., 2022)
Obu map	Permafrost percent (%) MAGT (°C)	1 km	MODIS; ERA5	2000-2016	APGC AWI (Obu et al., 2018)
NIEER map	Permafrost percent (%) MAGT (°C) ALT (m)	1 km	MODIS; AVHRR; EO-derived open data	2000-2016	NIEER (Ran et al., 2022)
Permafrost_CCI map (annual)	Permafrost percent (%) MAGT (°C) ALT (m)	0.01°	MODIS; other ESA CCI products; ERA5	1997-2021	ESA CCI (Westermann et al., 2024)
Land Cover					
GLCC	Global Ecosystems	1 km	AVHRR	1992-1993	USGS EROS (2018)
CAVM v2	Tundra land cover	1 km	AVHRR; MODIS; Elevation	1993/1995, 2000-2009	Mendeley Data (Raynolds & Walker, 2022)
Land Cover_CCI (annual)	Land covers	300 m	MERIS; AVHRR; SPOT-VGT; Sentinl-3	1992-2022	ESA CCI (Defourney et al., in preparation)
Land Surface Variables (repetitive)					
MEaSUREs PER-FT v2 (daily)	Freeze-thaw state	6 km	AMSE-E, AMSR2	2002-2021	NSIDC (Kim et al., 2021)
MOD10CM (monthly)	Snow Cover (%)	0.05°	MODIS	2000-2025	NSIDC (Hall & Riggs, 2021)
GLASS LST (monthly)	All-weather Land Surface Temperature	1 km	MODIS	2000-2020	GLASS (Li et al., 2024)
GLASS NDVI (8-day)	Reconstructed NDVI	5 km	MODIS	2000-2024	GLASS (Xiong et al., 2023)
Topography					
SRTM GL30	DEM	30 arc second	SRTM; GTOPO30	2000	WorldClim (Fick & Hijmans, 2017)
Soil					
SoilGrid v2	Soil organic carbon Bulk density Coarse fragment Vol. water content	1 km	Available soil profiles; environmental covariates	N/A	ISRIC (Poggio et al., 2021)
GPM v2	Peat in soil	30 arc second	Scientific publications; external sources	N/A	UNEP GPI (Greifswald Mire Centre, 2022)

NSIDC–National Snow and Ice Data Center (USA); APGC AWI–Arctic Permafrost Geospatial Centre, Alfred Wegener Institute (Germany); NIEER–Northwest Institute of Eco-Environment and Research (China); ESA CCI–European Space Agency Climate Change Initiative; GLCC–Global Land Cover Characterization (USA); USGS EROS–U.S. Geological Survey/Earth Resources Observation and Science Center (USA); CAVM–Circumpolar Arctic Vegetation Map (Alaska Geobotany Center, USA); ISRIC–International Soil Reference and Information Centre (also known as World Soil Information); UNEP GPI–United Nations Environmental Programme, Global Peatlands Initiative; GLASS–Global Land Surface Satellite product suite (China).

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

2.1.1 Permafrost maps

The digitized IPA Circum-Arctic Permafrost Map is available at the U.S. National Snow and Ice Data Center (NSIDC) (Brown et al., 2002). It includes vector boundaries of four permafrost regions (continuous, discontinuous, sporadic, and isolated patches), as well as ground ice contents and landforms. However, it lacks fine-scale spatial detail and has high uncertainty especially along its southern boundaries (Wang et al., 2023). The IPA map serves as the reference for this study. Another reference is the Ice-rich Yedoma Permafrost map (Strauss et al., 2022). It is a vector shapefile that delineates the Yedoma domain across the Circum-Arctic.

Using MODIS thermal imagery from 2000 to 2016, Obu et al. (2019) published the first km-scale global permafrost map based on the Temperature at the Top of Permafrost (TTOP) approach. The TTOP is a simple thermal equilibrium model of permafrost ground that infers ground temperatures from thaw and freezing degree days derived from EO-based land surface temperatures. As an equilibrium scheme, it does not account for the transient response of ground thermal regime to changing climate forcing.

With a comprehensive set of 1,000+ boreholes and ground observation sites, environmental variables, and satellite thermal imagery from 2000 to 2016, Ran et al. (2022) produced the 1 km permafrost maps for the north hemisphere. The maps were derived as the ensemble means of four statistical, machine learning and deep learning models, with each model going over 1,000 runs of resampled training dataset. The dataset includes the mean annual ground temperature (MAGT) at the depth of zero annual amplitude, the active layer thickness (ALT), and the probability of permafrost existence that is a proxy of permafrost fraction (%) within each pixel. The dataset is released at the Northwest Institute of Eco-Environment and Research (NIEER), China, hereafter referred to as the NIEER map.

More recently, the European Space Agency (EPA) Climate Change Initiative (CCI) released the 1 km annual Permafrost_CCI datasets from 1997 to 2021 (Westermann et al., 2024). Land surface thermal properties were derived from MODIS LST products supplemented with ERA-5 reanalysis data. An ensemble ground thermal model numerically solves the heat conduction equation within the thermal profile from ground surface downwards. The Permafrost_CCI maps include the 0.01º mean annual ground temperature (MAGT), annual ALT, and permafrost fraction (%), which is the in-pixel fraction of permafrost existence averaged across simulations from seven ensemble model members. Consequently, the permafrost fraction data contains seven discrete values ranging from 14% (1/7) to 100%.

2.1.2 Land cover maps

Several published maps are utilized to characterize land covers in the study region. The Global Land Cover Characterization (GLCC) dataset provides a historical basemap of global vegetation cover (USGS EROS, 2018). Derived from 1 km 10-day NDVI composites of AVHRR imagery collected in 1992-1993, the GLCC Global Ecosystems map classifies the Earth's surfaces according to ecological characteristics, grouping land covers into homogeneous categories based on vegetation

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

structures, phenology, and primary productivity. Although outdated, it provides a valuable reference for diverse Earth surfaces from a consistent, ecosystem-based perspective.

Under the ESA CCI project, the annual 300 m global land cover maps, LandCover_CCI, were released for the period 1992–2022 (Defourny et al., in preparation). These map products are generated using satellite image series from the Medium Resolution Imaging Spectrometer (MERIS), together with other sensors like AVHRR, SPOT-Vegetation and Sentinel-3. The maps include 22 land cover classes and 15 subclass composition levels. Such detailed information on biophysical properties may better characterize ecosystem impacts on permafrost thaw.

Particularly for the Arctic tundra above the latitudinal tree line, a satellite-classified Circumpolar Arctic Vegetation Map (CAVM) was released (Raynolds et al., 2022). This map identifies three vegetated tundra types based on Arctic flora: graminoid, prostrate dwarf shrub (< 15 cm), and erect dwarf shrub (15-40 cm). as well as barren tundra and wetland tundra. The CAVM tundra falls mostly in the continuous permafrost region.

2.1.3 Satellite-derived land surface variables

The monthly 0.05° MOD10CM snow cover (%) products from July 2000 to June 2025 were downloaded from NSIDC (Hall & Riggs, 2021). Data is missing for six months (August 2000, June 2001, March 2002, December 2003, February 2016, and October 2022). For these missing months, a 10-year average of the corresponding MOD10CM products is used as replacement.

The 8-day, 5 km Normalized Difference Vegetation Index (GLASS NDVI) dataset for the period 2000–2024 is available in the China Global Land Surface Satellite (GLASS) products suite (Xiong et al., 2023). This dataset provides global NDVI time series reconstructed from MODIS surface reflectance. The reconstruction effectively reduces cloud-induced missing data and noises in the NDVI time series.

The monthly 1 km Land Surface Temperature (GLASS LST) dataset is also available in the GLASS product suite (Li et al., 2024). The dataset provides global monthly LST for the period 2000–2020. The LST is derived using ensemble machine learning of multiple satellite datasets, including MODIS LST (MOD11) and other MODIS products, as well as ERA reanalysis and in situ measurements. The January 2000 layer is missing; it is replaced with the 10-year average of the January layers, 2002-2009.

Through the NASA MEaSUREs (Making Earth System Data Records for Use in Research Environments) program, the 6 km daily Polar Enhanced Resolution Freeze–Thaw (PER-FT v2.1) product was released at NISDC (Kim et al., 2021). For the period 2002-2021, this dataset provides the daily freeze/thaw state derived from passive microwave sensors of AMSR-E and AMSR2. It characterizes the seasonal and interannual freeze/thaw state of the land surface with an enhanced resolution across northern high latitudes.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

2.1.4 Environmental confounding layers

In a warming climate, changes on land surface contribute to permafrost thaw. However, the causal pathways may vary due to confounding environmental factors such as topography and soil properties. Elevation data at 30 arc-second resolution is available at WorldClim (Fick and Hijmans 2017). The DEM in this data product is from the near-global Shuttle Radar Topography Mission (SRTM). The GTOPO30 DEM used in areas where SRTM data is unavailable (i.e., north of 60 °N).

Soil properties in the active layer influence water draining capacity and thermal conductivity, leading to thermal offset at the permafrost table (Jorgenson et al., 2010). The SoilGrids database characterizes global soil properties using state-of-the-art machine learning based on 230,000 soil profile observations and a pool of 400 environmental covariates worldwide (Poggio et al., 2021). The 1 km resolution soil maps are produced across six standard depth intervals: 0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm. This study examines four soil parameters from the SoilGrids database: soil organic carbon content, volumetric fraction of coarse fragments, bulk density, and volumetric water content that fairly represents soil moisture availability.

Another soil parameter considered in this study is peat in the active layer, representing areas with a naturally accumulated layer of peat soil in which plant materials are partially decomposed due to the cold and water-saturated conditions. In 2022, the UNEP Global Peatlands Initiative conducted global peatlands assessment to provide the first comprehensive global overview of the world's peatlands (UNEP, 2022). The Global Peatland Map 2.0 (GPM2.0) was produced based on extensive scientific literature and external data sources on soil and peatland research. The GMP2.0 raster layer includes two binary classes (peat dominated vs. peat in soil mosaic) at 30 arc-second resolution.

2.2 Data ensemble approaches

2.2.1 Building an ensemble environmental dataset

Three land cover maps listed in Table 1 (LandCover_CCI, GLCC Global Ecosystems, and CAVM) are available for the study region. Due to their different classification schemes, the class names and corresponding land cover definitions do not fully align. Land cover classes across these maps may represent distinct land types despite having similar names, or conversely, share similar characteristics under different class names. This study integrates these three maps to build an ensemble land cover product.

The LandCover_CCI serves as the primary map due to its higher resolution (300 m) and more recent release. By applying a majority filter to annual maps in 2018–2022, the land cover at each 300×300 m^2 pixel is determined as the most frequently classified type over this five-year period. The resulting map is then visually investigated and refined using the GLCC Global Ecosystems map as a reference. Some classes are merged or reclassified to maintain consistency in ecological characteristics. For example, deciduous coniferous forest, defined in the LandCover_CCI map as "tree cover, needleleaved, deciduous, closed to open >15%", is the most dominant class (24.33%) within the study region. Visual comparison between the two

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

global products indicates that the corresponding GLCC classes include “small leaf mixed woods” and “deciduous and mixed boreal forest”. The second-largest class, evergreen coniferous forest, covers 12.58% of the study region. Its corresponding classes in the two products are more diverse, reflecting the geographically heterogeneous distribution of this land cover type. Based on visual comparison, three LandCover_CCI classes and five GLCC classes are grouped into this ensemble category. The re-coding between the two global maps is summarized in Table 2.

Tundra is a key biome in permafrost regions. It is primarily distributed across the Arctic, with isolated patches of alpine tundra in high mountains. Tundra has not been well mapped in the LandCover_CCI product, where it is broadly classified into barren tundra, wooded tundra and upland tundra. A substantial portion of Arctic tundra mis-labeled as polar and alpine desert. To improve tundra classification, the CAVM map is used to represent tundra lands above the latitudinal tree line. All barren classes in the CAVM map are recoded as barren tundra; graminoid-dominated classes as herbaceous tundra; prostrated dwarf-shrub dominated classes as dwarf tundra; erect dwarf-shrub dominated classes as low shrub tundra; and wetland classes in the Arctic as wetland tundra. In subarctic regions beyond the extent of the CAVM map, alpine tundra defines the high-altitude biome and is represented by several classes in the GLCC Global Ecosystems map. In this study, an alpine tundra mask layer is created using the four tundra classes in the GLCC map (9,53,63,69), as listed in Table 2. Within this GLCC tundra mask, the LandCover_LC classes (bare areas, lichens and mosses, sparse vegetation) that matches the GLCC alpine tundra are recoded barren tundra, herbaceous tundra, and low shrub tundra, respectively. Table2 summarizes the ensemble classes after interactive examination of the three open-access maps.

Table 2. Class codes and land cover names of the ensemble land cover map based on three open-access maps.

Ensemble code	Ensemble land cover name	Percent cover (%)	Open-access land cover maps (code)		
			LandCover_CCI	GLCC Global Ecosystems	CAVM
10	Cropland and Herbaceous Cover	1.01	10, 11, 20	30, 31, 35, 36, 37, 38, 93	
30	Mosaic Cropland/Natural Vegetation	0.91	30, 40	19, 55, 56, 57, 58, 94	
60	Deciduous Broadleaf Forest	1.11	50, 60, 61, 62	5, 23, 25, 26	
70	Evergreen Coniferous Forest	12.58	70, 71, 72	3, 20, 21, 22, 62	
80	Deciduous Coniferous Forest	24.33	80	60, 61	
90	Mixed Forest	0.82	90	14	
100	Mosaic Tree/ Shrub/ Herbaceous Cover	5.30	12, 100, 110	7, 13, 16, 27, 29, 32, 33, 90, 92, 96	
120	Shrubland	5.98	120, 121, 122	17, 37, 64	
130	Grassland	4.34	130	40, 41, 42, 43, 91	
150	Sparse Vegetation	1.89	140*, 150*, 152, 153	2, 51, 52	
160	Forested Wetland	0.81	160		
180	Shrub/Herbaceous wetland	3.76	180	44, 45	
190	Urban	0.04	190	1	
200*	Bare land	0.39	200*	11,50	
201	Barren Tundra	9.01	201, 200	53,69	B1, B2a,

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Open Access Earth System Science Data Discussions

					B2b, B3, B4
202	Herbaceous Tundra	8.65	202, 140*		G1, G2, G3, G4
203	Dwarf Shrub Tundra	3.59	203		P1, P2
204	Low Shrub Tundra	8.81	204, 150*	9,63	S1, S2
205	Wetland Tundra	1.18	205		W1, W2, W3
210	Water	5.33	210	14	FW
220	Glacier Ice	0.14	220	12	GL

*Three LandCover_CCI classes within the GLCC Alpine tundra mask are recoded: “140–Lichens and mosses” to “202–Herb tundra”; “150–Sparse veg” to “204–Low shrub tundra”; “200–Bare areas” to “201–Barren tundra”.

The per-pixel land surface variables, including vegetation cover, snow cover, LST, and freeze/thaw state, are ensembled from published EO data series as listed in Table 1. Frequent satellite observations allow us to detect their seasonal cyclical patterns and long-term means. For the 5 km, 8-day GLASS NDVI series (2000–2024), the annual peakNDVI is defined as the maximal NDVI value within one calendar year, and its 25-year average is used to represent the peak greenness. The annual ΣNDVI is calculated by accumulating all NDVI values greater than 0 within one year. Its 25-year average is used to represent the total greenness within one year. Additionally, seasonal Mann-Kendall trend analysis (Hirsch et al., 1982) is applied to the 8-day NDVI series over the 25-year period. For pixels with significant trend ($p < 0.05$), the Sen’s slope at each pixel is used to represent the NDVI trend. This method adopts a harmonic trend analysis to preserve seasonality when quantifying trends in both directions. Similarly, for the 1 km, monthly GLASS LST series (2000–2020), all LST values in the time series are averaged to represent the mean land surface temperature at each pixel. A Seasonal Mann-Kendal trend analysis is again applied to detect the significant LST trends ($p < 0.05$).

A snow year is defined as a 12-month period from July of one year through June of the following year. Using the 0.05° (approximately 5 km) monthly MOD10CM snow cover series from July 2000 to June 2025, the average snow cover at each pixel is calculated to represent the overall snow cover condition. A seasonal Mann-Kendall trend analysis is applied to detect harmonic trends in the 25-year span. For pixels with significant trends ($p < 0.05$), the Sen’s slope at each pixel is calculated to represent the snow cover trend.

The last satellite-derived land surface variable in this study is the freeze/thaw state. Using the daily, 6 km FER-FT (V2.1) data series, the total number of frozen days is counted for each calendar year. At each pixel, these values are averaged over 2002–2021 to represent the general freeze-thaw condition. A regular Mann-Kendal trend analysis is applied to the annual frozen-day series to determine whether significant trends are present.

Soil properties and topography are considered environmental confounders, as they remain relatively static factors that indirectly affect permafrost thaw by constraining the spatial distributions of the aforementioned land surface variables. Among the four soil variables in the 1 km SoilGrids database, SOC refers to soil organic carbon content in the fine earth fraction; FC denotes the volumetric fraction of coarse fragments (> 2 mm); BD is the bulk density of the fine earth fraction; and VMC is the volumetric water content at 33kPa, a common measure of soil moisture availability after excess gravitational

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

water has drained out. For each soil variable, a weighted average is calculated from the SoilGrids layers across five depth intervals (0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, and 60-100 cm). Soil peat content is available from the 1 km GMP2.0 global peat map, which delineates peatland soils into two binary categories: peat dominated soil and peat-soil mosaic. The remaining are defined as non-peat lands. Elevation and slope are extracted from the 1 km DEM data.

In total, sixteen environmental variables are ensembled in four regimes: thermal (LST, LST trend, and Frozen Days), biophysical (land cover, peakNDVI, sumNDVI, NDVI trend), and hydrological (snow cover, snow cover trend), and environmental constraints of soil (SOC, CF, BD, VMC, peat) and topography (elevation, slope). To construct the ensemble dataset, all data layers are resampled to 1 km grid. Each variable is standardized to a range of [0,100] using a min-max linear stretch. Exceptionally, the ensemble land cover map remains categorical, retaining its original class codes 10–220 (21 classes). The peat map is re-coded as 0 (no peat), 50 (peat in soil) and 100 (peat dominated). This ensemble dataset is used to model the environmental influences on permafrost thaw.

2.2.2 Building an ensemble permafrost dataset (PP, MAGT, ALT)

At the time this study was conducted, four permafrost products were available for the northern high latitudes: the IPA map, the Obu map, the NIEER map, and the annual ESA Permafrost_CCI maps in the period 1997–2021 (see Table 1 for details). The IPA map serves as a reference of this study. All three other maps measure the fractions of permafrost existence within each pixel, hereafter referred to as permafrost percent (PP) maps. The Obu and NIEER products each provide a single PP map. For the annual Permafrost_CCI product, a 25-year mean filter is applied to extract the overall permafrost existence. Geographically, the three PP maps show good agreement in areas of continuous permafrost. However, in the southern, non-continuous permafrost regions, substantial discrepancies exist, particularly along the southern ends.

In this study, the three PP maps are integrated to generate an ensemble dataset representing permafrost distributions. Areas with PP > 0% in at least two of the three maps are counted as permafrost existence, thereby establishing a new permafrost base map. The median of the three PP layers is used as the percent cover of permafrost on the base map.

An ensemble MAGT map is also derived from the three permafrost products. The Obu and NIEER products estimate the MAGT at top of the permafrost table (i.e., the bottom of the active layer). The annual Permafrost_CCI provides ground surface temperature (GST) estimates at multiple depths (0-30 cm, 1 m, 2 m, 5 m, and 10 m) from an EO-constrained thermal model. In permafrost regions with ALT < 1 m, e.g., in the continuous permafrost zone of the IPA map, the GST at 1 m depth is only slightly higher than that at 2 m depth. Therefore, this study adopts the GST map at 2 m depth as a proxy for MAGT, and a general MAGT layer is extracted by averaging all layers across the 25-year span. To reconcile the product-level differences, the median of these three MAGT products (Obu, NIEER, Permafros_CCI) is extracted to represent the MAGT across the study region. In addition, a Mann-Kendell trend analysis is performed on the annual Permafrost_CCI MAGT time series to assess the MAGT trend in the 25-year span. Permafrost regions exhibiting a significant positive MAGT trend ($p < 0.05$) are considered more sensitive to thaw.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

The Permafrost_CCI product also includes the annual ALT, which is defined as the depth of maximum seasonal thaw simulated by its thermal model. In this study, the Permafrost_CCI ALT layers from 1997 to 2021 are averaged to represent the general state of active layer thickness. The resulting map shows that the active layer thickness across most areas of the northern high latitudes (> 45 ºN) is less than 2 m, supporting our assumption that GST at 2 m depth can serve as a proxy for MAGT. Similarly, a Mann-Kendell trend analysis is conducted to assess the 25-year ALT trend. Permafrost areas exhibiting a significant positive ALT trend ($p < 0.05$) are considered more sensitive to thaw. The NIEER product also provides a general ALT map. However, it does not fully cover the study region, and its spatial patterns differ substantially from those of the Permafrost_CCI ALT. Therefore, the NIEER ALT is not further considered in this study.

2.2.3 Extracting a permafrost thaw index (PTI) data layer

Permafrost thaw is a multi-faceted process controlled by its thermal regime. Permafrost becomes susceptible to thaw when its ground temperature approaches the melting point of ice. Permafrost with a higher MAGT is more vulnerable to thaw than permafrost in very cold regions. Meanwhile, permafrost thaw is strongly regulated by long-term climatic change. As climate warming increases air and land surface temperatures, heat is transferred downward into the ground layer of permafrost, causing the MAGT to rise. Along with permafrost thaw, the ALT also becomes thicker. Therefore, in addition to permafrost coldness represented by a static MAGT value, investigations of permafrost thaw should also consider changes in both MAGT and ALT to better characterize thermal stability. The 25-year Permafrost_CCI time series of MAGT and ALT products enable us to extract their relatively long-term trends.

Climate warming and permafrost thaw dynamics are spatially heterogeneous. By integrating multi-source permafrost products in a spatial framework, this study proposes a permafrost thaw index (PTI) using a two-tier ranking system based on two permafrost variables derived above: (1) the ensemble MAGT of multi-decadal mean, and (2) the MAGT|ALT trend extracted from the 25-year Permafrost_CCI time series of MAGT and ALT (Table 3). Within each tier, Rank 1 represents the most stable conditions, whereas Rank 3 indicates the highest thawing potential.

Table 3. Permafrost thaw index (PTI) ranking system and sampling points of each PTI class.

Permafrost type	MAGT (ºC)	MAGT|ALT* trend	PTI Rank	Samples
Cold	< -3	None	11	6,371
		One is positive	12	8,776
		Both are positive	13	4,348
Cool	[-3, -1.5]	None	21	4,087
		One is positive	22	4,758
		Both are positive	23	4,391
Warm	>-1.5	None	31	5,362
		One is positive	32	9,094
		Both are positive	33	8,842
Isolated patches**			40	5,296

* The symbol "|" represents the logic operator "or". ** Pixels with PP <10% in the permafrost base map are directly counted as Rank 40.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

The ensemble MAGT serves as the tier-1 variable in the PTI ranking system. It defines three levels of permafrost conditions following a classification scheme defined in Ran et al., (2022). Rank 1 represents the areas with MAGT < -3°C, or cold permafrost in the scheme. Rank 2 is the areas with MAGT between -3°C and -1.5°C, or cool permafrost. Previous studies (Obu et al., 2019; Ran et al., 2022) have shown that permafrost may persist in areas with MAGT slightly above 0 °C, especially in regions undergoing transitional degradation. To be inclusive, Rank 3 in this study covers the areas with MAGT between -1.5°C and 3°C as warm permafrost. Marginal areas in the permafrost base map with MAGT > 3°C are considered non-permafrost and are excluded from the analysis.

The MAGT|ALT trend from the Permafrots_CCI data series serves as the tier-2 variable in the ranking system, capturing areas with significantly increasing trends over the past 25 years. In this tier, Rank 1 is assigned to the areas showing no significant trend in either MAGT or ALT. Rank 2 is to the areas where one of the two variables exhibits a significantly increasing trend. Rank 3 is to the areas where both MAGT and ALT show significantly increasing trends.

Isolated patches (PP < 10%) in the ensemble PP map are highly fragmented along the southmost margin of the northern high-latitude permafrost. At such low permafrost fractions, these areas are most likely undergoing degradation. These pixels are therefore classified as the most vulnerable in the PTI framework. A separate Rank 40 is assigned to represent these areas that are extremely vulnerable to thaw.

With the ensemble MAGT (multi-decadal mean) and MAGT|ALT trends, a conceptual PTI layer is built using the rankings scheme in Table 3. The map provides a general indication of permafrost thaw in a warming climate. Permafrost thaw, however, is a complex process interacting with environmental conditions. Permafrost stability at the local scale varies with environmental characteristics, such as thermal conditions, land cover, snow cover, soil properties, and topography, as included in the ensemble dataset in Table 1. The environmental influences must be systematically evaluated to properly quantify the thaw potential and assess permafrost stability.

2.2.4 Machine learning of the ensemble dataset for PTI prediction

Samples are randomly collected on the permafrost base map. The PTI and associated environmental variables at sample points are extracted from the conceptual PTI map and the ensemble environmental dataset. A minimum point-to-point distance of 25 km is enforced, and points within a 10 km buffer along the polygon boundaries of each PTI class are excluded to reduce edge effects. Exceptionally, for PTI =40, the minimum point-to-point distance is reduced to 10 km and the edge buffer to 5 km due to the highly fragmented patterns of the isolated patches. The sample dataset contains approximately 4,000 to 9,000 sample points for each PTI class (Table 3).

With this sample dataset, a widely used and powerful machine learning algorithm, eXtreme Gradient Boosting (XGBoost), is applied to perform PTI classification by integrating a complex set of environmental variables. XGBoost is an ensemble learning method that implements gradient-boosted decision trees for supervised classification (Chen and Guestrin, 2016). By

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

combining numerous weak decision trees, it constructs an optimized predictive model capable of handling large and heterogenous training datasets. Its ensemble learning framework spreads the training among model ensembles (e.g., random trees and boosted trees) and multi-source ensembles to achieve generalized prediction.

Not all environment variables may contribute significantly to PTI prediction. To mitigate overfitting and reduce data redundancy, we implement a rigorous feature selection procedure, Recursive Feature Elimination with Cross-Validation (RFECV), to identify the environmental influencers. RFECV is a wrapper-based method that iteratively trains the model and prunes the least important features, thereby identifying the optimal subset of variables that maximizes predictive performance (Guyon et al., 2002). The XGBoost classifier is the base estimator for the RFECV process, with 16 environmental variables as the input features. XGBoost is initialized using the full set of features. After model training, feature importance is quantified from the gain metric. The feature contributing least to model performance is eliminated, and the model is trained again using the remaining features. A Stratified 5-Fold Cross-Validation strategy is employed, with classification accuracy serving the evaluation metric of each feature. Through iterative elimination and re-evaluation, the optimal set of environmental variables are identified for the final XGBoost classification. This procedure enhances the robustness of feature selection and minimizes the bias associated with sample partitioning.

To further improve model stability, the XGBoost classification is run 100 times to generate an ensemble prediction of the ranked PTI classes. In each run, 80% of the samples are randomly selected for training and the remaining 20% are used for validation. The final output of the XGBoost model is an optimized PTI map representing permafrost thaw vulnerability across the study region.

3 Results and Discussion

3.1 Spatial distributions of permafrost (PP, MAGT, ALT)

Permafrost in the study region is predominantly distributed across North America (northern Canada and Alaska, USA) and North Asia (particularly Siberia, Russa), with a smaller portion extending into Europe west of the Ural Mountains (Fig.1). The three published permafrost maps, including the Obu map (Fig.1a), the NIEER map (Fig.1b), and the Permafrost_CCI map (Fig.1c), aligned well in the extensive continuous permafrost (90–100%) at higher latitudes. For the non-continuous permafrost southward, their permafrost areas varied substantially. The Permafrost_CCI map had the greatest coverage of non-continuous permafrost, followed by the Obu map, whereas the NIEER map showed the least coverage. Therefore, the Permafrost_CCI map had the greatest coverage of total permafrost extent. Both the Obu and the NIEER maps relied primarily on thermal imagery as their main data source; their spatial extents of permafrost were similar, with the NIEER map slightly lower at the southern margin. These spatial differences indicate considerable uncertainties among these open-access permafrost datasets.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

The ensemble PP map (Fig.1d) of this study leveraged these three open data products. It excluded areas where permafrost was only identified in one of the three published maps. At each pixel, selecting the median value among the three maps to represent permafrost existence reduced outliers. In Fig.1d, discontinuous (50–90%) and sporadic (10–50%) permafrost zones become more pronounced, facilitating the exploration of thaw vulnerability in these sensitive regions. Fig.1d also excludes some noisy isolated patches (PP < 10%) situated at the southmost ends, which are particularly evident in the Permafrost_CCI and Obu maps.

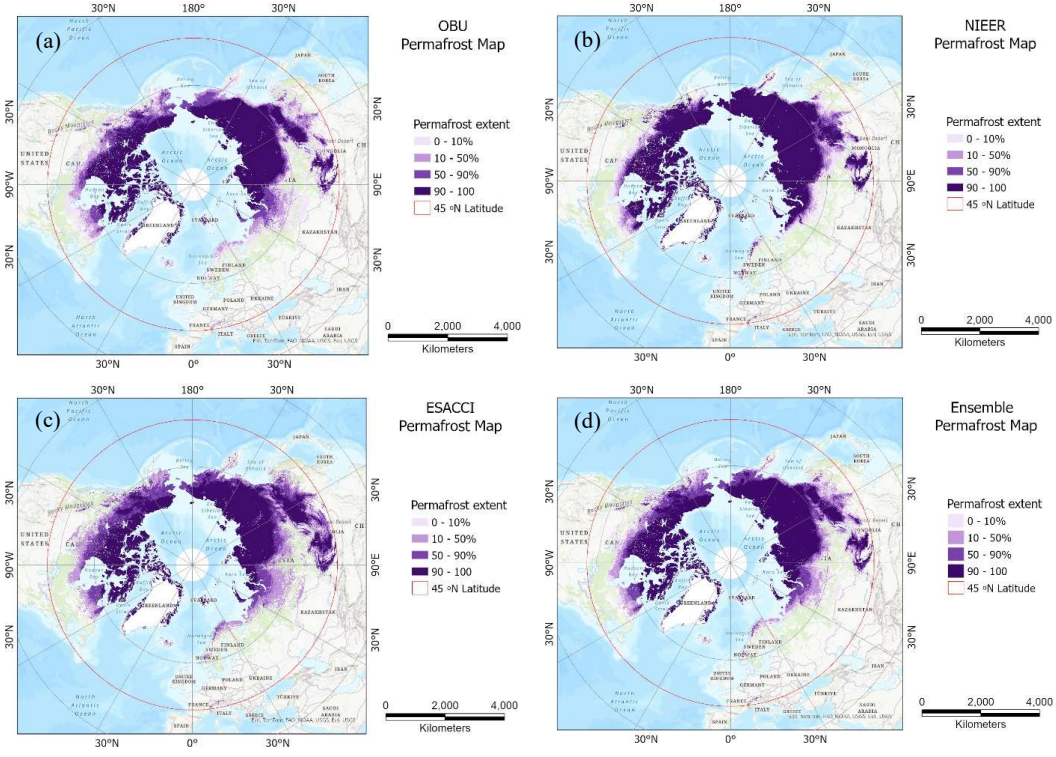
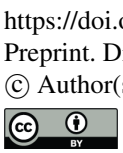


Figure 1 Permafrost percent maps: OBU (a), NIEER (b), Permafrost_CCI (c), and the ensemble PP map of this study (d).

In the MAGT maps (Fig.2), the color legend follows the classification scheme of cold permafrost (MAGT < -3ºC), cool permafrost (-3ºC < MAGT < -1.5ºC) and warm permafrost (MAGT > -1.5ºC). Yellow pixels indicate exceptionally high MAGT values (MAGT >3 ºC) that are most likely non-permafrost. Overall, the three published MAGT maps (Obu map in Fig.2a, NIEER map in Fig.2b, and the 25-year average of Permafrost_CCI maps in Fig.2c) revealed similar spatial patterns, with permafrost ground temperatures increasing from the High Arctic southward. In regions of cool and warm permafrost, MAGT values showed greater deviations among the three products. Be noted the Permafrost_CCI MAGT was approximated as permafrost ground surface temperature (GST) at 2 m depth. Interestingly, cool permafrost was limited among all three

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026



Open Access Earth System Science Data Discussions

open datasets. The NIEER MAGT map (Fig.2b) identified a relatively higher extent of cool permafrost. However, in the high-altitude mountain permafrost regions, e.g., clusters of continuous permafrost in southern Siberia, Russa, it classified these areas as cool permafrost, whereas the other two MAGT maps (Fig.2a and Fig.2c) recognized them as cold permafrost, which is typical of continuous permafrost. The ensemble MAGT map (Fig.2d) integrated these three products to more clearly delineate the zonal categorization of cold, cool, and warm permafrost. Areas with MAGT > 3ºC were excluded from subsequent data analysis.

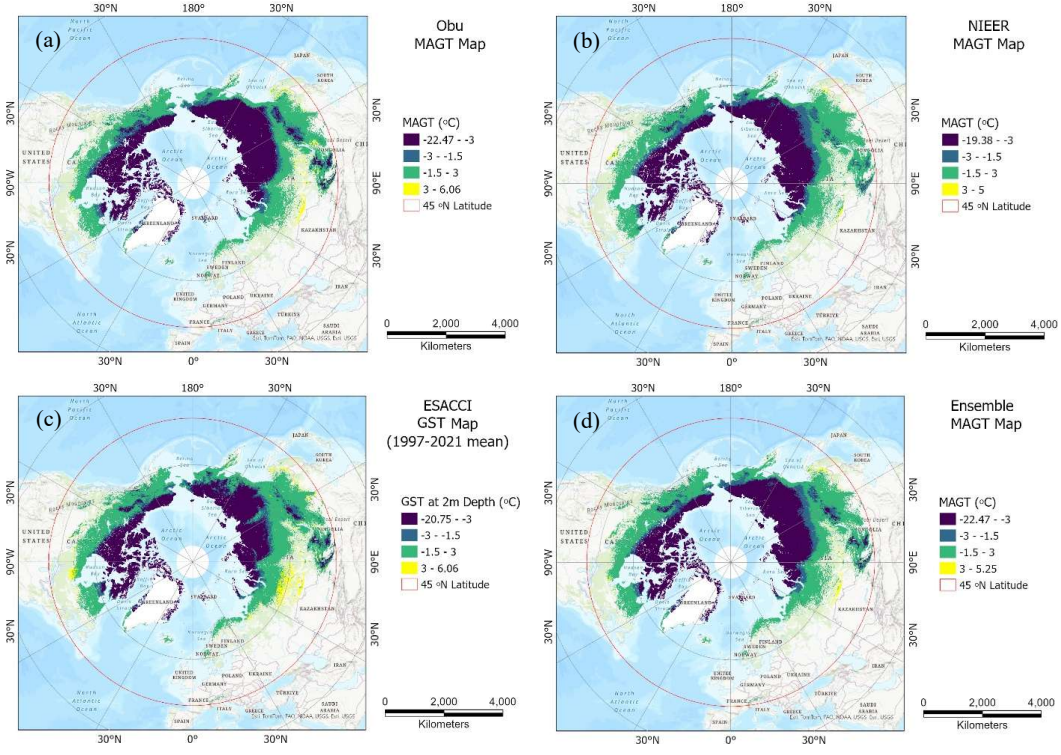


Figure 2 MAGT maps: OBU (a), NIEER (b), the 25-year average annual Permafrost_CCI GST at 2 m depth (c), and the ensemble MAGT of this study (d).

The NIEER ALT (Fig.3a) is available only in areas with PP > 50%. The 25-year average Permafrost_CCI ALT (Fig.3b) is available for all permafrost lands across the study region. Both maps in Fig.3 indicate that active layer thickness in the northern high-latitude permafrost is generally less than 2 meters that increases southward from the High Arctic. However, their absolute ALT values varied substantially, making the two maps not spatially comparable. Therefore, this study did not generate an ensemble ALT map. Fig.3b is used as the representative ALT layer for the study region.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

The Mann-Kendall trends of MAGT and ALT were derived from the annual Permafrost_CCI products. In Fig.4, the permafrost basemap (in grey) is underlaid to delineate permafrost lands (PP > 0%). Most permafrost lands in the study region exhibit significantly increasing trend in MAGT (Fig.4a). Only patches of permafrost reveal significant increase of active layer thickness (Fig.4b), yet the majority of the region remains relatively stable. While spatially limited, these patches distribute across all permafrost types from continuous permafrost in the High Arctic to sporadic and isolated patches to the south. These trends suggest ongoing permafrost warming and potential thaw, consistent with findings widely reported in the current literature (as reviewed in Smith et al., 2022).

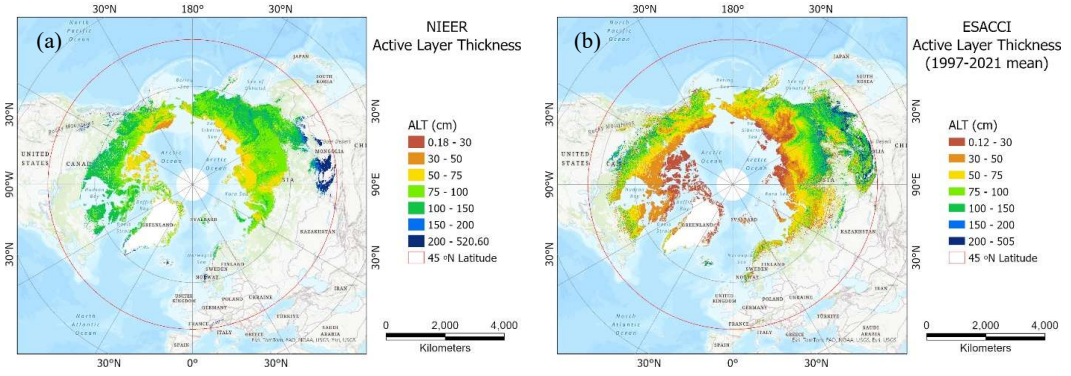


Figure 3 ALT maps: NIEER (a) and the 25-year average of annual Permafrost_CCI (b).

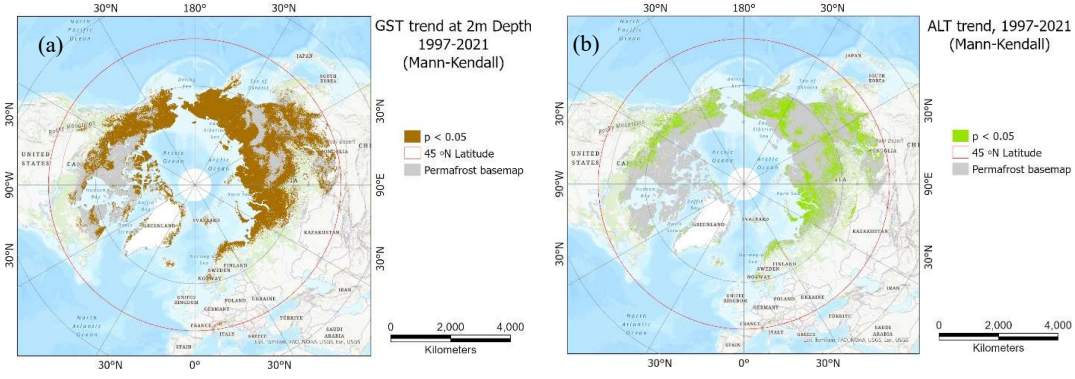


Figure 4 Permafrost trends in GST at 2 m depth (a) and ALT (b), derived from the 1997-2021 Permafrost_CCI products. The trend maps are overlaid on the permafrost base map (in grey).

3.2 Spatiotemporal characteristics of environmental variables

3.2.1 Satellite-derived land cover and land surface variables

An ensemble land cover map (Fig.5a) was developed by integrating the LandCover_CCI map, the CAVM Arctic Vegetation map, and the GLCC Global Ecosystem map. The color scheme primarily follows LandCover_LLC while the class names

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Open Access Earth System Science Data Discussions

are aligned with the GLCC. 21 classes are defined in the ensemble land cover map. The spatial distributions transition from herbaceous croplands in the south to broadleaf and needleleaf forests northward, and finally to tundra in the Arctic above the latitudinal tree line, as marked in the figure. High mountain areas exhibit a transition from coniferous forests to alpine tundra above the altitudinal tree line. Deciduous coniferous forest has the largest coverage (24.33%) and is primarily distributed across North Asia. The 2nd largest class is evergreen coniferous forest (12.58%) that is mainly found in North America. The five arctic/alpine tundra classes together account for 31.25% of permafrost land in the study region. Inland water and glacier ice are masked out in this study. Aside from a small coverage of non-vegetation classes such as barren tundra, Fig.5a can be considered a vegetation cover map with their class codes increasing from croplands (10) to tundra classes (201–205), as listed in Table 2.

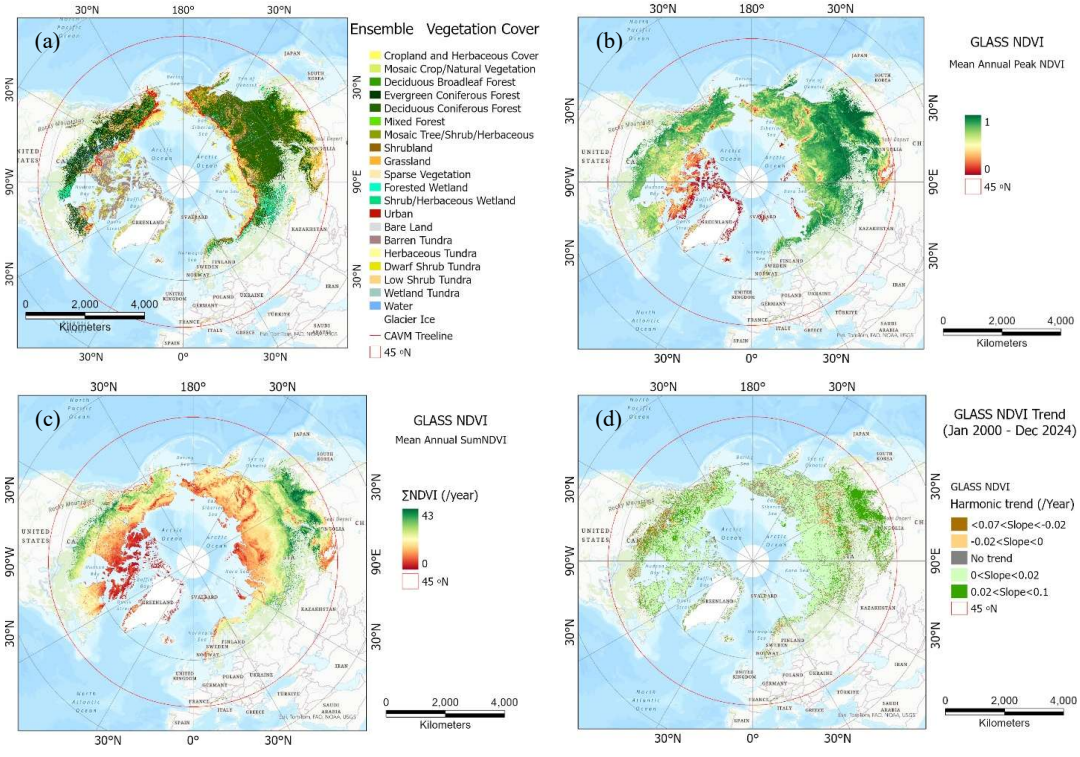


Figure 5 The ensemble land cover map with the CAVM tree line marked (a), and vegetation greenness maps of peak NDVI (b), ∑NDVI (c), and the harmonic NDVI trend (d).

Quantitative bio-characteristics in permafrost lands were examined using vegetation greenness, approximated by the 8-day GLASS NDVI series from 2000 to 2024. The mean annual peak NDVI (Fig.5b) generally aligns with biome distribution, with low NDVI values in Arctic and alpine tundra and high NDVI in boreal forests southward. Compared with the ensemble

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

land cover map, deciduous boreal conifers in North Asia exhibit higher NDVI than evergreen boreal conifers in North America. In Fig.5c, the annual accumulated NDVI (∑NDVI) characterizes seasonal dynamics across the study region. Even within the same biome, for example boreal taiga in Siberia, ∑NDVI at higher latitudes is substantially lower than that in more southerly areas. Through the Seasonal Mann-Kendall analysis of NDVI series in the 25-year span, the NDVI harmonic trends also vary spatially (Fig.5d). Most permafrost regions experienced a weak increase in peak NDVI (slope <0.02 yr^{-1}). Boreal forests at the southern end of the North Asian permafrost exhibit stronger increases with slopes > 0.02 yr^{-1}, although decreasing NDVI trends are observed in isolated areas. Harmonic slopes between -0.02 and 0.02 yr^{-1} failed to pass the seasonal Mann-Kendall test ($p < 0.05$) and were thus categorized as no trend. Such areas are very limited in extent and spatially fragmented across the permafrost region. Overall, the spatial pattern in Fig.5d is consistent with previous studies (Myers-Smith et al., 2020; Liu et al., 2024) reporting significant and spatially dynamic vegetation greening in the North Hemisphere. These findings suggest that vegetation in cold lands has been actively and positively responding to climate change.

Regarding thermal conditions, the mean annual LST for 2000–2020 ranges from -22 ºC in the High Arctic to 9 ºC at the southern end of permafrost (Fig.6a). The latitudinal and elevational gradients are clearly observed. In Fig.6b, areas with significant LST changes during 2000–2020 are shown in purplish tones. A large portion of Siberia and Alaska experienced significant LST increases, consistent with the widely recognized pattern of global warming particularly at high latitudes. Areas with the most pronounced warming are located in north-central and west Siberia, where the northernmost coastal region (e.g., Kara Sea) reveal the warming rates up to 0.2 ºC yr^{-1}. The warming rate slows down southward to about 0.1 ºC yr^{-1}. No significant negative LST trends are observed across the study region, indicating an overall warming climate over the past decades. It is noteworthy that extensive permafrost regions in Canada and subarctic North Asia (> 60 ºN) show no increase in LST, despite their relatively high temperatures on land surface.

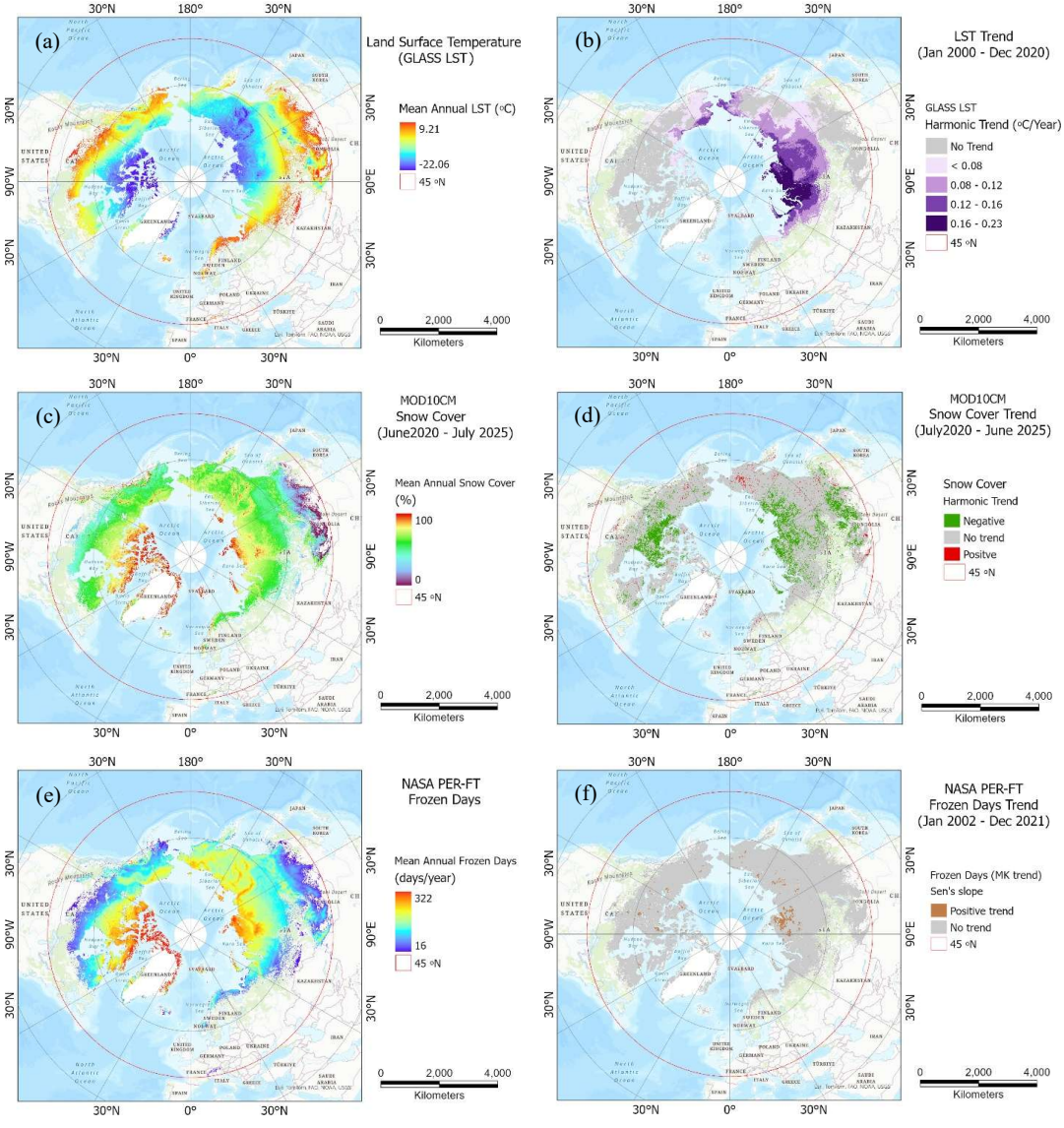
This study considered snow cover to be a critical hydro-thermal variable for characterizing permafrost landscapes. The mean annual snow cover (Fig.6c) for the 2000–2025 snow years, derived from the monthly MOD10CM snow cover products, has a clear latitudinal decrease from the High Arctic (up to 100%) to southern margin of permafrost (close to 0%). Mountainous areas also exhibit distinctively higher snow cover. In Fig.6d, majority of the study region remains stable in inter-annual snow covers. Clusters of areas with significant trends in snow cover change are observed, and the harmonic slopes are generally small. Therefore, the trends are presented in a binarily manner as negative or positive. In areas with significant trends, decreasing snow cover dominates–e.g., in Arctic Canada, central Siberia, and subarctic North Asia–indicating reduced snow cover over the past decades. Areas with increasing trends are limited and sporadically distributed, lacking a coherent geographic pattern across the study region.

The mean numbers of annual frozen days were derived from the 6 km NASA PER-FT products for the period 2002–2021, which show a clear decrease from the High Arctic southward (Fig.6e). The High Arctic tundra of North America has the greatest number of frozen days–up to 322 days per year. Influenced by topography, high mountains also have more frozen

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Open Access Earth System Science Data Discussions

days than surrounding lowlands. The Mann-Kendall trend analysis indicates the number of frozen days across most permafrost areas does not exhibit significant trends (Fig.6f). particularly, o significant negative trends were observed. Only a few isolated clusters show a significant positive trend, indicating an increase in frozen days over the 20-year span. Due to their limited spatial extent, the trends in frozen days were not further examined in this study.



https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Figure 6 Satellite-based land surface variables and their multi-decadal trends: LST (a, b), snow cover (c, d), and frozen days (e, f).

3.2.2 Confounding environmental conditions

In the SOC map (Fig.7a), areas with high soil organic content largely correspond to vegetated wetlands identified in the ensemble land cover map (Fig.5a), such as northern Ontario, Canada and western Siberia, Russa. Above the tree line, herbaceous and shrub tundra tend to have higher organic matter, for example in northern Alaska, USA. In contrast, subarctic regions (> 60 ºN) are dominated with relatively low organic soil, particularly along the southern margin of the permafrost. Soil coarse fragment content (Fig.7b) is higher in subarctic North Asia. Mountainous areas in eastern Siberia and Alaska also exhibit high coarse fragment content. Bulk density (Fig.7c) and volumetric water content (Fig.7d) display contrasting spatial patterns. Areas across Alaska generally have low bulk density and high water content, whereas much of North Asia and north-central Canada exhibits high bulk density and low water content.

The GMP Peatland map (Fig.7e) shows contrasting peatland distributions between North Asia and North America. In the High Arctic region of North Asia, i.e., Siberia (> 60 ºN), extensive peatland areas are present. In contrast, peatlands in North America are more limited, primarily clustered in tundra lands of Alaska and wetlands in northwest Canada. The GMP map also classified some areas as peat-rich soil (peat in soil mosaic) that is mostly documented in Canada. Such soil was not identified in North Asia, except for fragmented distributions along the southern end such as in the grassland steppe of Mongolia.

Peat soil is more ice-rich than other soils in permafrost regions. Particularly in north Asia, the peatland distribution generally aligns with that of ground ice content from the IPA map (Fig.7f), which depicts frozen water within the soils and rocks of permafrost regions (Brown et al., 2002). Notably, the ice-rich Yedoma permafrost mapped by Strauss et al. (2022), as outlined in Fig.7f, is largely associated with peatlands across Sebria and Alaska. Ground ice content in permafrost could also be influenced by other soil properties. Fine-grained soils (silt and clay) retain more water and can support greater ice formation than coarse-grained soils (sand and gravel). Poor drainage conditions, where water cannot easily escape, promote soil saturation and enhance ice formation, potentially leading to more rapid and widespread thaw upon climate warming. Fig.7f illustrates the spatial relationships between ground ice content and soil properties. Due to substantial spatial uncertainties in the IPA map, ground ice content was not incorporated into the modelling process of this study.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

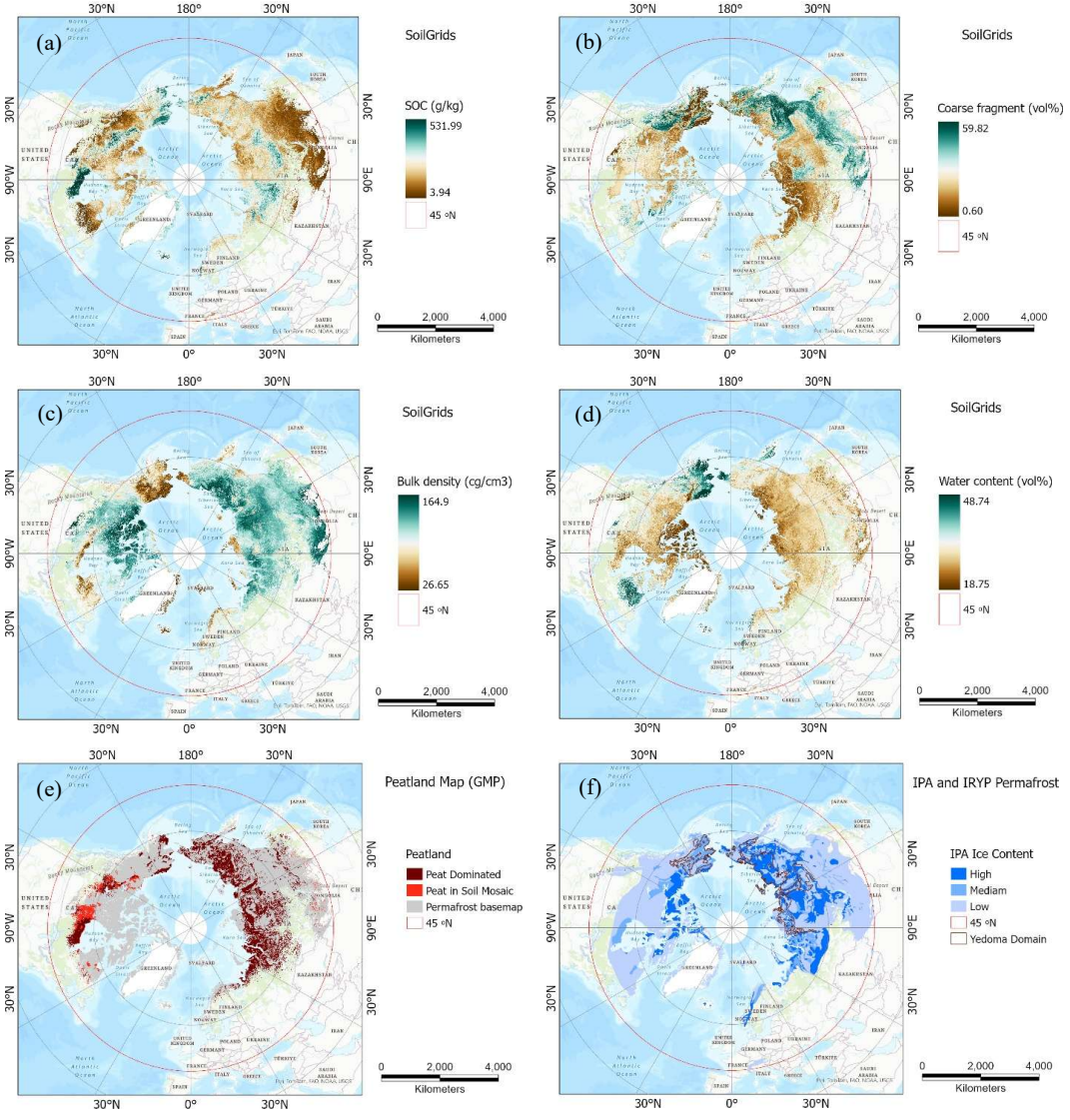


Figure 7 SoilGrids maps of soil organic content (a), coarse fragment (b), bulk density (c), and volumetric water content (d), along with the GMP peatland (e) and the IPA ice content (f).

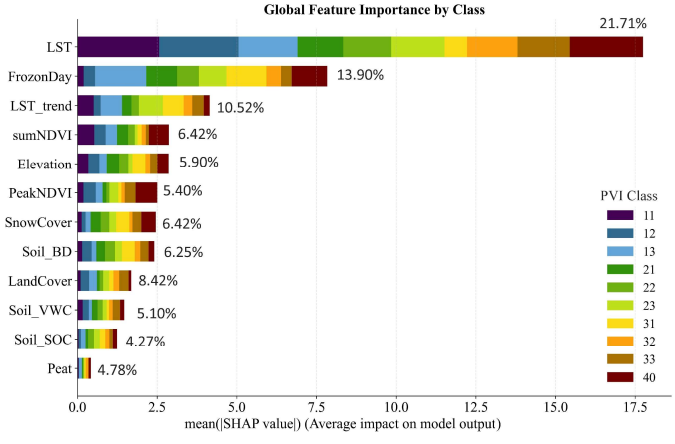
https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

3.3 Permafrost Thaw Index

3.3.1 Influencing environmental variables

The optimal feature set for the XGBoost classifier was determined to include 12 environmental variables. Four variables (slope, NDVI trend, snow trend, and soil coarse fragment) were excluded because of their low importance scores and potential collinearity with other predictors. With the RFECV feature selection procedure for XGBoost classification, the stratified 5-fold cross-validation yielded a peak overall accuracy (OA) of 57%. Although this value might appear moderate compared with binary classification benchmarks, it reflects substantial predictive capability for a complex multi-class problem involving 10 ranked PTI classes. Notably, it significantly exceeds the theoretical random baseline accuracy of 10% (over 5-fold).

The PTI model of XGBoost classification was constructed using the remaining 12 variables. The contribution of individual variables to PTI prediction was evaluated with the SHAP (Shapley Additive ExPlanations) method (Lundberg and Lee 2017). In the SHAP bar plot (Fig.8), the global importance of each variable is aggregated across all PTI classes and presented in decreasing order. The x-axis, representing the mean absolute SHAP value (mean |SHAP value|), indicates the average absolute contribution of each variable across 100 model iterations. This metric reflects the overall impact of a variable on each PTI class prediction. For comparison, the XGBoost feature importance (gain) of each variable is also marked on the bar plot. Thermal conditions of permafrost exert the dominant factor (46%) on PTI prediction, comprising contributions of 21.71% from LST, 13.90% from frozen days, and 10.52% from the LST trend. Vegetation-related variables represent the 2nd most influential group (21.15%), with roughly equal contributions from ΣNDVI, peak NDVI, and vegetation type (land cover). Snow cover accounts for 6.42% of the explained importance. Soil properties collectively contributed 20.4%, with bulk density showing slightly higher importance, followed by volumetric water content, peat presence, and soil organic content. With a contribution of 5.9%, elevation also acts as an important confounding factor in shaping permafrost characteristics.



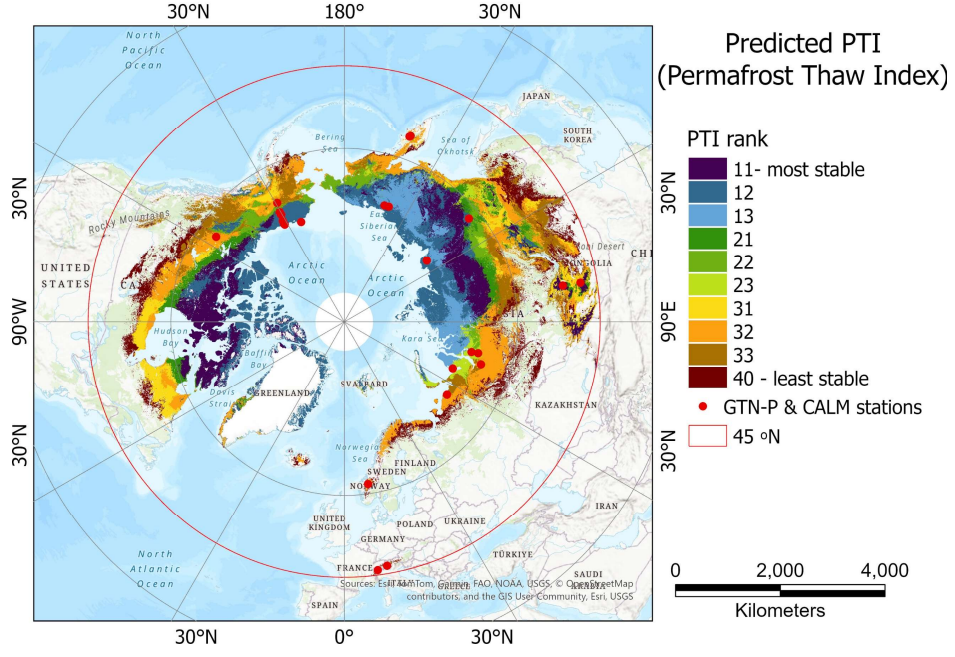
https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Figure 8 SHAP bar plot illustrating the importance of the 12 selected variables for PTI prediction, with XGBoost feature importance (gain, %) also indicated.

3.3.2 PTI map

On the PTI map (Fig.9), permafrost regions are classified in a two-tier system with 10 ranks of thaw vulnerability. A 3x3 smoothing filter was applied to reduce noises. The most stable rank (PTI = 11) represents cold permafrost without any temporal trends. Notably, these areas are not only located in the northmost High Arctic. Permafrost at Rank 40 clusters along the southmost extent that are mostly isolated patches (0-10%), indicating areas that are immediately subjected to thaw under current environmental conditions. Also along the southern margin are Rank 33, which represents warm permafrost with significantly increasing trends in both MAGT and ALT. The remaining ranks indicate intermediate thaw potential, generally increasing from the High Arctic toward the south. In the high mountains of the subarctic, alpine environments play an important role in regulating the vulnerability of permafrost thaw.

Statistically, the total permafrost coverage in the northern high latitudes (> 45 °N) is approximately 18.69 million km^2. Regarding the thaw potentials, only about 15.54% remains the most stable (Rank 11), while the three cold permafrost classes (PTI = 11, 12, 13) together account for 41.41% of permafrost regions. In contrast, Rank 40 comprises 11%, and the three warm permafrost classes (PTI = 31, 32, 33) collectively contain 34.86%. Notably, nearly half of permafrost in the stud region is experiencing high thaw pressure.



https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Open Access Earth System Science Data Discussions

Figure 9 Predicted two-tier PTI map. The first digit represents the MAGT-based permafrost types (cold, cool, warm). The 2nd digit indicates the temporal trend of MAGT and ALT. The 26 paired GTN-P/CALM stations are marked in the figure.

3.3.3 PTI map accuracy and quality assessment

The validity and quality of the predicted PTI map were evaluated using both randomly selected sampling points across the study region, and globally archived permafrost stations containing continuous borehole records since the 2000s.

The PTI map accuracy was evaluated during XGBoost model development, in which 80% of the sample dataset (listed in Table 3) was used as the training set and the remaining 20% as the validation set. Across the outputs of 100 XGBoost iterations, the coefficient of variation of the overall accuracy (OA) was only 0.24%, indicating the strong stability and robustness of the XGBoost model. After the ensemble PTI prediction was achieved, a confusion matrix (Table 4) was constructed. The size of validation points for each PTI class is listed in the table. The XGBoost classifier in this study achieved an OA of 91.79%. In addition to the OA, class-specific performance was evaluated with precision and recall metrics. Given a PTI class, precision measures the proportion of correct predictions among all predicted instances of this class (false positive), whereas recall measures the proportion of correctly predicted instances among all actual occurrences of this class (false negative). Among the 10 PTI classes, the 1st-level ranks represent cold permafrost (PTI = 11, 12, 13) that achieve the highest accuracies, followed by the 2nd-level in cool permafrost (PTI = 21, 22, 23) and the 3rd-level in warm permafrost (PTI = 31, 32, 33). As indicated by the shaded cells in Tabe 4, misclassifications primarily occur within the same-level, tier-2 ranks. For example, Rank 11 is most often misclassified as 12 or 13, and Rank 32 is misclassified as 31 or 33. The isolated patches (Rank 40), as a single class, achieve high accuracies in both precision and recall (94-95%), with only a small portion of samples misclassified as Ranks 31–33.

Overall, precision and recall values are well balanced across the PTI classes in the confusion matrix. These high accuracies suggest that the XGBoost model effectively captures the underlying relationships between the 12 environmental variables and the PTI ranks across the study region. The derived PTI map can be applied to assess the thaw potential of the northern high-latitude permafrost regions.

Table 4. Confusion matrix of the predicted PTI map

		Validation points										Precision (%)
	PTI	#11	#12	#13	#21	#22	#23	#31	#32	#33	#40	
Classified	#11	**1232**	73	5	15	2	0	1	0	0	0	92.77
	#12	57	**1683**	27	4	15	0	0	2	1	0	94.27
	#13	14	24	**805**	1	0	7	0	0	0	0	94.59
	#21	13	2	2	**691**	27	9	14	12	9	0	88.70
	#22	9	1	0	23	**854**	7	0	23	5	0	92.62
	#23	6	1	8	17	18	**795**	1	8	10	0	92.01
	#31	1	1	0	14	4	1	**953**	56	13	8	90.68
	#32	2	4	0	15	32	4	56	**1586**	101	19	87.19
	#33	0	1	0	2	12	12	12	80	**1618**	20	92.09

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

#40	0	0	0	0	0	0	14	25	25	**1041**	94.21
Recall (%)	92.35	94.02	95.04	88.36	88.59	95.21	90.68	88.50	90.80	95.68	
Overall accuracy			91.79%								

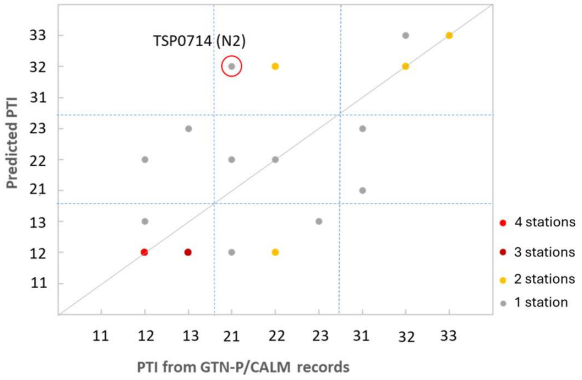
Upon the PTI ranking system shown in Table 3, the PTI at a given location can be explained by its MAGT and the trends in MAGT and ALT. The quality of the PTI map was further evaluated using established permafrost monitoring stations where long-term in situ borehole records have been archived. Currently, two major global networks are available for monitoring permafrost conditions and changes worldwide: the Global Terrestrial Network for Permafrost (GTN-P, 2018) and the Circumpolar Active Layer Monitoring (CALM) program (Nelson et al., 2021). Together, GTN-P and CALM provide complementary datasets that are essential for quality assessment of the predicted PTI map. All stations located in the northern high latitudes (> 45ºN) were examined. MAGT measurements are available at GTN-P sites, whereas ALT measurements are available at CALM sites. The two measurements were paired when the two sites were geographically close (although the collection years often vary). At each paired station, the average MAGT was calculated from the time series records to represent its general thermal condition. To assess MAGT and ALT trends, only stations with records spanning at least three years were counted. A total of 26 paired stations were identified for this study. The paired MAGT and ALT records were mostly collected during 2007-2016. Their locations are marked in the Figure 9.

The PTI ranks at these stations were extracted using the same ranking system as presented in Table 3. Their PTI ranks ranged from 11 to 33, with no station having a PTI rank of 40. These values were compared with the predicted PTI at each station (Fig.10). The PTI values at 9 stations matched the predicted PTI values exactly (lying on the 1:1 line), while those at 7 stations differed by only one rank (close to the 1:1 line). The comparisons in Fig.10 yield a ranked correlation coefficient (Spearman's r) of 0.69, which is statistically significant based on a Student's t test ($p < 0.001$). The strong goodness of fit indicates the validity of the predicted PTI map for assessing permafrost thaw conditions of permafrost in the study region.

As marked in the figure, the station with the greatest discrepancy is the paired GTN-P TSP_0714 and CALM N2 site, located in a mountainous region of Norway (61.68ºN, 8.37ºE). At an elevation of 1,861 m, it is a typical alpine permafrost. According to the ensemble environmental dataset, the station lies in barren tundra with 86% permafrost percent cover. The 2007–2016 borehole time-series records at this station were relatively complete, with MAGT missing in 2007 and ALT missing in 2012. The ALT remained stable over time, which was consistent with the satellite-based estimation (i.e., no ALT trend). The average MAGT from borehole records was -1.83ºC, yet the satellite-based ensemble MAGT was only -0.37ºC. More specifically, the 9-year borehole MAGT records showed a uniquely decreasing trend ($R^2 = 0.52$), whereas the satellite-observed MAGT trend in this study increased uniformly across the study region. Although permafrost warming has been observed globally ((Biskaborn et al., 2019), limited studies suggest that high-elevation permafrost zones may exhibit short-term cooling trends influenced by local factors such as snow and terrain (Luetschg et al., 2008). Future studies incorporating longer borehole records may help better explain the MAGT trends at this site.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

590



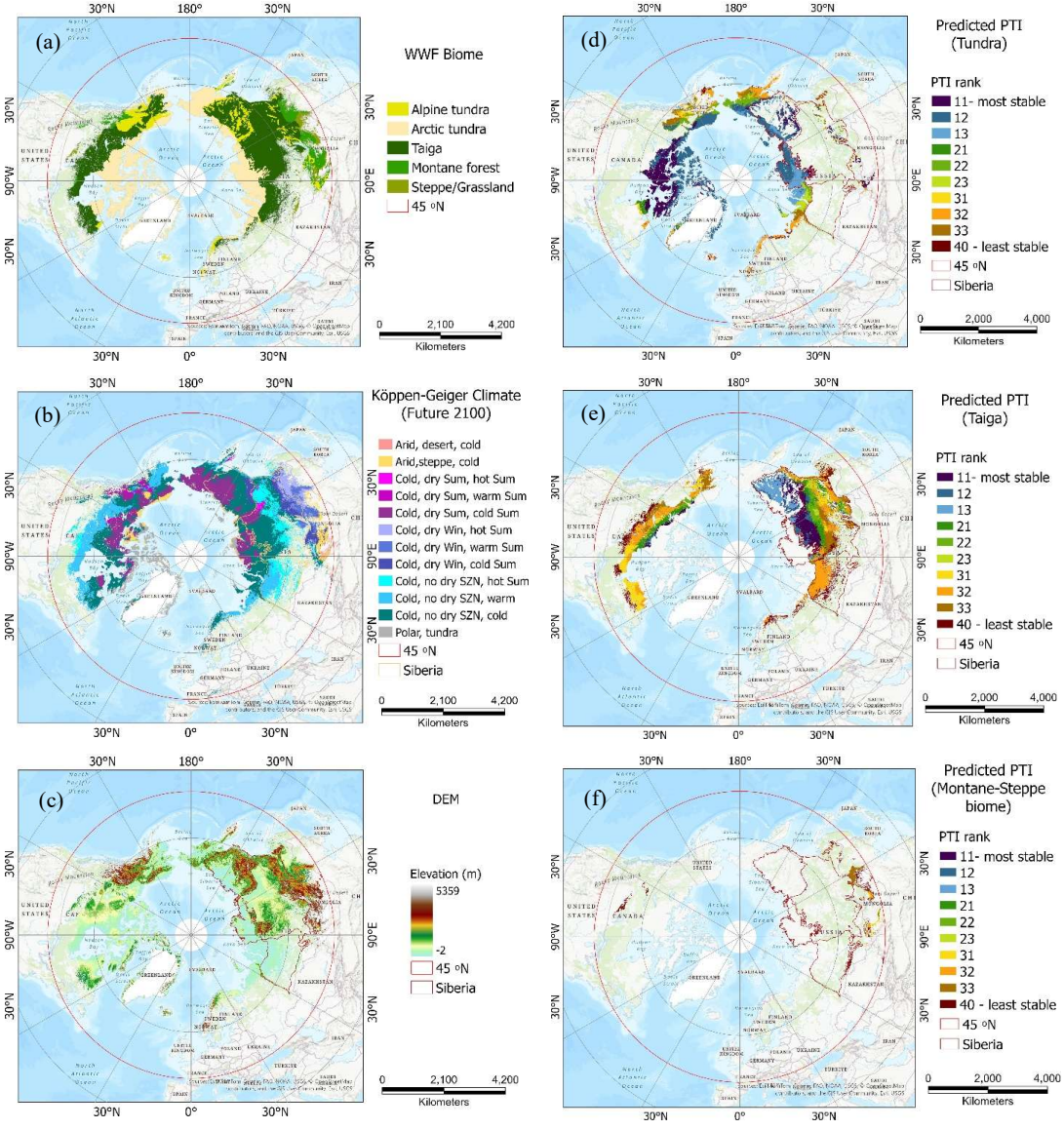
595 **Figure 10 Comparison of satellite-estimated and ground-observed PTI at GTN-P and CALM stations with MAGT and ALT records from 2007–2016. Coloured dots indicate multiple stations.**

3.3.4 Eco-climatic characteristics of permafrost thaw

The derived PTI enables exploration of latitudinal and altitudinal transitions of permafrost thaw in relation to its eco-climatic characteristics. The WWF biomes (90 ecoregions) in permafrost of the study region were grouped into tundra, Taiga, 600 montane forest, and subalpine–temperate biomes (Fig.11a). Tundra consists of Arctic tundra in the High Arctic and alpine tundra at high elevations. Taiga, or boreal forest, is a cold-climate forest dominated by coniferous trees. Located in the montane-subalpine zones, montane forest is characterized by high-attitude coniferous trees below the tree line. The southmost ecosystems in permafrost regions are a mixture of forest steppe, temperate broadleaf forests, shrublands, and grasslands, labeled as Steppe/Grassland in Fig.11a.

605 Although it is not feasible to entail all 90 WWF ecoregions, Siberia was selected to illustrate the eco-climate influences. Permafrost in North Asia is predominantly distributed in four WWF ecoregions within the vast area of Siberia, corresponding to its geographic subdivisions: West Siberian taiga, North-Central (Taimyr-Central) Siberian tundra, East Siberian taiga, and Northeast Siberian taiga. For clearer visualization of permafrost in Siberia, the four WWF ecoregions are outlined in the figure.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026



610 **Figure 11 Eco-climate characteristics in permafrost lands: WWF ecoregions (a), Köppen-Geiger climate (b), and elevation (c). Panels on the right show PTI maps across different biomes: taiga (d), tundra (e), and alpine–temperate biome (f). The four Siberian ecoregions are outlined in each panel.**

Climate zonation in this study adopts the 1-km Köppen–Geiger projected 2100 climate (Beck et al., 2018). In Fig.11b, permafrost lands fall within cold climate regions and are further sub-categorized into ten zones based on seasonal conditions

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

(dry/wet; hot/warm/cold). For example, Arctic tundra is predominantly characterized by polar tundra climates with dry and cold summers, whereas boreal taiga within the Arctic circle (>66.5 ºN) commonly experiences cold summers with or without a distinct dry season. These sub-categorical variations may exert different influences on permafrost thaw. Topography (Fig.11) plays an important role in both permafrost distribution and the regulation of ecosystem–climate zonation. The mountain ranges are clearly visible in the figure. Topographic patterns are particularly used to explain permafrost conditions in alpine tundra and montane forests.

In Fig.11d-f, the PTI distributions of three biomes are visually compared: tundra, taiga, and a mixed montane-temperate biome. Arctic tundra occupies the northmost permafrost regions. Its PTI is dominated by ranks 11, 12 and 13 in the High Arctic (Fig.11d). The most stable permafrost is in the Canadian Low Arctic tundra north of Hudson Bay. Other Arctic tundra regions in both continents are classified as coastal tundra in the WWF framework. They exhibit slightly lower stability (PTI = 12). The southern half of the Taimyr-Central Siberian tundra shows increased vulnerability to thaw (PTI = 13). The lowland Arctic tundra in the northwestern Siberia and the far northern European coast is the most vulnerable, with PTI ranging from 13 to 22 in the Yamal-Gydan tundra and reaching 32-33 in the Northwest Russa–Novaya Zemlya tundra.

Alpine tundra is a high-elevation, treeless biome. In the mountain ranges of Siberia, permafrost within alpine tundra is predominantly stable (PTI =11, 12). Despite their lower latitudes (45-50 ºN), te Altai-Sayan meadows and tundra along the high-mountain borders of Russa, China, Mongolia, and Kazakhstan exhibit highly stable permafrost conditions (PTI =11). In contrast, permafrost in other alpine tundra lands, such as the Bering tundra in northeastern Russia and the Interior Yukon-Alaska tundra spanning Alaska and northwestern Canada, faces greater thaw pressure with dominant PTI ranks of 22 and 32.

Boreal taiga (Fig.11e) is the largest biome in the northern high-latitude permafrost and is dominated by cold-climate, coniferous forests. In North Asia, permafrost in the taiga is mostly distributed across Siberia and exhibits clear geographic variation in PTI ranks. In Northwestern Siberia, permafrost is relatively stable at PTI=13 within the predominant taiga, or even higher stability (PTI = 11) within the Cherskii-Kolyma mountain tundra. In East Siberia, permafrost displays a clear latitudinal transition from the most stable (PTI = 11) in the Central Siberia Plateau to graduate shift of 2nd- and 3rd-level ranks toward the south. Strong local variability is observed in the mountain ranges northeast of Lake Baikal, where higher elevations correspond to lower thaw vulnerability within the Trans-Baikal Bald Mountain tundra. Extensive permafrost areas in other taiga ecoregions—including West Siberia, northern Europe, Alaska, and northeast Canada—generally have high, 3rd-level ranks (PTI = 31, 32, 33) of thaw. The most vulnerable rank (PTI = 40) occurs at the southernmost extent of the taiga biome.

Fig.11e displays the PTI ranks in a mixture of montane forests, steppe and grasslands, each with limited coverage at the southern extent of permafrost. Montane forests are alpine ecosystems dominated by coniferous and mixed forests, functioning as transitional zones (ecotones) between the cold boreal taiga and warmer temperate forests. Owing to their southern latitudinal locations, permafrost in this region bears the highest thaw vulnerability with the PTI ranks from 31 to 40. This pattern corresponds to the lowest permafrost percent (PP) cover in open-access permafrost maps. An exception occurs

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

in the Khangai Mountains of central Mongolia, where the most stable permafrost (PTI = 11) is observed at high elevations. In the WWF classification, this area is identified as the Khangai Mountains alpine meadow. However, our ensemble land cover map identifies it as grassland and sparse vegetation, consistent with the WWF-classified Gobi Lakes Valley desert steppe to the south. With elevation exceeding 3,000 m at the summit of the Khangai Mountains, this area may be more appropriately classified as alpine tundra where permafrost remains stable.

Table 5 summarizes the dominant PTI ranks of WWF ecoregions and their environmental influences. 18 of the 90 WWF ecoregions have a geographic extent exceeding 200,000 km^2 spanning four biomes—Arctic tundra, alpine tundra, taiga, and montane forest. Upon the SHAP bar plot in Fig.8, the five most influential environmental variables for PTI prediction are LST, frozen days, LST trend, ΣNDVI, and elevation in a decreasing order of SHAP impact. Permafrost across the Low, Middle, and High Arctic tundra of North America exhibits the highest stability (PTI = 11, 12), characterized by extremely low LST, negligible LST trends, long frozen days, and sparse vegetation cover. In contrast, Arctic tundra in North Asia experiences higher LST, significantly increasing trends (also shown in Fig.6b) and denser vegetation cover. Particularly, Bering tundra has an average LST of -6.6°C and ΣNDVI of 9.4, both substantially higher than other Arctic tundra. Its dominant PTI ranks are 32 and 22, indicating that its permafrost is under considerable pressure of thaw. Alpine tundra in the two continents situates at comparable elevations. However, the Interior Yukon-Alaska alpine tundra exhibits much higher LST (-4.0°C), fewer frozen days (168 days), and greater ΣNDVI (13.4), corresponding to dominant PTI ranks of 33 and 32. In contrast, the Cheriskii-Kolyma Mountain tundra in Northeast Siberia shows markedly colder conditions with LST of -10.8°C), longer frozen days (212 days), and lower ΣNDVI (7.1), contributing to greater permafrost stability with dominant PTI ranks of 12 and 11.

Permafrost stability in boreal taiga across the two continents reveals great geographic variability. Among all taiga ecoregions, the West Siberia taiga has the most stable permafrost, whereas the Midwestern Canadian Shield forest is the least stable. The latter is not explicitly classified as taiga in the WWF system; however, its immerse, largely intact boreal forest on rocky terrain, and the characteristic coniferous evergreen trees under a cold climate, places it within the taiga category. Geographically, it transitions southward from the Northern Canadian Shield taiga and has considerably higher temperatures (LST = -1.8°C), fewer frozen days (139 days), and higher vegetation cover (ΣNDVI = 15.8) than other taiga ecoregions. Consequently, it experiences the highest thaw (PTI = 40, 32) within the taiga biome.

Montane forests grow at high elevations and act as a critical ecotone between boreal taiga and warmer hardwood forests along steep vertical gradients. The Northern Cordillera forest in North America has an average elevation of 1,407 m, while the Trans-Baikal conifer forest averages 1210 m. Both forests have a PTI rank of 33. The Da Hinggan-Dzhagdy Mountain conifer forest in the south experiences even greater thaw pressure (PTI = 40).

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Table 5. The top two PTI ranks and key environmental controls on permafrost in major WWF ecoregions (> 200,000 km^2).

	Biome	WWF Ecoregion	Top 2 PTI ranks		Primary influencing variables*				
			1st PTI / Percent (%)	2nd PTI# / Percent (%)	Elevation (m)	LST (°C)	Frozen days	LST trend (°C/year)	ΣNDVI
North America	Arctic tundra	Low Arctic tundra	#12 / 80.7		278	-13.9	258	0.01	1.0
		Middle Arctic tundra	#11 / 68.6		181	-11.4	227	0.01	2.7
		High Arctic tundra	#11 / 59.7		220	-8.9	201	0.01	6.1
	Alpine tundra	Interior Yukon-Alaska alpine tundra	#33 / 36.4	#32 / 24.0	840	-4.0	168	0.01	13.4
	Taiga	Interior Alaska-Yukon lowland taiga	#33 / 35.4	#32 / 29.4	254	-3.2	163	0.05	14.2
		Northern Canadian Shield taiga	#21 / 34.2	#32 / 31.2	325	-5.6	168	0.00	9.8
		Eastern Canadian Shield taiga	#31 / 66.6		371	-4.6	161	0.00	10.0
		Midwestern Canadian Shield Forest	#40 / 44.6	#32 / 37.1	329	-1.8	139	0.00	15.8
	Montane forest	Northern Cordillera Forest	#33 / 86.3		1407	-3.8	168	0.02	12.8
North Asia	Arctic tundra	Taimyr-Central Siberian tundra	#12 / 54.7		236	-11.0	226	0.16	6.1
		Yamal-Gydan tundra	#13 / 44.0	#23 / 36.5	37	-7.0	200	0.18	8.0
		Bering tundra	#32 / 36.6	#22 / 33.0	326	-6.6	191	0.05	9.4
	Alpine tundra	Cherskii-Kolyma Mountain tundra	#12 / 44.4	#11 / 30.8	881	-10.8	212	0.04	7.1
	Taiga	East Siberian taiga	#11 / 21.2	#32 / 17.8	486	-6.5	180	0.09	13.2
		Northeast Siberian taiga	#13 / 48.6	#11 / 21.8	519	-9.9	199	0.05	9.2
		West Siberian taiga	#32 / 65.6		75	-3.1	158	0.12	13.7
	Montane forest	Trans-Baikal conifer forests	#33 / 53.0		1210	-1.7	144	0.00	20.0
		Da Hinggan-Dzhagdy Mountain conifer forest	#40 / 46.8	#33 / 19.0	624	-2.1	142	0.01	20.7

* All variables, except the LST trend, represent annual means derived from multi-decade satellite observations.
# The 2nd PTI is listed only when the 1st PTI accounts for less than 50% of the ecoregion.

3.4 Satellite-based open data for permafrost research

Permafrost thaw is a nonlinear and spatially heterogeneous process that should be examined through a multi-fold framework. As global temperatures rise, permafrost warms and begins to thaw. At the same time, climate change alters environmental conditions such as snow cover, vegetation, and soil moisture, exerting direct or indirect influences on permafrost dynamics. Snow functions as an insulating layer that traps heat in the ground during winter, while vegetation cover reduces surface albedo and enhances heat absorption. Local topography influences snow accumulation and solar radiation, and soil

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

properties regulate thermal conductivity. These environmental factors interact nonlinearly, collectively shaping permafrost distributions. Up-to-date information of these land surface variables are crucial to timely investigating permafrost degradation.

Satellite remote sensing enables consistent, repeated and spatially continuous observation of the Earth's surfaces. Open-access image products of key land surface variables from various satellite sensors have been available at a global scale. This study takes advantage of those multi-decadal open data product series, including land surface temperature, land cover, NDVI, snow cover, and frozen days. Many of these products integrate imagery from multiple satellites and sensors, as well as long-term ERA5 climate reanalysis data. As demonstrated in this study, these multi-source, continuous open datasets characterize the environmental drivers of permafrost thaw. Together with in situ permafrost measurements, these environmental datasets provide essential constraints and validation sources for models simulating permafrost dynamics. Moreover, these large datasets facilitate machine learning and data assimilation approaches that integrate observations with process-based models to improve predictive performance. With the expanding availability of satellite-derived open data products, we anticipate their increasing applications in monitoring permafrost degradation from regional to global scales.

Despite their global availability, satellite-derived open data products inevitably contain uncertainties due to limitations in image acquisition, processing, integration and validation. Data gaps, for example, are a persistent challenge in these products because of poor atmospheric conditions such as frequent cloud cover, low solar angles, and long winter darkness in high latitudes. Snow cover and surface moisture further complicate the retrieval of variables such as vegetation and land surface temperature. Sensor-specific noises, such as orbital drift, instrument degradation, and cross-calibration inconsistencies among successive satellite sensors, may also introduce discontinuities in long-term data series. To address these limitations, various gap-filling approaches have been implemented in the published open datasets, including temporal interpolation, spatial smoothing, and multi-source data fusion. Although these methods improve completeness of the data products, they may introduce additional uncertainty and potential bias. Consequently, modelling permafrost distribution and degradation using these multi-source datasets is inherently noisy. Taking advantage of large datasets, machine learning and deep learning methods within a broader field of artificial intelligence (AI) are capable of exploiting noisy observation and error characterization in open data, and therefore, enhances model robustness, interpretability and reliability. Adopted in the XGBoost model in this study, the ensemble learning framework spreads the training among model ensembles (i.e., random trees and boosted trees) and multi-source ensembles (i.e., environmental variables) to achieve generalized prediction. Nevertheless, for satellite-based open datasets, rigorous quality control, uncertainty assessment, and transparent documentation of data processing procedures are essential to ensure their robust application for reliable permafrost assessment.

Satellite imagery cannot directly map permafrost beneath the active layer, which can exceed 2 meters in thickness in the northern high latitudes (Hugelius et al., 2013; Westerman et al., 2024). This study confirms that thermal dynamics play a dominant role in permafrost thaw. However, land surface temperatures derived from thermal sensors do not exhibit a simple

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

linear relationship with permafrost ground temperature. Different modelling approaches have been incorporated in the three open-access permafrost maps (Obu, NIEER, and Permafrost_CCI) used in this study. Substantial discrepancies are observed among these products, particularly along the southern margins of permafrost. In this study, we integrate these three permafrost maps to generate an ensemble permafrost base map delineating permafrost distribution across the northern high latitudes. Furthermore, uncertainties persist in the estimates of the mean annual ground temperature (MAGT) and active layer thickness (ALT), which form the basis of the PTI ranking system developed in this study. Future advances in satellite observations permafrost modelling may help reduce uncertainties in MAGT and ALT estimates. Improved permafrost products are essential for more accurately documenting permafrost thaw, particularly in regions with higher PTI ranks that exhibit greater sensitivity to climate warming.

Finally, thawing permafrost has direct societal impacts across Arctic and subarctic regions such as infrastructure damage, altered water resources, and ecosystem transformation. Policymakers and local communities require timely and reliable information to support risk assessment and mitigation planning. Recent advances in satellite-based open data have highlighted the importance of shared data frameworks in fostering coordinated permafrost monitoring across geopolitical boundaries. Long-term satellite archives are invaluable for detecting trends, abrupt shifts, and emerging hotspots of permafrost degradation. Ensuring the continuity of future satellite missions and free accessibility of satellite-derived data products is essential to advance permafrost research in a warming world.

4 Data Availability

The ensemble datasets derived in this study are publicly available. The datasets include: (1) permafrost thaw index, (2) land cover, (3) permafrost percent distribution (PP), and (4) mean annual ground temperature (MAGT) across permafrost regions in the Northern High Latitudes (> 45ºN). The datasets are provided in GeoTIFF format and can be downloaded at https://doi.org/10.5281/zenodo.19148960 (Guo et al., 2026).

5 Conclusion

This study develops an ensemble-based Permafrost Thaw Index (PTI) to assess thaw potentials across the northern high latitudes (>45°N) using multi-decadal satellite-derived open datasets. Primary findings include:

1) An ensemble permafrost dataset is constructed by integrating three open-access permafrost maps (Obu, NIEER, Permafrost_CCI). The dataset includes harmonized estimates of permafrost percent (PP), mean annual ground temperature (MAGT) and active layer thickness (ALT) to represent current state of permafrost, covering 18.7 million km^2 in the northern high latitudes (> 45 ºN).

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

2) Permafrost thaw is primarily controlled by thermal conditions. Land surface temperature (LST), frozen days, and LST trends collectively account for nearly half of the predictive importance, followed by vegetation variables (ΣNDVI, peak NDVI, and land cover). Snow cover, elevation and soil properties contribute additional explanatory power. These results confirm that while ground thermal dynamics dominate permafrost thaw, the biophysical and soil-hydrological conditions exert substantial secondary influences, shaping the spatial variability of thawing.

3) The high accuracy (92%) of XGBoost classification indicates that the ensemble environmental dataset effectively captures the environmental determinants of thaw vulnerability. The resulted PTI map reveals substantial spatial heterogeneity in permafrost stability. Continuous permafrost in the High Arctic generally exhibits high stability (PTI = 11, 12, 13), covering about 41.4% of all permafrost. Nearly half of the permafrost to the south is subject to high thaw pressure (PTI = 31, 32, 33, 40).

4) Biome-level comparisons further highlight the contrast eco-climatic controls to permafrost thaw at a regional level. Arctic tundra in North-central Canada exhibits generally stable permafrost, characterized by low LST, long frozen season, and sparse vegetation. In contrast, Bering tundra in Russa and Alpine tundra in Alaska, USA experience higher thaw pressure associated with warmer LST and stronger warming trends. Within the boreal taiga, the East Siberian taiga in the Arctic region shows the most stable conditions, whereas the Midwestern Canadian Shield forest represents the most vulnerable taiga biome. Montane forests, functioning as elevational ecotones, commonly display high PTI ranks due to their southern latitude and warming sensitivity, though high-elevation alpine tundra retains stable permafrost despite its lower latitude.

This study underscores the value of satellite-based open data for permafrost research. Multi-source, long-term EO datasets enable spatially continuous assessment of environmental drivers and support ensemble modelling approaches to improve robustness and interpretability. Nevertheless, uncertainties remain in satellite retrievals, ground thermal modelling, and cross-product discrepancies. Ensuring the continuity of future satellite missions, maintaining open access and quality control of EO data products are essential for long-term permafrost monitoring.

Author Contributions

DG and CW conceptualized and designed the study. DG prepared the datasets and conducted data analysis. CZ contributed to methodological design, implemented the modelling process, and contributed to writing. SZ reviewed and improved the structure of the paper.

Competing interests

The contact authors have declared that neither they nor their co-authors have any competing interest.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

This research relies on a range of open-access datasets. The authors gratefully acknowledge the national and international programs and agencies that generate, curate, and freely provide these essential datasets, as listed in Table 1.

Financial support

This research was funded by the Key project of the National Natural Science Foundation of China (NSFC) [No. 42430412], the International (Regional) Cooperation and Exchange of the National Natural Science Foundation of China [No. W2412013], and the Basic Research Operating Funds for Provincial Undergraduate Institutions in Heilongjiang Province [No. 2025-KYYWF-ZR0126].

References

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood E.F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Nature Scientific Data, 5, 180214, https://doi.org/10.1038/sdata.2018.214, 2018.

Biskaborn, B. K., Lanckman, J. P., Lantuit, H., Elger, K., Streletskiy, D. A., Cable, W. L., and Romanovsky, V. E.: The new database of the Global Terrestrial Network for Permafrost (GTN-P). Earth Syst. Sci. Data, 7(2), 245–259, 2015.

Biskaborn, B.K., Smith, S.L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D.A., Schoeneich, P., Romanovsky, V.E., Lewkowicz, A.G., and Abramov, A.: Permafrost is warming at a global scale. Nat. Commun. 10, 1–11, 2019.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., and Zemp, M.: The concept of Essential Climate Variables in support of climate research, applications, and policy. Bulletin of the American Meteorological Society, 95(9), 1431–1443, 2014.

Brown, J., Ferrians, O., Heginbottom, J.A., and Melnikov, E.: Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.7265/skbg-kf16, 2002.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Brown, J., Ferrians, O.J. Heginbottom, Jr., J.A., and Melnikov, E.S. (eds.): Circum-Arctic map of permafrost and ground-ice conditions. Washington, DC: U.S. Geological Survey in Cooperation with the Circum-Pacific Council for Energy and Mineral Resources. Circum-Pacific Map Series CP-45, scale 1:10,000,000, 1 sheet, 1997.

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

Defourny, P., Lamarche, C., Brockmann, C., Boettcher, M., Bontemps, S., De Maet, T., Duveiller, G. L., Harper, K., Hartley A., Kirches, G., Moreau, I., Peylin, P., Ottlé, C., Radoux J., Van Bogaert, E., Ramoino, F., Albergel, C., and Arino, O.: Observed annual global land-use change from 1992 to 2020 three times more dynamic than reported by inventory-based statistics, in preparation, 2023.

European Space Agency (ESA): Climate Change Initiative (CCI) Essential Climate Variables [Dataset]. ESA CCI Open Data Portal. https://climate.esa.int/en/data/, 2024.

Fick, S.E. and Hijmans, R.J.: WorldClim2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302-4315, 2017.

Greifswald Mire Centre: Global Peatland Map 2.0. Underlying dataset of the UNEP Global Peatland Assessment - The State of the World's Peatlands: Evidence for action toward the conservation, restoration, and sustainable management of peatlands, Global Peatlands Initiative, United Nations Environment Programme (UNEP), Nairobi, 2022.

GTN-P: GTN-P global mean annual ground temperature data for permafrost near the depth of zero annual amplitude (2007-2016) [dataset]. PANGAEA, https://doi.org/10.1594/PANGAEA.884711, 2018.

Guo, J.F., Wang, C.Z., and Zang, S.Y.: Ensemble Dataset of Permafrost Thaw Conditions, Northern High Latitudes (>45°N) [data set], Zenodo, https://doi.org/10.5281/zenodo.19148960, 2026.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V.: Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46, 389–422, 2002.

Hall, D.K. and Riggs, G.A.: MODIS/Terra Snow Cover Monthly L3 Global 0.05Deg CMG, Version 61. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/MODIS/MOD10CM.061, 2021.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5, 3-13, 2013.

Jorgenson M.T., Romanovsky, V., Harden, J., Shur, Y., O'Donnell, J., Schuur, E.A.G., Kanevskiy, M., and Marchenko, S.: Resilience and vulnerability of permafrost to climate change. Canadian Journal of Forest Research, 40, 1219-1236, 2010.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Kim, Y., Kimball, J.S., Du, J., and Glassy, J.: MEaSUREs Polar Enhanced Resolution Freeze/Thaw Data Record from AMSR-E and AMSR2, Version 02 [2002 to 2020]. Boulder Colorado USA: National Snow and Ice Data Center. https://doi.org/10.5067/WM9R9LQ2SA85, 2021.

Li, B., Liang, S., Ma, H., Dong, G., Liu, Z., He, T., and Zhang, Y.: Generation of global 1 km all-weather instantaneous and daily mean land surface temperatures from MODIS data. Earth Syst. Sci. Data, 16, 3795-3819, 2024.

Liu, C., Huang, H., Liu, C., Wang, X., and Wang, S.: Comparative evaluation of vegetation greenness trends over circumpolar Arctic tundra using multi-sensors satellite datasets. International Journal of Digital Earth, *17*(1), 2334009, 2024.

Luetschg, M., Lehning, M., and Haeberli, W.: A sensitivity study of factors influencing warm/thin permafrost in the Swiss Alps. Journal of Glaciology, 54(187), 696–704, 2008.

Lundberg, S.M., and Lee, S.I.: A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (Eds.), Advances in Neural Information Processing Systems, 30, 4765–4774. Red Hook, NY: Curran Associates, Inc., 2017.

Myers-Smith, I.H., Kerby, J.T., Phoenix, G.K., Bjerke, J. W., Epstein, J.W., et al.: *Complexity revealed in the greening of the Arctic*. *Nature Climate Change*, 10(2), 106–117, 2020.

Natali, S., Watts, J.D., Rogers, B.M., Potter, S., Ludwig, S.M., et al.: Large loss of CO2 in winter observed across the northern permafrost region. Nat. Clim. Chang. 9, 852–857, 2019.

Nelson F E, Shiklomanov N I, Nyland K E.: Cool, CALM, collected: the Circumpolar Active Layer Monitoring program and network. Polar Geography, 44(3): 155-166, 2021.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H.H., et al.: Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale. Earth Sci. Rev. 183, 299–316, 2019.

Obu, J., Westermann, S., Kääb, A., and Bartsch, A.: Ground Temperature Map, 2000-2016, Northern Hemisphere Permafrost [dataset]. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA, https://doi.org/10.1594/PANGAEA.888600, 2018

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T. ., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., and Kassem, K.R: Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience, 51(11), 933–938, 2001.

Poggio, L.P., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil, 7, 217-240, 2021.

Ran, Y., Li, X., Cheng, G., Che, J., Aalto, J., Karjalainen, O., Hjort, J., Luoto, M., Jin, H., Obu, J., Hori, M., Yu, Q., and Chang, X.: New high-resolution estimates of the permafrost thermal state and hydrothermal conditions over the Northern Hemisphere. Earth System Science Data, 14, 865–884. 2022.

https://doi.org/10.5194/essd-2026-299
Preprint. Discussion started: 20 May 2026

Raynolds, M., and Walker, D.: Raster Circumpolar Arctic Vegetation Map, Mendeley Data (Version 2) [Dataset]. Mendeley Data, https://doi.org/10.17632/c4xj5rv6kv.2, 2022.

Smith, S.L., O'Neill, H.B., Isaksen, K., Noetzli, J., and Romanovsky, V.E.: The changing thermal state of permafrost. Nature Reviews Earth & Environment, 3, 10–23, 2022.

Strauss, J., Laboor, S., Schirrmeister, L., Fedorov, A. N., Fortier, D., et al.: Database of Ice-Rich Yedoma Permafrost Version 2 (IRYP v2) [dataset]. PANGAEA. https://doi.org/10.1594/PANGAEA.940078, 2022.

Trofaier, A.M., Westermann, S., and Bartsch, A.: Progress in space-borne studies of permafrost for climate science: Towards a multi-ECV approach. Remote Sensing of Environment, 203, 55–70, 2017.

U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center: USGS EROS Archive – Land Cover Products – Global Land Cover Characterization (GLCC) (archived July 11, 2018). U.S. Geological Survey, https://doi.org/10.5066/F7GB230D, 2018.

UNEP: Global Peatlands Assessment – The State of the World's Peatlands: Evidence for action toward the conservation, restoration, and sustainable management of peatlands (Main report). Global Peatlands Initiative, United Nations Environment Programme, Nairobi, 2022.

Wang, C., Guo, D., and Zang, S.: Thermal remote sensing for mapping the sub-Arctic permafrost and refining its southern limits. International Journal of Applied Earth observations and Geoinformation. 118, 103235, 2023.

Westermann, S., Barboux, C., Bartsch, A., Delaloye, R., Grosse, G., et al.: ESA Permafrost Climate Change Initiative (Permafrost_cci): Permafrost Version 4 Data Products. NERC EDS Centre for Environmental Data Analysis (CEDA), http://catalogue.ceda.ac.uk/uuid/7479606004d9465bad949671501e5f21, 2024.

Xiong, C., Ma, H., Liang, S., He, T., Zhang, Y., Zhang, G., and Xu, J.: Improved global 250 m 8-day NDVI and EVI products from 2000-2021 using the LSTM model. Scientific Data, 10, 800. https://doi.org/10.1038/s41597-023-02695-x, 2023.