

Duration of frozen days show a strong decline in the Northern Hemisphere mainly driven by autumn temperature increase

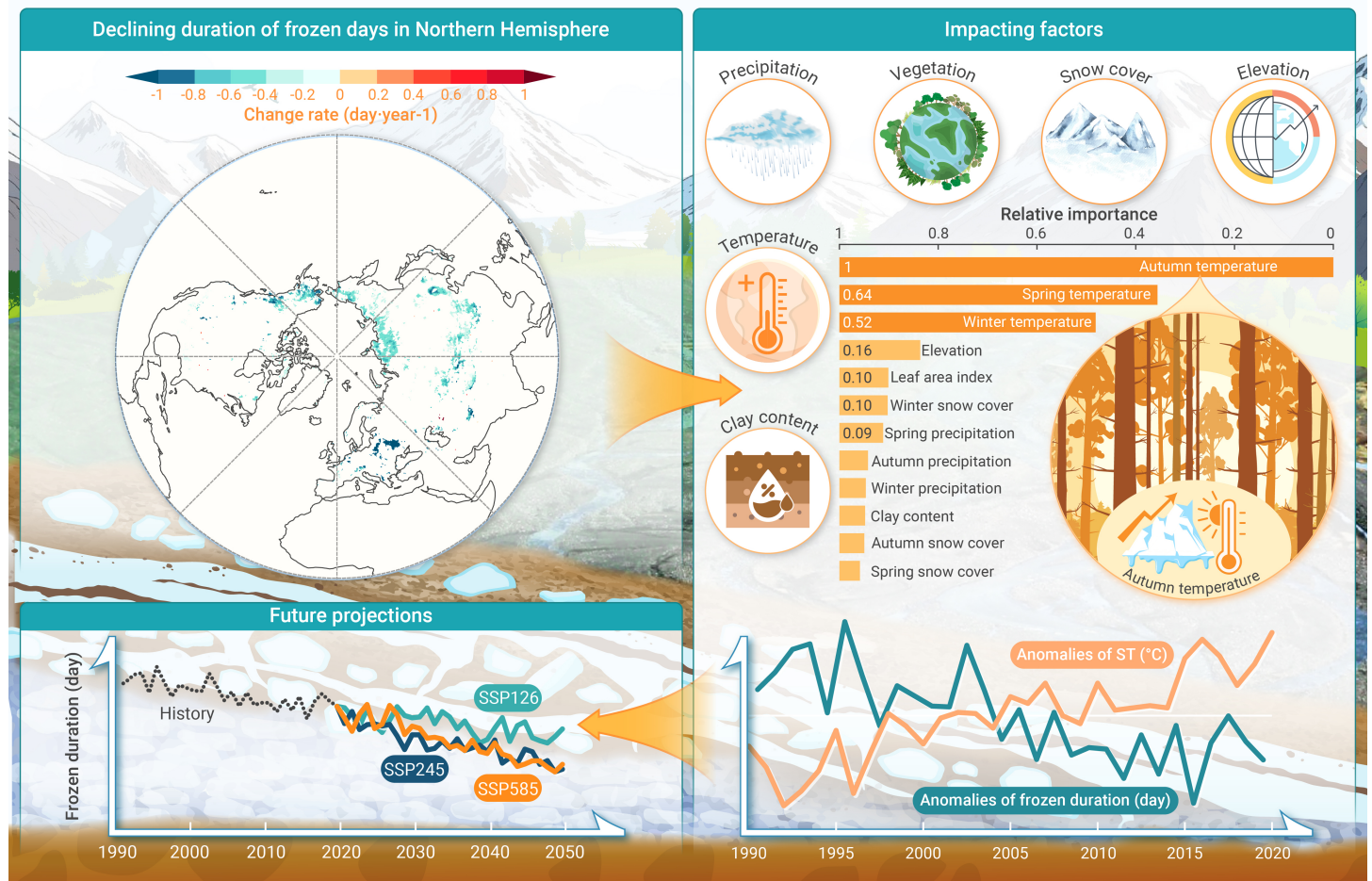
Qiangqiang Yuan,^{1,9} Wen Zhong,^{1,2,9} Qianqian Yang,³ Yanfei Peng,^{1,4} Tobias Bolch,⁴ Yuan Wang,⁵ Linwei Yue,⁶ Huanfeng Shen,⁷ and Liangpei Zhang^{8,*}

*Correspondence: zlp62@whu.edu.cn

Received: August 7, 2024; Accepted: January 15, 2025; Published Online: January 17, 2025; <https://doi.org/10.59717/j.xinn-geo.2024.100118>

© 2025 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Frozen days in the Northern Hemisphere have decreased by 0.17 days per year since 1990.
- The greatest declines were observed in Belarus, Ukraine, Alaska, and the Yukon region in Canada.
- Rising autumn temperatures, spring precipitation and vegetation changes primarily drive this decline.
- The reduction in frozen days is expected to accelerate for the highest and moderate emission scenarios.

Duration of frozen days show a strong decline in the Northern Hemisphere mainly driven by autumn temperature increase

Qiangqiang Yuan,^{1,9} Wen Zhong,^{1,2,9} Qianqian Yang,³ Yanfei Peng,^{1,4} Tobias Bolch,⁴ Yuan Wang,⁵ Linwei Yue,⁶ Huanfeng Shen,⁷ and Liangpei Zhang^{8,*}

¹School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China

²Chongqing Ecological Environment Monitoring Center, Chongqing 401147, China

³Department of Geography, Faculty of Social Sciences, Hong Kong Baptist University, Hong Kong SAR 999077, China

⁴Institute of Geodesy, Graz University of Technology, Graz 8010, Austria

⁵School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing 210044, China

⁶School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China

⁷School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China

⁸State Key Laboratory of Information Engineering, Survey Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

⁹These authors contribute equally

*Correspondence: zlp62@whu.edu.cn

Received: August 7, 2024; Accepted: January 15, 2025; Published Online: January 17, 2025; <https://doi.org/10.59717/j.xinn-geo.2024.100118>

© 2025 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: Yuan Q., Zhong W., Yang Q., et al. (2025). Duration of frozen days show a strong decline in the Northern Hemisphere mainly driven by autumn temperature increase. *The Innovation Geoscience* 3:100118.

Thawing permafrost releases methane and carbon dioxide to the atmosphere, contributing to positive feedback loop in global warming. Therefore, accurately monitoring changes in the permafrost freeze–thaw status is imperative. However, the spatiotemporal evolution and potential driving factors remain elusive. Here, we investigated the freeze–thaw status and driving factors by developing novel machine learning models trained on satellite and in situ observations in the Northern Hemisphere. We find that the frozen duration decreased on average by 0.17 days/yr since 1990 with the highest decrease of approximately up to 1.0 days/yr in parts of Belarus and Ukraine, followed by the Yukon region in Canada and Alaska. This decrease is primarily driven by temperatures in boreal autumn and spring and by precipitation and vegetation cover in boreal spring. The frozen duration is projected to decline further with reduction rates doubling until 2050 for the highest and moderate emission scenarios.

INTRODUCTION

As a key component of the cryosphere, permafrost¹ exhibits strong responses to global environmental changes, rendering it a reliable indicator and regulator of climate change.^{2,3} Widely distributed across the Northern Hemisphere, permafrost subjects vast regions to annual freeze–thaw cycles.⁴ Surface freeze–thaw cycles, an important attribute of permafrost regions, exert indispensable influences on water–heat exchanges, vegetation growth, and animal survival.^{5,6} With an exceptionally rich carbon reservoir, permafrost influences the carbon separation between Earth's surface and the atmosphere by altering soil properties during freeze–thaw cycles.^{7–9} Presently, the escalating environmental concerns arising from global climate change are increasingly evident, as glaciers recede and permafrost degrades.^{10,11} The Sixth Assessment Report by the Intergovernmental Panel on Climate Change highlights that the active layer in permafrost continually thickens primarily due to human-induced climate change.¹² Accelerated release of soil carbon during permafrost degradation contributes to a positive feedback loop in global warming.¹³

Accurate acquisition of annual freeze–thaw cycle variations proves advantageous in refining numerical models for simulating the thermal and hydrological dynamics of permafrost.¹⁴ Reliable monitoring of annual freeze–thaw cycles also aids in classifying the climatic sensitivity of permafrost, facilitating the implementation of refined protection strategies for permafrost with differing sensitivities.¹⁵ Surface freeze–thaw cycle can be estimated from in situ measurements and satellite remote sensing data. In situ measurements currently represent the most accurate approach for monitoring the frozen state of the surface and hence the surface freeze–thaw cycles. However, given their point-based nature, in situ instruments struggle to capture surface freeze–thaw variations across extensive regions. Moreover, difficulties in deploying these stations in climatically extreme and topographically complex cold regions results in relatively limited information on the frozen state and freeze–thaw cycles in these frigid zones. By contrast, microwave remote

sensing facilitates area-wide measurements and stands as the most effective means in monitoring the frozen state of the surface, owing to its strong penetrative capabilities, all-weather functionality, high temporal resolution, and wide coverage.^{16,17} However, microwave remote sensing is susceptible to interference from snow cover and vegetation. The synergy of accurate ground data at point level and satellite observations at surface level shows promising perspectives for monitoring variables in earth science at a large scale. Successful applications have been made in multiple fields. For instance, satellite-derived aerosol data have been integrated with ground-level particulate matter concentration measurements to achieve high-accuracy, spatially continuous monitoring of air pollution.¹⁸ Similarly, leveraging machine learning models to fuse active and passive microwave satellite data with in situ ground measurements, enhanced accuracy and temporal resolution in global surface soil moisture mapping was achieved.¹⁹ Comparable methods have also been employed for estimating sea surface temperatures²⁰ and snow-related parameters.²¹

Surface freeze–thaw cycles are influenced by a multitude of environmental variables. Temperature exerts the most direct influence on freeze–thaw cycles, while water infiltration also introduces heat into the soil, delaying the onset of soil freezing.^{22,23} Vegetation coverage and height can also influence freeze–thaw through the insulation effect of vegetation canopy.²⁴ Although snow cover also has insulating properties, its reflectivity considerably weakens the impact of solar radiation, resulting in a dual influence on soil freeze–thaw dynamics. Concerning soil properties, soil texture and organic matter content affect the stability post soil freezing.²⁵ Although numerous studies explored the influencing factors of freeze–thaw cycle changes, studies considering the combined effects of multiple factors are limited. In addition, amidst the backdrop of global warming, the future trends of freeze–thaw cycle with its driving factors remain uncertain. Investigating the influence of these environmental variables is pivotal for an enhanced understanding of the mechanisms driving surface freeze–thaw cycles and further exploration of their response to global environmental changes. However, the comprehensive investigations on these factors have not been fully conducted, limited by complexity of multiple data and inner interactions of these factors in conventional statistical modeling.

Nowadays, machine learning methods have been extensively employed in remote sensing research across various domains due to their robust learning and prediction capability. Leveraging machine learning to extract concealed information within remote sensing data enables more efficient modeling of the relationship between surface parameters and remote sensing data.²⁶ Ensemble learning algorithms achieve great robustness by constructing frameworks that combine multiple learners, surpassing the performance of individual learners.²⁷ These algorithms have demonstrated promising results in the retrieval of parameters, such as snow depth,²⁸ soil moisture,²⁹ thickness of the permafrost active layer,³⁰ and soil organic carbon in permafrost³¹ from satellite remote sensing data. Additionally, ensemble learning exhibits high flexibility and computational efficiency in integrating heterogeneous data

from multiple sources.^{32,33} Furthermore, ensemble learning, without relying on empirical assumptions about biophysical mechanisms, can unveil complex relationships between physical quantities.³⁴ Presently, it has been applied to investigate the impact of climate-driven factors on net primary productivity³⁵ and surface water coverage.³⁶ Hence, the application of ensemble learning to multisource data not only allows integrating satellite and in situ surface freeze–thaw cycles monitoring data to enhance the monitoring accuracy but also allows for a more comprehensive exploration of the climatic driving factors behind these cycles.

By combining a stacking model and in situ measurements, this study corrected the daily freeze–thaw status products across the Northern Hemisphere from the Making Earth System Data Records for Use in Research Environments (MeASURES) project. On this basis, the long-term trends for the period 1990–2020 and spatial variations of Northern Hemisphere annual freeze–thaw cycles were analyzed. Moreover, the driving factors behind spatiotemporal changes were analyzed with an explainable machine learning model, i.e., XGBoost-SHAP model. ST, precipitation, vegetation, snow cover, soil texture, and other environmental variables were utilized as inputs with frozen duration as the output, and an XGBoost model was employed to fit the relationship between the frozen duration and environmental variables. Subsequently, Shapley Additive exPlanations (SHAP) values³⁷ were used to assess the marginal contributions of each environmental variable and analyze their impact on the frozen duration. Finally, future projections of the annual freeze–thaw cycles under different emission scenarios were modeled using climate model data from 2020 to 2050. Our primary contributions include the following: (1) enhancing the precision of current freeze–thaw cycle products by integrating satellite and in situ observations, (2) discerning the response of freeze–thaw cycles to environment and climate changes with interpretable machine learning, and (3) revealing prospective variations in future freeze–thaw cycles under diverse climate change scenarios.

MATERIALS AND METHODS

Data collection and integration

Microwave remote sensing data. The microwave remote sensing freeze–thaw state product employed in this study was sourced from the MeASURES project³⁸ initiated by National Aeronautics and Space Administration (NASA). The primary objective of this project is to establish a consistent time series by integrating diverse data sources, thereby furnishing a foundational dataset for the analysis of prolonged trends and comprehension of Earth system dynamics. For the analysis of long-term variations in freeze–thaw cycles across the Northern Hemisphere, this study mainly leveraged the MeASURES global daily surface freeze–thaw state records. Spanning the period from 1979 to 2020, these records are derived from 37 GHz brightness temperature data acquired by a scanning multichannel microwave radiometer, a special sensor microwave/imager, and a special sensor microwave imager/sounder, employing an enhanced seasonal threshold method³⁸ to augment adaptability. Additionally, ERA5 ground-level air temperature data are incorporated to formulate climate indices,³⁹ delineating the operative spatial extent of daily freeze–thaw state records and excluding regions unaffected by freezing temperature constraints. The data are projected utilizing the EASE-Grid methodology with a spatial resolution of 25 km. The temporal scope of the MeASURES freeze–thaw products under consideration in this study encompasses the period from August 1990 to August 2020, ascertained through the computation of annual freeze–thaw transition times from the extensive temporal series of daily freeze–thaw state products.

In situ temperature measurements. The daily ground air temperature data were provided by the global historical climatology network and can be obtained from the National Centers for Environmental Information under the National Oceanic and Atmospheric Administration (NOAA). It offers observational data from 1832 to the present and is commonly used for evaluating the accuracy of global freeze–thaw cycle monitoring.⁴⁰ The study utilizes daily average ground temperature data from August 1990 to August 2020. When the temperature is below 0°C, the station is classified as in a frozen state; otherwise, it is considered in a thawed state. On the basis of the statistical analysis of Northern Hemisphere station data, the period from August to December each year is designated as the freezing transition period, while the

period from January to July the following year is designated as the thawing transition period. A five-day moving window is employed to mitigate noise interference. The first day within a consecutive five-day window with temperatures below 0°C during the freezing transition period is identified as the first freeze day, i.e., start date of freeze (SOF), and similarly, the first day within a consecutive five-day window with temperatures above 0°C during the thawing transition period is identified as the first thaw day, i.e., start date of thaw (SOT). The temperature data from multiple ground stations within each grid cell are averaged before determining the freeze–thaw status and obtaining the freeze–thaw transition time to align with microwave remote sensing data. Finally, the surface freeze–thaw cycle monitoring results obtained from ground stations are used to construct and evaluate the accuracy of the surface freeze–thaw states estimation model.

Digital elevation model (DEM) data. Considering the impact of terrain on the accuracy of surface freeze–thaw state monitoring, this study utilizes DEM data as auxiliary parameters for the machine learning model. The ETOPO1 global topographic model⁴¹ supplied by the National Geophysical Data Center under NOAA is utilized. By integrating various datasets from NASA, the Scientific Committee on Antarctic Research, the Mediterranean Science Commission, etc., this model offers high-resolution terrain and bathymetric data at a spatial resolution of 1 arc-minute. In this study, the topographic model data are resampled to match the resolution of microwave remote sensing data.

CMIP6 climate model data. The World Climate Research Program initiated the international Coupled Model Intercomparison Project (CMIP)⁴² in 1995, aiming to advance climate science through global collaboration among research institutions. CMIP6,⁴³ the latest iteration launched in 2014, boasts the most comprehensive experimental design and the highest number of participating models. Monthly near-surface air temperature, precipitation, leaf area index (LAI), and snowfall data from CMIP6 were utilized in this study. The monthly data are aggregated into seasonal scale to form explanatory variables for the response model. Seasonal averages for near-surface air temperature, precipitation, snowfall in autumn/winter/spring, and the average LAI during the growing season are obtained. The median of five climate models is used as the experimental input to enhance the reliability of input data and leverage the diversity of CMIP6 models (Table S2).

CMIP6 categorizes future scenarios based on different anthropogenic radiative forcing, known as shared socioeconomic pathway (SSP).⁴⁴ In this study, SSP126, SSP245, and SSP585 were chosen to represent low, medium, and high carbon emission scenarios, respectively, for future predictions of surface freeze–thaw cycles. Future predictions of the frozen duration were conducted under these three carbon emission scenarios to analyze the impact of human socioeconomic activities on freeze–thaw cycles. The data from these five climate models were resampled to the same resolution to achieve consistency among multiple climate models and match with MeASURES freeze–thaw state products.

Additional auxiliary data. Various auxiliary data, including land cover type, climate zone, multiyear permafrost range, and soil properties, were also collected. Given the interference of water bodies in microwave remote sensing for freeze–thaw cycle monitoring, MODIS land cover product (MCD12C1)⁴⁵ is employed to exclude water areas from the effective monitoring range. This product classifies land cover types into 17 categories according to the International Geosphere–Biosphere Program definition, with a spatial resolution of 0.5°.

The average monthly temperature in regions controlled by ice cap climate remains below 0°C throughout the year; thus, Köppen–Geiger global climate zone data⁴⁶ are used to identify these regions, exempting these areas from surface freeze–thaw cycle monitoring. These regions are mainly found in Greenland, the Antarctic continent, and some small islands in the Arctic Ocean. Köppen–Geiger climate zone data classify climates into five primary classes and 30 secondary classes based on monthly average temperature and precipitation, with a spatial resolution of 0.5°.

For a detailed analysis of surface freeze–thaw cycle variations, PERZONES North Hemisphere multiyear permafrost range data⁴⁷ are utilized. Derived from a temperature balance model with satellite ST, climate reanalysis data, and land cover as input, these data define the extent of permafrost. Using the Arctic polar stereographic projection, the data have a spatial resolution of 1

km². Areas with a permafrost probability above 90% are categorized as continuous permafrost zones, while other areas within the permafrost range are classified as discontinuous permafrost zones. Continuous permafrost zones exhibit large, continuous permafrost areas, while discontinuous permafrost zones showcase scattered, island-like permafrost distribution.

This study employs the SoilGrids250m product⁴⁸ to provide soil properties information. Based on machine learning algorithms, this product provides global 250 m high-resolution soil information using soil cover data, topography, climate data, lithology data, and other relevant variables. The data can be obtained from the International Soil Reference and Information Centre. In this study, clay content at a depth of 5 cm is used to represent soil texture information.

SOF and SOT products calibration and trend analysis

MeASURES freeze–thaw cycles products were corrected using in situ observations. First, the remote sensing data is spatiotemporally matched with site observations based on latitude, longitude, and time for the construction of the stacking model. For each specific date and station, we retrieved the corresponding pixel values based on the station's latitude and longitude at the same period. In consideration of the influence of geographical location and topography on surface freeze–thaw cycle monitoring, latitude, longitude, and elevation are used as auxiliary parameters. The feature variables of the training samples include the SOF/SOT, latitude, longitude, and elevation from MeASURES products, with the target variable being the SOF/SOT determined from site observation data. The base learners of the stacking model consist of three models: random forest (RF)⁴⁹, XGBoost⁵⁰, and LightGBM⁵¹. RF belongs to the bagging algorithm, while XGBoost and LightGBM belong to the boosting algorithm. All three algorithms achieve high accuracy, and the combination of different types of base learners provides rich information. The metalearner used to combine these three base learners in the second layer is a Lasso regression model (Figure S1), which addresses overfitting issues in standard linear regression by introducing L1 regularization. The two-layer structure of stacking further improves the accuracy of freeze–thaw cycle monitoring.

Ten-fold cross-validation is employed to evaluate model performance in this study. The evaluation metrics include root mean square error (RMSE), mean absolute error (MAE), R, and A. The accuracy of the stacking-based correction method is compared with the accuracy of the original product before correction to validate further the effectiveness of the proposed method. In addition, various ensemble learning models are established using the same construction form for comparison. The ensemble learning models involved in the comparison include the three base models used in this Stacking model construction: RF, XGBoost, and LightGBM. The model comparison results can be found in Text S1 and Table S1.

The daily freeze–thaw state sequences for each grid point in the Northern Hemisphere were inverted using the proposed stacking model, providing spatial distributions of the SOF, SOT, and frozen duration from 1990 to 2020. Subsequently, the Theil–Sen Median method was employed for trend analysis of freeze–thaw cycle variations (Text S2).

XGBoost-SHAP model and driving factors analysis

Building upon the accurate spatiotemporal freeze–thaw cycle information obtained by the stacking model, we further employ the XGBoost-SHAP algorithm to investigate the driving factors of surface freeze–thaw cycles. Initially, relevant environmental variables associated with freeze–thaw cycles are identified. The XGBoost model is then utilized to model the relationship between the frozen duration and various environmental variables, creating a response model for the Northern Hemisphere's surface freeze–thaw cycle to environmental changes. Subsequently, the SHAP model interpreter³⁷ is applied to analyze the impact of each environmental variable on the frozen duration. Finally, using the environmental response model fitted by XGBoost, future changes in freeze–thaw cycles from 2020 to 2050 are predicted on the basis of climate model data under three socioeconomic scenarios: SSP126, SSP245, and SSP585. Specifically, three steps are conducted (Figure S1).

First, environmental variables selection. The considered environmental variables include temperature, precipitation, snowfall, and LAI from CMIP6 climate model, altitude from the ETOPO1 global topographic model, and clay

content from SoilGrids250m. To analyze the environmental response differences between the transition period in spring/autumn and the stable frozen duration in winter, the study considers temperature, precipitation, and snowfall data for autumn, winter, and spring separately in the model. The final input variables include seasonal temperature, precipitation, and snowfall, along with growing season LAI, altitude, and clay content.

Second, XGBoost-SHAP-based environmental response model construction. To obtain a universal pattern of freeze–thaw cycle changes, the study uses the frozen durations for each grid point in the Northern Hemisphere from 1990 to 2020, which were obtained from the stacking model as the target variable. The chosen environmental variables serve as explanatory variables in constructing the environmental response model. In consideration of accuracy and interpretability, the XGBoost algorithm is used to fit the response model of the frozen duration to various environmental variables. Ten-fold cross-validation is employed to assess model performance using metrics such as RMSE, MAE, and R for precision evaluation. To validate further the effectiveness of the chosen algorithm, various models are built for comparison, including the traditional multiple linear regression algorithm and two ensemble learning algorithms: RF and LightGBM. The comparison results can be found in Figure S2. SHAP values are used to analyze the positive and negative impacts of each environmental variable on frozen duration changes and the differences in impact caused by different feature values of the same variable. By calculating the average absolute value of SHAP for each environmental variable across all samples, the study determines the overall importance of each variable. The relative importance is obtained by dividing the importance of each environmental variable by the importance of the most important variable, providing a visual representation of the differences in importance between different variables. A bee swarm plot is used for visual analysis to illustrate the degree and direction of the impact of environmental variables simultaneously. By contrast, the study models each grid point's data separately using XGBoost and conducts SHAP value analysis to analyze the most important driving factors for the interannual variation of frozen durations in different spatial locations.

Third, assuming that the altitude and soil properties in the Northern Hemisphere remain unchanged during the study period, the study uses climate model data from CMIP6 to predict changes in temperature, precipitation, snowfall, and vegetation under different development scenarios for 2020–2050. In consideration of the socioeconomic scenarios of SSP126, SSP245, and SSP585, representing low, medium, and high carbon emission scenarios, the environmental response model based on XGBoost is applied to predict possible frozen duration changes. Then, Theil–Sen Median analysis is employed to investigate trends in future freeze–thaw cycle changes.

RESULTS AND DISCUSSIONS

Accuracy of corrected freeze–thaw cycle products

Figure 1 illustrates the spatial distribution of the accuracy of the corrected SOF and SOT products, highlighting a comparison with the precision of the original product. The original product exhibits relatively lower accuracy in Northwestern China, Western Russia, Northern Europe, Northwestern United States, with an overall accuracy (Supplementary material) below 90%. By contrast, our product corrected by in situ measurements and the stacking model presents an enhanced accuracy of freeze–thaw cycle monitoring across the Northern Hemisphere. Specifically, the stacking-based correcting model achieves a correlation coefficient (R) of 0.96 and MAE of 5.88 days in SOF and an R of 0.95 and MAE of 9.55 days in SOT (Text S1 and Table S1). The average overall accuracy (A) improved from 90.63% to 93.77%.

Spatiotemporal pattern of SOF and SOT

Figures 2A–B displays the spatial distribution of the average freeze–thaw transition times from 1990 to 2020, as inferred by the stacking model. The figure indicates a strong latitudinal correlation between the SOF and the SOT. In the Northern Hemisphere, the SOF typically occurs in October–November each year, while the SOT generally occurs from February to April. Regions at high latitudes, such as northern Russia, northern Canada, and northern Alaska, experience an earlier SOF and a later SOT, with most entering a frozen state around September–October and thawing starting in April–May of the following year. Early SOF and late SOT are also observed in some high-alti-

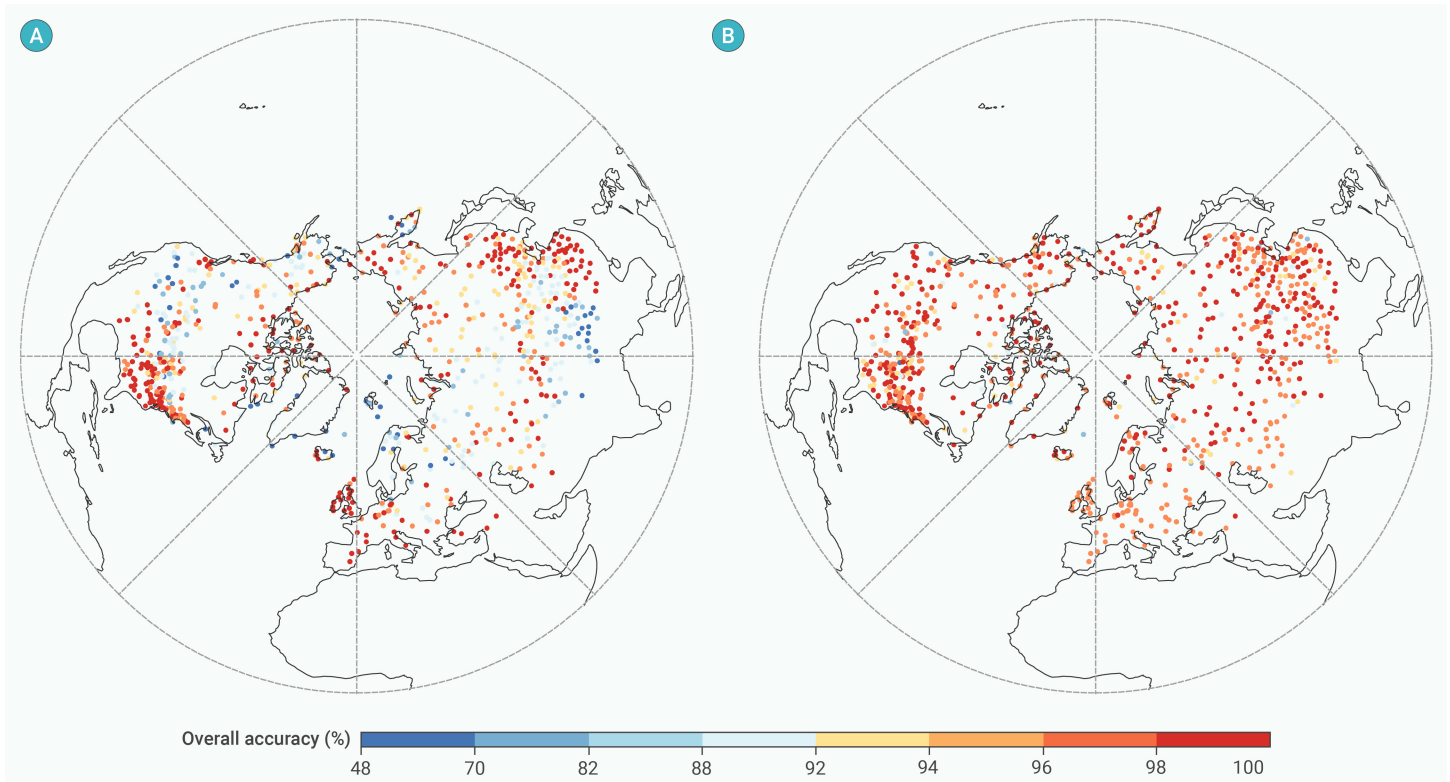


Figure 1. Spatial distribution of accuracy for (A) the original MeASURES product and (B) the product corrected by the stacking model.

tude zones in mid-to-low latitudes, such as Qinghai–Tibet Plateau, Mongolian Plateau, and Rocky Mountains.

The frozen duration for each winter was defined as the time interval between SOF and SOT, and the average frozen duration for 1990–2020 is presented in Figure 2C. Trend analysis with the Theil–Sen median algorithm shows that the average frozen duration in the Northern Hemisphere over the past three decades is generally 145 ± 77 days. Regions at higher latitudes and elevations tend to have longer frozen durations, while areas below 30° N latitude typically experience shorter freeze durations (Figure S3). Figure 2D illustrates the trend analysis results of the frozen duration change rate from 1990 to 2020, with only regions exhibiting significant (p -value < 0.05) changes shown. The average reduction rate of the frozen duration in the Northern Hemisphere is 0.17 days per year, resulting from a delayed SOF and an earlier SOT. The trend of the delayed SOF is more pronounced, with an average delay rate of 0.10 days per year. Regions with significant changes generally show a trend of reduction. Among them, the most significant reduction is observed in the eastern parts of Belarus and Ukraine, with reduction rates mostly ranging from 0.8 days to 1.0 days per year. Following are the Yukon region in Canada and Alaska in the United States, with reduction rates mostly ranging from 0.6 days to 0.8 days per year. Regions in northern Russia, northern China, and eastern Kazakhstan show reduction rates mostly ranging from 0.4 days to 0.6 days per year. Overall, Asia exhibits the largest region of significant reductions in frozen durations in the Northern Hemisphere.

Figure 3A displays the interannual variation of the frozen duration in the Northern Hemisphere from 1990 to 2020 in comparison with the annual average ST provided by NOAA. A close negative correlation ($r = -0.77$) between the frozen duration and annual average ST can be observed. Over the study period, the annual average ST exhibits a gradual upward trend with intermittent fluctuations, coinciding with a diminishing trend in frozen duration. The winter of 1995–1996 recorded the lengthiest frozen duration, whereas the winter of 2015–2016 experienced the shortest period of freezing. Prior to 2005, frozen durations predominantly exceeded the average, while post-2005, they generally fell below the average, aligning consistently with the pattern of changes in annual average ST.

Frozen duration changes among different permafrost types, including

continuous permafrost zones, discontinuous permafrost zones, and nonpermafrost zones (seasonally frozen ground), were further investigated, as illustrated in Figure 3B. The reduction rate in frozen duration for continuous permafrost zones in the Eurasian and North American continents is relatively similar, while significant differences exist between Eurasian and North American continents for discontinuous permafrost zones and nonpermafrost zones. For the Eurasian continent, continuous permafrost zones exhibit the most pronounced changes in frozen duration, while discontinuous permafrost zones show relatively smaller variations. For the North American continent, discontinuous permafrost zones exhibit the most significant changes in frozen duration, while nonpermafrost zones show the smallest changes. Overall, continuous permafrost zones experience the fastest reduction in frozen duration, with a rate of 0.21 days per year.

Driving factors of freeze–thaw dates variations

We constructed a XGBoost–SHAP-based environmental response model to identify environmental variables that dominate the change in frozen duration. Environmental variables including temperature, precipitation, snow accumulation, LAI, altitude, and clay content were used as explanatory variables, while frozen duration was used as target variable. Evaluation results showed that the constructed response model can explain 97% of the changes of frozen duration.

Figure 4A shows the importance of each environmental variable in the prediction of duration. Overall, autumn temperatures notably influence the frozen duration by predominantly affecting the timing of the initial freeze, consequently dictating the freezing period's duration. Snow cover considerably affects freezing and thawing.¹⁵ Snow exhibits a high albedo, reflecting substantial solar radiation and diminishing its impact on ST. Simultaneously, snow acts as a thermal insulator, hindering heat exchange between the ground and atmosphere, thereby providing insulation for the surface. Our analysis shows that winter snowfall considerably impacts the frozen duration, likely due to its relatively consistent volume, leading to a more stable effect on the freezing period's duration. Conversely, variations in snowfall during spring and autumn yield more pronounced fluctuations, resulting in uncertain implications for the freezing period's length. Among precipitation factors, spring rainfall exerts the most prominent influence on the frozen

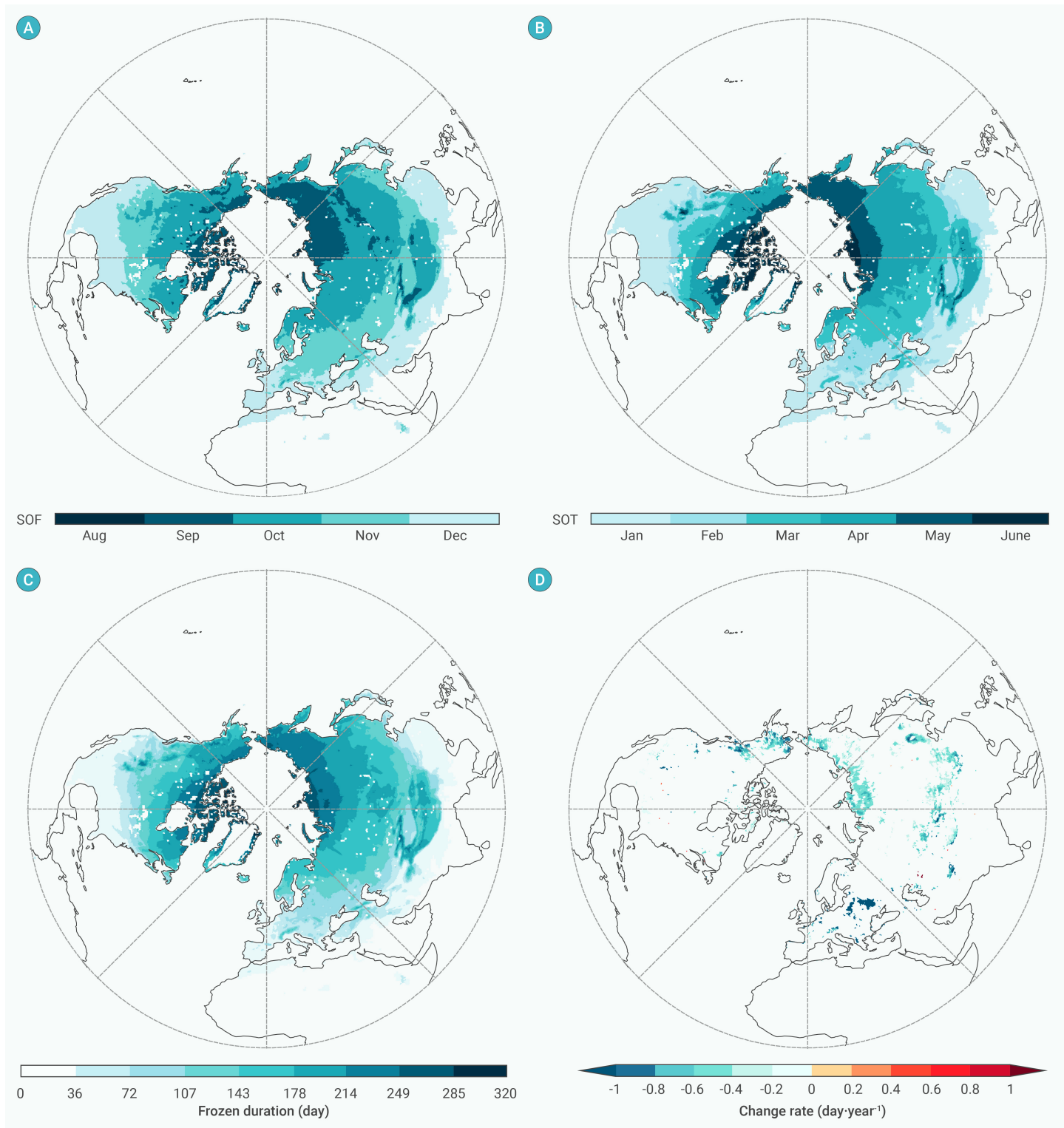


Figure 2. Spatial distribution and change rate of freeze-thaw status (A) Spatial distribution of SOF. (B) Spatial distribution of SOT. (C) Spatial distribution of average frozen duration per year of 1990–2020. (D) Change rate of frozen duration during 1990–2020.

duration. As temperatures gradually rise in spring, the relatively cool soil temperatures combined with the infiltration of spring rain introduce atmospheric heat into the soil, expediting its thawing process. This effect primarily influences the timing of the initial thaw, subsequently shaping the freezing period's length. A similar phenomenon was also reported in previous research, which found that increased rainfall can enhance soil moisture, leading to increased thermal conductivity and more heat transfer into the ground, potentially warming the permafrost.¹⁵ From the perspective of spatial differences, regions primarily affected by temperature encompass the largest area,

constituting 37% of the total. Following this are regions predominantly influenced by precipitation, encompassing 33% of the total area. The extent of areas primarily affected by snow coverage exceeds those primarily influenced by vegetation coverage. When considering seasonal distinctions, regions mainly affected by autumn temperatures constitute the largest portion at 14% of the total, highlighting autumn temperatures as the most crucial influencing factor for interannual variations in the frozen duration across different geographical locations.

The swarm plots illustrated in Figure 4B depict the effects of various envi-

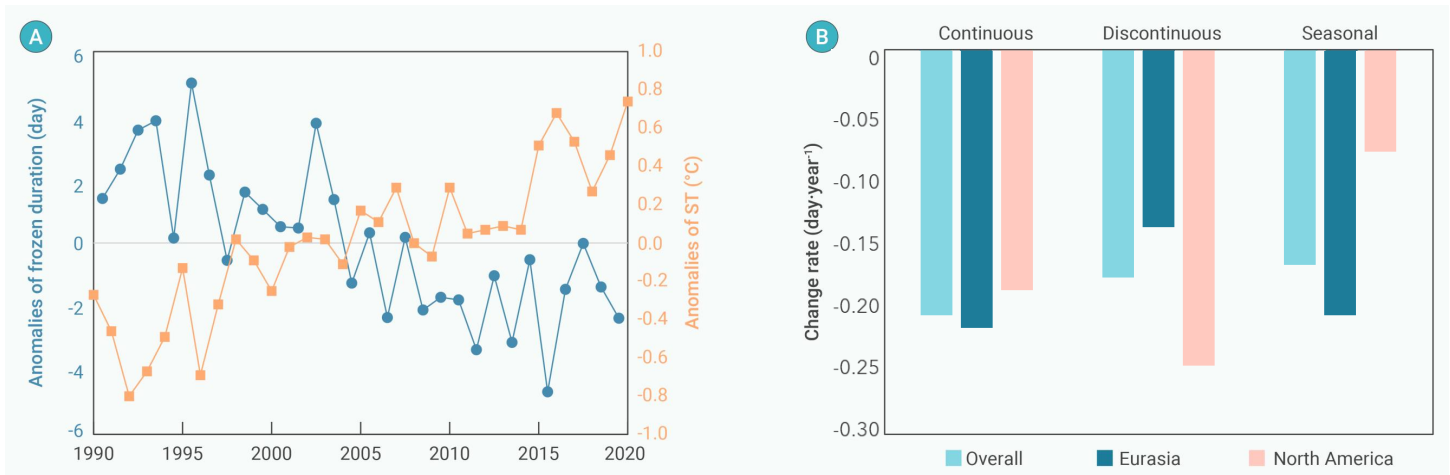


Figure 3. Impact of ST and permafrost type (A) Relationship between frozen duration and ST. (B) Average change rate in frozen duration for different types of permafrost. Continuous, discontinuous, and seasonal represent continuous permafrost zones, discontinuous permafrost zones, and nonpermafrost zones (seasonally frozen ground), respectively.

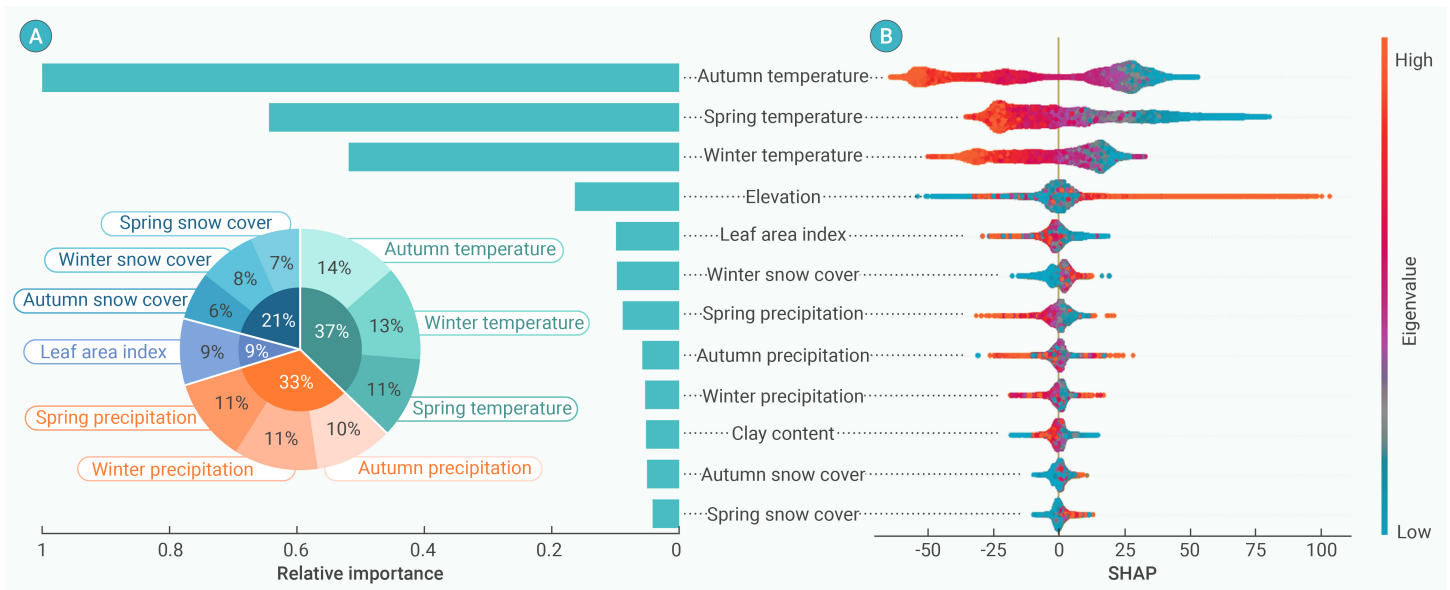


Figure 4. Variable importance in the environmental response model (A) Importance of each environmental variable in the environmental response model. The pie chart in the bottom left corner illustrates the statistical analysis of the most significant influencing factors across different geographical locations. (B) The effects of various environmental variables at different feature values on the freezing period.

ronmental variables at different feature values on the frozen duration. As expected, high temperature values often exhibit a negative impact on the frozen duration, whereas low temperature values tend to have a positive effect. The high-value range of autumn temperatures manifests the most pronounced negative impact on the frozen duration, whereas the low-value range of spring temperatures exhibits the most distinct positive impact on it. However, several other environmental variables show considerable uncertainty in their impact on the frozen duration. In such cases, employing ensemble learning models proves more accurate than multiple linear regression when exploring the effects of these environmental variables on the frozen duration. Our analysis shows that notably positive impacts on the frozen duration often belong to high elevation ranges. Additionally, samples with lower LAI values tend to display a significantly positive impact on the frozen duration, while those with high LAI values tend to exhibit a more negative impact. This phenomenon might be attributed to dense vegetation cover exerting a more pronounced insulating effect on the surface, potentially leading to increased overall STs during winter and consequently shortened frozen duration. Snow accumulation samples demonstrating a distinctly positive impact often correspond to high snow accumulation, while those displaying a notable negative impact usually have low snow accumulation. The differences in the impact brought about by varying winter snow accumulations are

more pronounced than those from spring and autumn snow quantities, which yield more uncertainties. As for spring precipitation, the majority of samples with high precipitation levels generally exhibit a negative impact, while a small fraction displays a positive influence. In terms of clay content, samples with lower clay content show considerable uncertainty in their impact on the frozen duration, whereas those with high clay content typically exhibit a negative impact. This result might be attributed to high clay content in soil, enhancing water retention and subsequently weakening interaggregate bonding after freezing, leading to structural instability and impeding the formation of perennial frozen soil.

Future freeze–thaw cycle change under different scenarios

The environmental response model for the frozen duration constructed by XGBoost predicts the spatiotemporal changes in the frozen duration from 2020 to 2050 under three SSPs: SSP126, SSP245, and SSP585. Comparative analysis indicates that in contrast to the current phase of the frozen duration from 1990 to 2020, the spatial distribution of the frozen duration under three future development paths shows no significant change but exhibits an overall reduction (Figure 5). Among these scenarios, the sustainability-oriented development path, SSP126, demonstrates the least reduction. Under this trajectory, the average frozen duration in the Northern Hemisphere is gener-

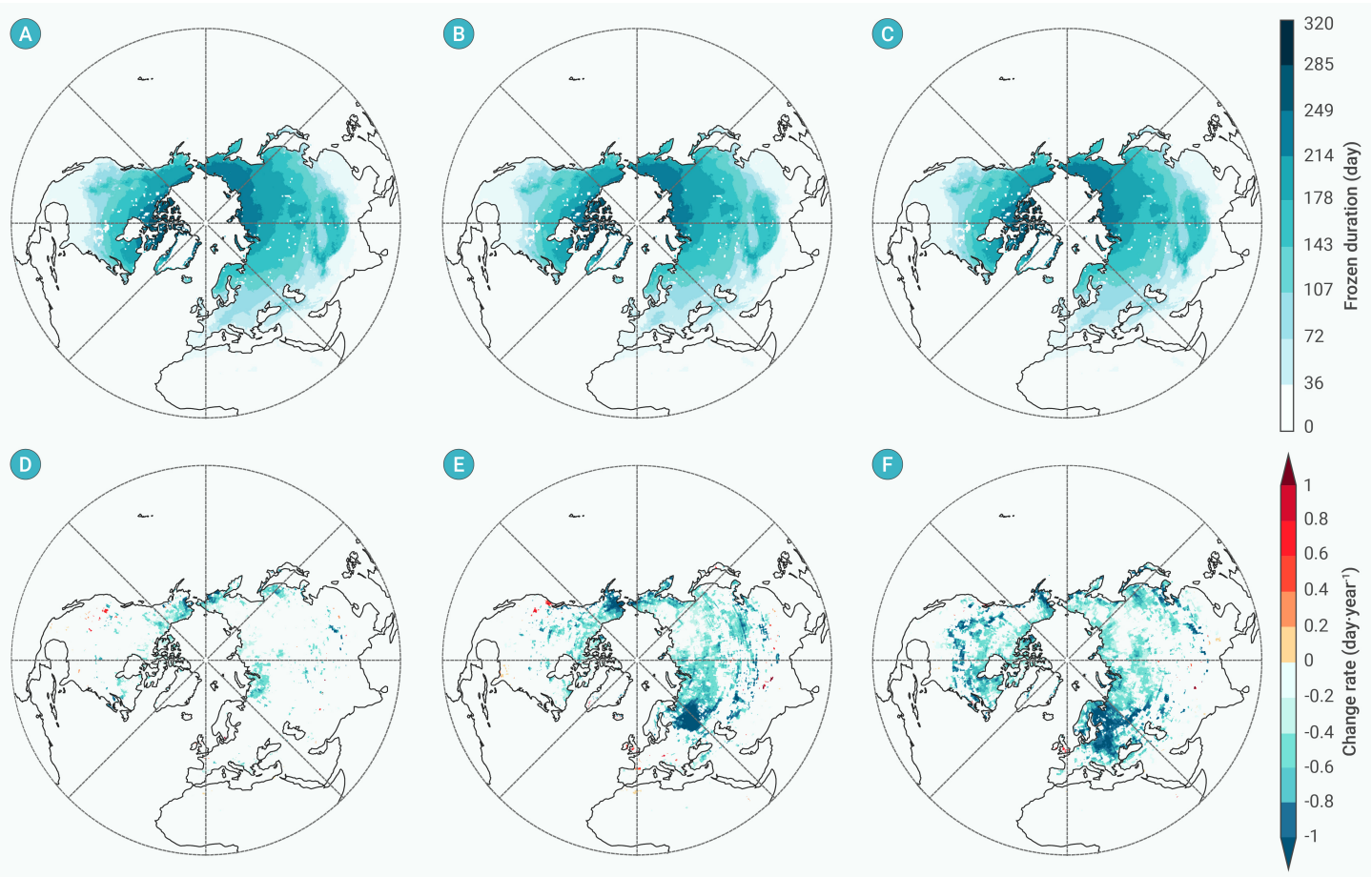


Figure 5. Freeze–thaw status change under three different emission scenarios Average duration of the freezing period (A–C) and its rate of change (D–F) under three different emission scenarios: A and D for SSP126, B and E for SSP245, C and F for SSP585.

ally approximately 139 ± 74 days. In comparison with the development path following historical pattern SSP245, the regions considerably affected by a reduction in the frozen duration under SSP126 are approximately 62% smaller. However, under the high-emission development path, i.e., SSP585, the regions considerably affected by a reduction in the frozen duration are approximately 33% larger than SSP245. Under SSP126, regions notably affected by a reduction include the Yukon region in Canada, Alaska in the United States, and the far eastern and northwestern regions of Russia. Within SSP245, the regions considerably affected by the reduction have expanded to encompass much of Russia, with additional notable reductions in eastern Mongolia and northern China regions. Meanwhile, under SSP585, the considerably affected reduction areas extend further into Europe, and considerable reductions are also observed in the northern regions of the contiguous United States. Overall, the advantages of the low-emission pathway SSP126 in preserving the cryosphere environment will become evidently pronounced in the next 30 years. In this development trajectory, the considerably affected regions by the reduction in the frozen duration are notably smaller, and the degree of reduction is comparatively modest.

The frozen durations under all three development paths continues to decrease following historical trends (Figure 6). Notably, the frozen duration's change is more gradual under SSP126, with an average reduction rate of 0.14 days per year. By contrast, the changes in the frozen duration under SSP245 and SSP585 appear more similar, potentially due to the limited immediate impact of the high-emission path (SSP585) on the cryosphere within the next 30 years. In terms of the rate of change in the frozen duration, SSP585 exhibits a higher reduction rate at 0.37 days per year compared with SSP245 with a rate of 0.29 days per year. Therefore, over an extended period, SSP585 might present more noticeable differences in the extent of reduction in the frozen duration. Overall, the period from 2020 to 2050 shows a continuing trend of decreasing frozen durations across the three development scenarios. However, this trend is notably more gradual under the sustainability-oriented

development path SSP126. Consequently, further advancement and implementation of sustainable development strategies hold importance in preserving the cryosphere environment. The climate model forecast data used in this study have a relatively short time range. CMIP6 climate models can provide predictions extending until the year 2100. Utilizing longer time-series climate model data can forecast long-term freeze–thaw cycle variations, thereby offering more valuable information support for devising strategies to address unstable permafrost changes. Additionally, the surface freeze–thaw cycle forecasts based on climate model data do not account for the influence of sudden events on the environment. Hence, they can only provide a reference for the overall trend in the freezing period changes.

CONCLUSION

This study employed the stacking algorithm for satellite–ground fusion calibration, significantly improving the accuracy of the original freeze–thaw products. Based on this enhanced dataset, we analyzed the spatial distribution and long-term trends of the freeze–thaw period in the Northern Hemisphere. The results indicate that the spatial distribution of the freeze–thaw period is closely related to latitude and elevation. Over the past 30 years, regions exhibiting significant changes in the freeze–thaw period generally show a trend of shortening. Furthermore, the length of the freeze–thaw period demonstrates an inverse relationship with surface temperature changes. To explore the driving factors behind these trends, we applied the XGBoost–SHAP algorithm, which accurately describes how the surface freeze–thaw cycle responds to environmental variables, including temperature, precipitation, snow cover, vegetation, topography, and soil texture. Our analysis revealed that autumn temperature has the most significant influence on the changes in the freeze–thaw period. Finally, environmental response models built using XGBoost were employed to predict potential future changes in the freeze–thaw period under three different emission scenarios: SSP126, SSP245, and SSP585. The predictions suggest that the freeze–thaw period

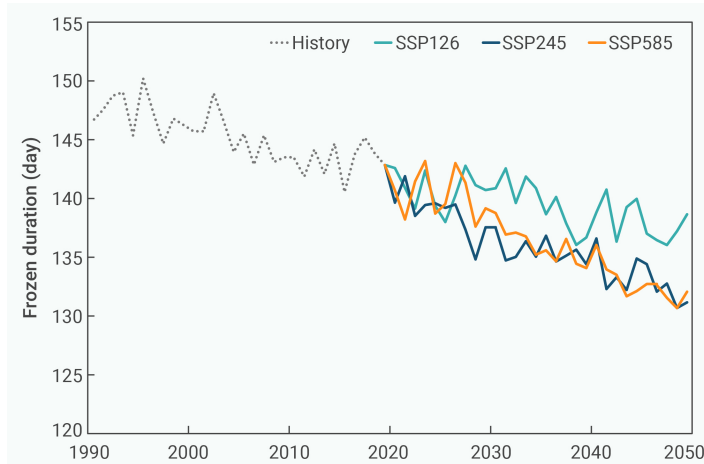


Figure 6. Variation trends of the frozen durations under three SSPs.

will continue to shorten over the next 30 years across all three pathways. However, the shortening trend is markedly slower under the more sustainable development pathway, SSP126. This study, by developing an integrated learning-based framework for monitoring and analyzing the surface freeze-thaw cycle, offers a more accurate and comprehensive understanding of the general trends in the Northern Hemisphere's freeze-thaw cycle. It provides valuable insights for developing strategies to adapt to and mitigate the impacts of changes in the freeze-thaw period.

REFERENCES

- Dobinski W. (2011). Permafrost. *Earth-Science Rev.* **108**:158–169. DOI:10.1016/j.earscirev.2011.06.007
- Schuur E. A., McGuire A. D., Schadel C., et al. (2015). Climate change and the permafrost carbon feedback. *Nature* **520**:171–179. DOI:10.1038/nature14338
- Biskaborn B. K., Smith S. L., Noetzi J., et al. (2019). Permafrost is warming at a global scale. *Nat. Commun.* **10**:264. DOI:10.1038/s41467-018-08240-4
- Running S. W., Way J., McDonald K., et al. (1999). Radar remote sensing proposed for monitoring freeze - thaw transitions in boreal regions. *Eos, Trans. Am. Geophys. Union* **80**:213–221.
- Gao B., Yang D., Qin Y., et al. (2018). Change in frozen soils and its effect on regional hydrology, upper Heihe basin, northeastern Qinghai–Tibetan Plateau. *The Cryosphere* **12**:657–673. DOI:10.5194/tc-12-657-2018
- Liu Z., Kimball J. S., Ballantyne A. P., et al. (2022). Respiratory loss during late-growing season determines the net carbon dioxide sink in northern permafrost regions. *Nat. Commun.* **13**:5626. DOI:10.1038/s41467-022-33293-x
- Turetsky M. R., Abbott B. W., Jones M. C., et al. (2020). Carbon release through abrupt permafrost thaw. *Nat. Geosci.* **13**:138–143. DOI:10.1038/s41561-019-0526-0
- Zimov S. A., Schuur E. A. G. and Chapin F. S. (2006). Permafrost and the Global Carbon Budget. *Science* **312**:1612–1613. DOI:10.1126/science.1128908
- Walter Anthony K. M., Anthony P., Grosse G., et al. (2012). Geologic methane seeps along boundaries of Arctic permafrost thaw and melting glaciers. *Nat. Geosci.* **5**:419–426. DOI:10.1038/ngeo1480
- Haeblerl W. and Beniston M. (1998). Climate change and its impacts on glaciers and permafrost in the alps. *Ambio* **27**:258–265.
- Teufel B. and Sushama L. (2019). Abrupt changes across the Arctic permafrost region endanger northern development. *Nat. Clim. Change* **9**:858–862. DOI:10.1038/s41558-019-0614-6
- IPCC (2022). Climate change 2022: impacts, adaptation, and vulnerability. Cambridge University Press.
- Smith S. L., O'Neill H. B., Isaksen K., et al. (2022). The changing thermal state of permafrost. *Nat. Rev. Earth Environ.* **3**:10–23. DOI:10.1038/s43017-021-00240-1
- Atchley A. L., Painter S. L., Harp D. R., et al. (2015). Using field observations to inform thermal hydrology models of permafrost dynamics with ATS (v0.83). *Geosci. Model Dev.* **8**:2701–2722. DOI:10.5194/gmd-8-2701-2015.
- Oldenborger G. A., Short N. and LeBlanc A.-M. (2022). Permafrost thaw sensitivity prediction using surficial geology, topography, and remote-sensing imagery: a data-driven neural network approach. *Can. J. of Earth Sci.* **59**:897–913. DOI:10.1139/cjes-2021-0117
- Gao H., Nie N., Zhang W., et al. (2020). Monitoring the spatial distribution and changes in permafrost with passive microwave remote sensing. *ISPRS J. Photogramm. Remote Sens.* **170**:142–155. DOI:10.1016/j.isprsjprs.2020.10.011
- Zhang T. and Armstrong R. L. (2001). Soil freeze/thaw cycles over snow-free land detected by passive microwave remote sensing. *Geophys. Res. Lett.* **28**:763–766. DOI:10.1029/2000GL011952
- Li T., Shen H., Yuan Q., et al. (2017). Estimating ground-level PM_{2.5} by fusing satellite and station observations: a geo-intelligent deep learning approach. *Geophys. Res. Lett.* **44**:11,985–911,993. DOI:10.1002/2017GL075710.
- Ma H., Zeng J., Zhang X., et al. (2024). Surface soil moisture from combined active and passive microwave observations: Integrating ASCAT and SMAP observations based on machine learning approaches. *Remote Sens. Environ.* **308**:114197. DOI:10.1016/j.rse.2024.114197
- Alerskans E., Zinck A.-S. P., Nielsen-Englyst P., et al. (2022). Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements. *Remote Sens. Environ.* **281**. DOI:10.1016/j.rse.2022.113220.
- Xiao X., Zhang T., Zhong X., et al. (2018). Support vector regression snow-depth retrieval algorithm using passive microwave remote sensing data. *Remote Sens. Environ.* **210**:48–64. DOI:10.1016/j.rse.2018.03.008
- Cuo L., Zhang Y., Bohn T. J., et al. (2015). Frozen soil degradation and its effects on surface hydrology in the northern Tibetan Plateau. *J. Geophys. Res. Atmos.* **120**:8276–8298. DOI:10.1002/2015JD023193
- Dai L., Guo X., Zhang F., et al. (2019). Seasonal dynamics and controls of deep soil water infiltration in the seasonally-frozen region of the Qinghai-Tibet plateau. *J. Hydrol.* **571**:740–748. DOI:10.1016/j.jhydrol.2019.02.021
- Heijmans M. M., Magnússon R. Í., Lara M. J., et al. (2022). Tundra vegetation change and impacts on permafrost. *Nat. Rev. Earth Environ.* **3**:68–84. DOI:10.1038/s43017-021-00233-0
- Lehrsch G., Sojka R., Carter D., et al. (1991). Freezing effects on aggregate stability affected by texture, mineralogy, and organic matter. *Soil Sci. Soc. Am. J.* **55**:1401–1406. DOI:10.2136/sssaj1991.03615995005500050033x
- Yuan Q., Shen H., Li T., et al. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **241**:111716. DOI:10.1016/j.rse.2020.111716
- Polikar R. (2012). Ensemble learning. *Ensemble machine learning: Methods and applications*, pp:1–34. DOI:10.1007/978-1-4419-9326-7_1.
- Yang J., Jiang L., Lemmetyinen J., et al. (2021). Improving snow depth estimation by coupling HUT-optimized effective snow grain size parameters with the random forest approach. *Remote Sens. Environ.* **264**:112630. DOI:10.1016/j.rse.2021.112630
- Xu Y., Wang L., Ma Z., et al. (2019). Spatially explicit model for statistical downscaling of satellite passive microwave soil moisture. *IEEE Trans. Geosci. and Remote Sens.* **58**:1182–1191. DOI:10.1109/TGRS.2019.2944421
- Ni J., Wu T., Zhu X., et al. (2021). Simulation of the present and future projection of permafrost on the Qinghai - Tibet Plateau with statistical and machine learning models. *J. Geophys. Res. Atmos.* **126**:e2020JD033402. DOI:10.1029/2020JD033402
- Siewert M. B. (2018). High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: a case study in a sub-Arctic peatland environment. *Biogeosciences* **15**:1663–1682. DOI:10.5194/bg-15-1663-2018
- Sagi O. and Rokach L. (2018). Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* **8**:e1249. DOI:10.1002/widm.1249
- Yang Q., Kim J., Cho Y., et al. (2023). A synchronized estimation of hourly surface concentrations of six criteria air pollutants with GEMS data. *npj Clim. Atmos. Sci.* **6**:94. DOI:10.1038/s41612-023-00407-1.
- Reichstein M., Camps-Valls G., Stevens B., et al. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature* **566**:195–204. DOI:10.1038/s41586-019-0912-1
- Liu X., Pei F., Wen Y., et al. (2019). Global urban expansion offsets climate-driven increases in terrestrial net primary productivity. *Nat. Commun.* **10**:5558. DOI:10.1038/s41467-019-13462-1
- Webb E. E., Liljedahl A. K., Cordeiro J. A., et al. (2022). Permafrost thaw drives surface water decline across lake-rich regions of the Arctic. *Nat. Clim. Change* **12**:841–846. DOI:10.1038/s41558-022-01455-w
- Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.
- Kim Y., Kimball J. S., Glassy J., et al. (2017). An extended global Earth system data record on daily landscape freeze-thaw status determined from satellite passive microwave remote sensing. *Earth Syst. Sci. Data* **9**:133–147. DOI:10.5194/essd-9-133-2017
- Kim Y., Kimball J. S., McDonald K. C., et al. (2011). Developing a global data record of daily landscape freeze/thaw status using satellite passive microwave remote sensing. *IEEE Trans. Geosci. Remote Sens.* **49**:949–960. DOI:10.1109/tgrs.2010.2070515
- Hu T., Zhao T., Zhao K., et al. (2019). A continuous global record of near-surface soil freeze/thaw status from AMSR-E and AMSR2 data. *Int. J. Remote Sens.* **40**:6993–7016. DOI:10.1080/01431161.2019.1597307
- Amante C. and Eakins B. W. (2009). ETOPO1 arc-minute global relief model : procedures, data sources and analysis.
- Meehl G. A., Boer G. J., Covey C., et al. (2000). The Coupled Model Intercomparison Project (CMIP). *Bull. Am. Meteorol. Soc.* **81**:313–318. DOI:10.1175/1520-0477(2000)081<0313:TCMIPC>2.3.CO;2
- Eyring V., Bony S., Meehl G. A., et al. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**:1937–1958. DOI:10.5194/gmd-9-1937-2016
- Meinshausen M., Nicholls Z. R. J., Lewis J., et al. (2020). The shared socio-economic

- pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev.* **13**:3571–3605. DOI:10.5194/gmd-13-3571-2020
45. Friedl M. A., Sulla-Menashé D., Tan B., et al. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **114**:168–182. DOI:10.1016/j.rse.2009.08.016
 46. MARKUS KOTTEK, JÜRGEN GRIESER, CHRISTOPH BECK, et al. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **15**:259–263. DOI:10.1127/0941-2948/2006/0130
 47. Obu J., Westermann S., Bartsch A., et al. (2019). Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km² scale. *Earth Sci. Rev.* **193**:299–316. DOI:10.1016/j.earscirev.2019.04.023
 48. Hengl T., Mendes de Jesus J., Heuvelink G. B. M., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **12**:e0169748. DOI:10.1371/journal.pone.0169748
 49. Belgiu M. and Drăguț L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **114**:24–31. DOI:10.1016/j.isprsjprs.2016.01.011
 50. Chen T. and Guestrin C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery.
 51. Ke G., Meng Q., Finley T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30. <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>

FUNDING AND ACKNOWLEDGMENTS

This work was supported by National Key R&D Program of China (2022YFB3903400), Fundamental Research Funds for the Central Universities (2042024kf0020). We thank F. Baldacchino for the suggestions on how to improve the manuscript. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Q. Yuan conceptualized the study. W. Zhong developed the methodology and conducted the experiments. W. Zhong, Q. Yang, and Y. Peng drafted the manuscript. Q. Yang and W. Zhong produced the visualizations. Q. Yuan, T. Bolch, Y. Wang, H. Shen, and L. Zhang reviewed, contributed to, and edited the manuscript. Q. Yuan solicited and received the funding. Q. Yuan, L. Zhang, and T. Bolch provided supervision. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AND CODE AVAILABILITY

The MeASURES freeze–thaw state product can be downloaded from <https://nsidc.org/data/nsidc-0477/versions/5>. The ground temperature data were provided by the global historical climate network can be downloaded from <https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily>. The ETOP01 Global Relief Model was provided by <https://ngdc.noaa.gov/mgg/global/relief/ETOP01/tiled/>. The CMIP6 climate model data can be found in <https://cds.climate.copernicus.eu/cdsapp#!/dataset/projections-cmip6?tab=form>. MODIS land cover product (MCD12C1) can be downloaded from <https://ladsweb.modaps.eosdis.nasa.gov/search/order/1/MCD12C1--61>. Köppen-Geiger global climate zone data can be found in <https://www.gloh2o.org/koppen/>. PERZONES North Hemisphere multi-year permafrost range data can be downloaded from <https://doi.pangaea.de/10.1594/PANGAEA.888600>. SoilGrids250m product can be downloaded at <https://soilgrids.org/>. The code used to perform the experiments is available from the corresponding author (zlp62@whu.edu.cn) upon request.

SUPPLEMENTAL INFORMATION

It can be found online at <https://doi.org/10.59717/j.xinn-geo.2024.100118>