High-precision estimation of pan-Arctic soil surface temperature from MODIS LST by incorporating multiple environment factors and monthly-based modeling

Hongxiang Guo [a,b], Wenquan Zhu [a,b,*], Cunde Xiao [c], Cenliang Zhao [a,b], Liyuan Chen [a,b]

[a] *State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China*
[b] *Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*
[c] *State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China*

ABSTRACT

Global warming has shown an "Arctic amplification effect" in recent decades, leading to pronounced changes in pan-Arctic soil surface temperature (SST). SST plays a direct role in energy exchange between soil and atmosphere and serves as an indicator of the land–atmosphere energy balance. Remote sensing land surface temperature (LST) data is able to indicate near-surface temperature, but influences from environment factors, such as vegetation and snow, can introduce biases between LST and SST. In this study, the importances of five environment factors (vegetation, snow, surface soil composition, topography, and solar radiation) to monthly mean SST estimation from MODIS LST in pan-Arctic were analyzed. Then a method for pan-Arctic monthly mean SST estimation from MODIS LST by incorporating these environment factors and monthly-based modeling based on random forest (RF) algorithm was proposed. The results reveal that all the selected environment factors contribute to monthly-based modeling, with vegetation exerting the greatest importance from May to October and snow in March and April. The root mean square error (RMSE) of pan-Arctic monthly SST estimated by the proposed method from 2003 to 2022 ranges from 0.89 to 1.88 °C, which is a 42.95——53.35 % reduction compared to the widely used season-based multivariate linear regression (MLR) models based solely on LST (RMSE between 1.56 and 4.03 °C). The accuracy is notably improved in areas with lower and no vegetation (grassy woodlands, grasslands, permanent wetlands, and barrens) in the cold season (September to the following April), and in higher vegetation (forests) areas in the warm season (May to August). The proposed method can contribute to producing high-precision monthly mean SST data from LST, estimating permafrost extent and active layer thickness, and understanding the land–atmosphere energy balance in pan-Arctic.

1. Introduction

Over the past several decades, global warming has exhibited an "amplification effect" in the Arctic, leading to significant changes in terrestrial environments across the pan-Arctic (AMAP, 2017, 2021; IPCC, 2022). Variations in soil surface temperature (SST) (0——10 cm depth) are important components of terrestrial environment changes (Biskaborn et al., 2019; Luo et al., 2016). SST is a crucial indicator of the surface-atmosphere energy balance and a primary climatic factors affecting the existence and distribution of permafrost (Hachem et al., 2012). Developing a high-precision SST estimation method for pan-Arctic is vital for producing high-precision SST data products, contributing to estimating permafrost extent and active layer thickness, and

understanding the stable state and dynamic changes of cryosphere-related environment elements, such as Arctic climate and eco-hydrology.

In permafrost numerical models, two common methods are used to build relationship between SST and air temperature. One is the near-surface air temperature conversion factor method, where the ratio of near-surface air temperature to SST, known as the n factor, is used to improve the SST parameters in the Stefan equation (Kurylyk and Hayashi, 2016). The n factor, calculated based on site-observed SST and air temperature, is often assigned according to land covers in studies with a larger spatial scope (Cai et al., 2020; Park et al., 2016). However, this approach may fail to account for the spatial heterogeneity of the n factor within the same land cover. The other method is the empirical

---

relationship method. The Kudryavtsev model (Peng et al., 2024) establishes relationships between annual mean and amplitude of SST with air temperature by empirical relationships. This method can eliminate the biases between air temperature and SST that are caused by snow and vegetation. However, parameters required for this empirical relationship, such as vegetation height, vegetation thermal conductivity, and snow thermal conductivity, are often difficult to be obtained accurately over a large spatiotemporal scale in pan-Arctic, which can impact the applicability of the method. Additionally, the spatial resolution of widely-used air temperature data, such as the ERA5-Land datasets with a 0.1° spatial resolution (Muñoz-Sabater et al., 2021), constrains their ability to finer SST estimation.

Land surface temperature (LST) derived from remote sensing has been found to have a high correlation with SST (Myhra et al., 2017; Westermann et al., 2015; Xu et al., 2020) and has a finer (1 km) spatial resolution. Consequently, models driven by LST data for mapping permafrost have attracted increasing attentions (Myhra et al., 2017; Westermann et al., 2015). However, LST derived from satellites observed longwave radiation is typically the instantaneous thermodynamic temperature of land surface (Reiners et al., 2023). The various land surface components (such as forests, grasslands and snow) can lead to biases between LST and SST (Fig. 1) (Niclòs et al., 2023). Estimating mean SST over a period from instantaneous LST data is the primary challenge for applying LST in permafrost mapping. Obu et al. (2019) filled the missing data in MODIS LST (MOD11A1 and MYD11A1) products with the ERA-interim reanalysis data and taking the 8-day average of four instantaneous MODIS LST as a 8-day mean LST to drive the TTOP model for permafrost mapping. Considering that winter snow can introduce biases between LST and SST, an empirical relationship was employed to calculate n factor in the cold season to derive SST, while LST was assumed to be equivalent to the SST in the warm season. Zou et al. (2014) proposed to establish the relationship between mean SST and four instantaneous MODIS LST data using a multivariate linear regression (MLR) model. The 8-day average MODIS LST products (MOD11A2 and MYD11A2) were used to estimate the mean SST by a MLR model in Tibetan Plateau because they are more reliable than daily products (Zou et al., 2017). Then SST data were used to drive TTOP

model for permafrost mapping. Cao et al. (2021) considered that changes in vegetation and snow would cause differences in the MLR models in different seasons, and established separate MLR models for warm season (May to August) and cold season (September to the following April).

In natural environment conditions, biases between SST and LST can be influenced by various environment factors, such as vegetation, snow, and soil composition (Niclòs et al., 2023). Lembrechts et al. (2022) trained random forest (RF) models based on site observed data from the SoilTemp dataset and multiple environment factors to estimate the biases between air temperature and SST at global scale. This idea, which estimates temperature biases between SST and air temperature by incorporating multiple environment factors, provides a new perspective for large-scale SST estimation based on remote sensing LST data. Compared to air temperature data, such as ERA5-Land data, MODIS LST data have a higher spatial resolution and can reflect finer spatial temperature variations. However, the biases between LST and SST are influenced by various environment factors, and some environment factors (such as vegetation and snow) change significantly over months, leading to variations in biases between LST and SST across months. To improve the estimation accuracy of large-scale mean SST based on LST data, it is necessary to optimize the method in two ways: Firstly, by integrating multiple environment factors related to SST; and secondly, by establishing separate SST estimation models for different months.

Considering that the SST in pan-Arctic is influenced by multiple environment factors, and these environment factors vary among different calendar months, this study aims to estimate high-precision SST across the pan-Arctic from MODIS LST by integrating various environment factor data and monthly-based modeling. Specifically, we focused on three main works: firstly, analyzing the contribution of multiple environment factors to SST estimation models for different months; secondly, establishing SST estimation RF models based on LST and multiple environment factor data for different months; and finally, comparing the SST estimation accuracies with existing methods, both overall accuracies and the accuracies in different land covers.
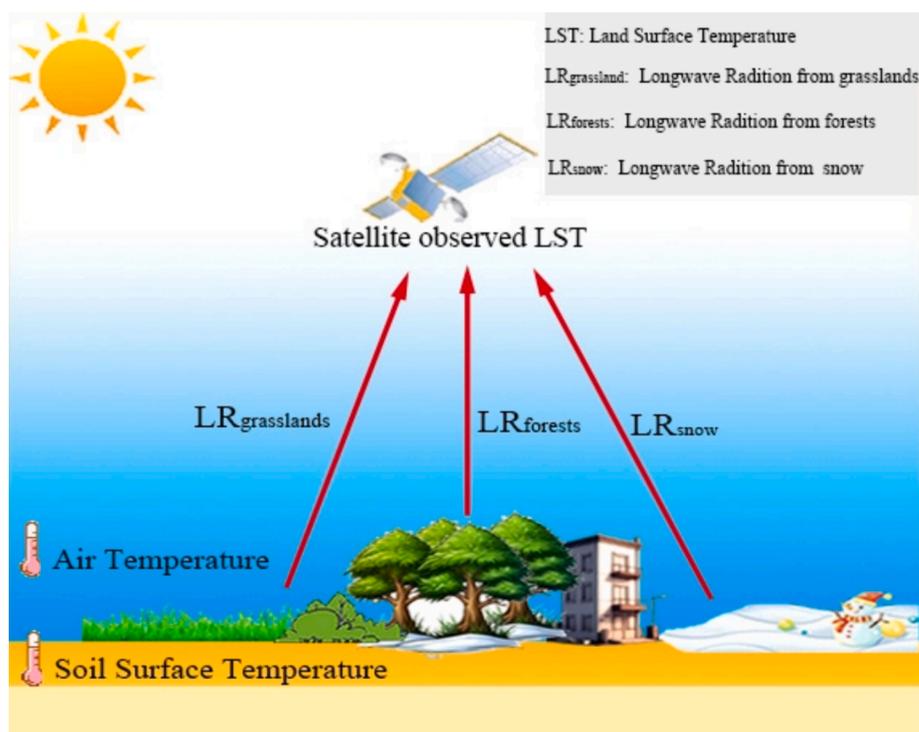

Fig. 1. Differences between air temperature, LST, and soil surface temperature.

2. Study area and data

2.1. Study area

The land area north of 60 °N latitude was selected as the study area. According to the ERA5 climate reanalysis data (Hersbach et al., 2017), the annual average air temperature in this area was between −11 and −8 °C during 2003——2022, with a warming rate of approximately 0.06 °C per year. During the warm season, many areas are covered with vegetation, and persistent snowpack tends to form in the cold season. This area is also one of the regions with the most extensively permafrost distribution, with a permafrost area exceeding 1.0×10^7 km^2, accounting for over half of the total permafrost area in the Northern Hemisphere (Brown et al., 2002). Furthermore, the permafrost in the study area has shown a degradation trend (Biskaborn et al., 2019; Luo et al., 2016; Smith et al., 2022).

2.2. Data

2.2.1. Site observed data

The SST site observed data were obtained from the SoilTemp and Global Terrestrial Network for Permafrost (GTN-P) databases. SoilTemp is a global soil temperature time series database (Lembrechts et al., 2020) which contains site observed soil temperature data from 60 countries and regions. The monthly average soil temperature site observed data have been calculated and published (Lembrechts et al., 2022). GTN-P, developed by the International Permafrost Association (IPA) based on the Global Climate Observing System (GCOS) and the Global Terrestrial Observing System (GTOS) in 1999, collects soil temperature site observed data in permafrost area globally (Biskaborn et al., 2015). Only months with more than 26 days of observed data were selected to calculate the monthly average values. The site observed data were obtained from 1453 soil temperature sites from 2003 to 2022 (Fig. 2).

2.2.2. Spatial datasets

Spatial datasets include MODIS LST data, environment factor (vegetation, snow, solar radiation, surface soil composition, topography) data and land cover data. MODIS LST data and environment factor data were used to estimate SST. Land cover data were used to validate the accuracies of SST estimations by regions.

The MODIS LST data were collected from the MOD11A1 (Wan et al., 2021b) and MYD11A1 (Wan et al., 2021a) products, which were derived from Terra and Aqua satellites data, respectively. Both of them includes two observed data: daytime data (Terra$_{day}$, Aqua$_{day}$) and nighttime data (Terra$_{night}$, Aqua$_{night}$).

Data related to five environment factors (vegetation, snow, surface soil composition, topography, and solar radiation) were collected (Table 1). Among these environment factors, vegetation, snow, and solar radiation vary among months. Vegetation data include normalized difference vegetation index (NDVI) and leaf area index (LAI), both of which are MODIS data products (Yuan et al., 2011). Snow data include two types of data products: one derived from MODIS data, including normalized difference snow index (NDSI) and NDSI snow cover (Hori et al., 2017), and the other from the ERA5-Land dataset (Muñoz-Sabater et al., 2021), including snow depth and snow density data. Solar radiation data were collected from MCD18A1 products, which includes shortwave solar radiation (DSR) data every 3 h (Wang, 2024). MODIS optical sensor data are missing in January, November, and December due to the polar night phenomenon, so the data derived from MODIS data for these months were excluded. In addition, MODIS data contain missing values and unreliable measurements due to clouds and atmospheric effects. Environment factors that do not vary over time include surface soil composition and topography. Surface soil composition sourced from the SoilGrids250 dataset, including soil bulk density, moisture, clay, silt, soil organic carbon (SOC), and sand content. The SoilGrids250 is produced through machine learning algorithms, which are trained based on a large number of soil profile site observed data and environment factor data samples (Hengl et al., 2017). Topography data were obtained from the ArcticDEM (Friedl and Sulla-Menashe, 2022), and the slope and aspect data were calculated based on this dataset. The ArcticDEM data is a digital surface model (DSM) of the Arctic automatically generated by photogrammetric software using optical stereo imagery, although there are some strip gaps.

Land cover data were used to evaluate the SST estimation accuracy in different regions. The MCD12Q1 data products (Friedl and Sulla-Menashe, 2022) were collected, which contain land cover data for each year from 2003 to 2022. The data were produced by a supervised classification and the data accuracies were further improved by post-processing.

3. Methodology

The method for estimating pan-Arctic SST by incorporating multiple environment factors and monthly-based modeling based on random forest algorithms (monthly-based RF) is shown in Fig. 3. Firstly, the site observed and environment factor data were preprocessed. Secondly, the sample dataset was established. The LST and environment factor data from pixels corresponding the sites' location were extracted as samples, among which 70 % for model training and 30 % for testing. Then, the random forest models were trained monthly from February to October, and the monthly SST data were predicted. Finally, the model's overall accuracies and its accuracies across different land cover types were evaluated and compared with the existing methods.

3.1. Data preprocessing and samples database establishment

Data preprocessing was necessary before establishing samples database. Monthly composites for LST, vegetation, snow, and solar radiation were performed by monthly averaging. Performing monthly composites could fill in some missing values and mitigate the impacts of unreliable measurements, making the monthly data more reliable than daily data. For ArcticDEM data, ArcGIS 10.5 was used to compute slope and aspect, and gaps in data were filled using the inverse distance weighting interpolation method (Setianto and Triandini, 2013). All of the LST and
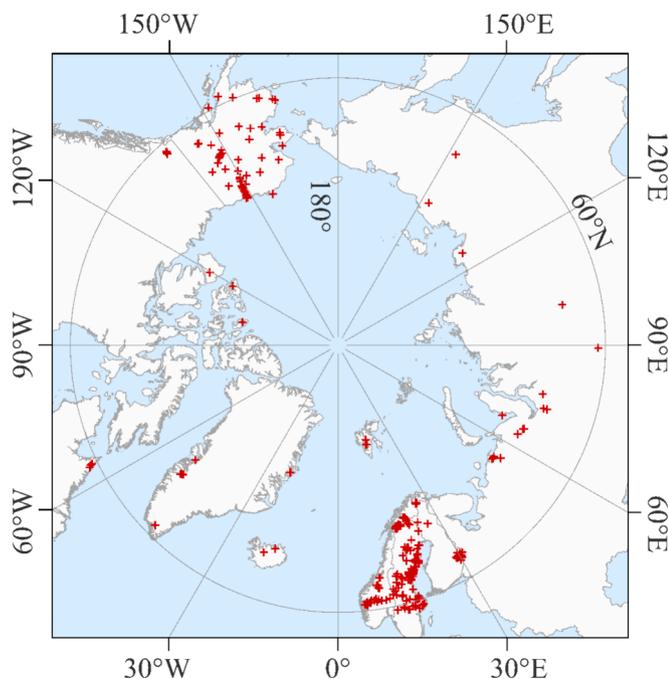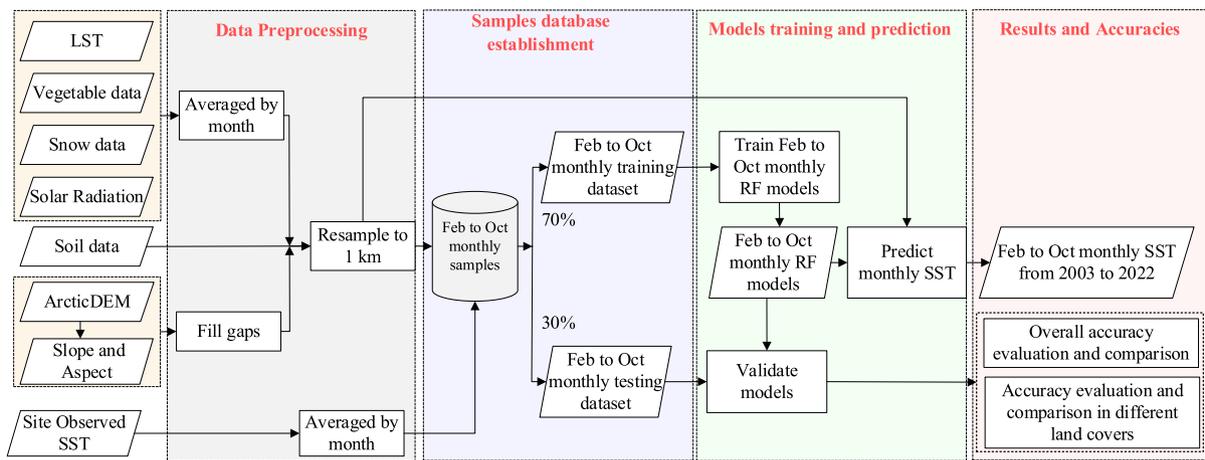

Fig. 2. Spatial distribution of sites.

Table 1
Spatial datasets.

Data categories	Data	Data source	Spatial resolution	Temporal resolution	Months covered	Years covered
MODIS LST	LST	MOD11A1 MYD11A1	1000 m	Daily	2–10	2003–2022
Vegetation	NDVI	MOD13A2 MYD13A2	1000 m	16 d	2–10	2003–2022
	LAI	MCD15A3H	500 m	4 d	2–10	2003–2022
Snow	NDSI snow cover NDSI	MOD10A1 MYD10A1	500 m	Daily	2–10	2003–2022
	Snow depth Snow density Precipitation	ERA5-Land	11132 m	Daily	2–10	2003–2022
Solar Radiation	DSR	MCD18A1	500 m	3 h	2–10	2003–2022
Surface soil composition	Bulk density Moisture Clay content Silt content SOC content Sand content	SoilGrid250	250 m	–	–	–
Topography	DSM Slope Aspect	ArcticDEM	2 m	–	–	–
Land Cover	Land Cover types	MCD12Q1	500 m	Yearly	–	2003–2022

Fig. 3. Flowchart of SST estimation method.

environment factor data were resampled by bilinear interpolation to obtain a raster dataset with 1 km spatial resolution.

Then, the raster pixel data corresponding to the location and observation time of each site data were extracted to establish the samples database. Pixels with missing data were excluded. For sites located within the same pixel, the observed data were averaged. All samples in

Table 2
Numbers of train and test samples for each month.

	Samples datasets	Forests	Shrublands	Grassy woodlands	Grasslands	Permanent Wetlands	Barrens	Total
Feb	training	84	144	63	311	32	9	643
	testing	36	63	32	134	16	5	286
Mar	training	83	151	65	307	16	15	637
	testing	36	57	32	139	7	8	279
Apr	training	86	150	55	304	16	14	625
	testing	31	53	32	141	6	6	269
May	training	86	151	63	295	20	19	634
	testing	31	60	18	151	8	8	276
Jun	training	92	146	87	304	17	19	665
	testing	51	63	58	132	7	8	319
Jul	training	95	152	104	298	14	19	682
	testing	49	65	36	130	7	7	294
Aug	training	106	135	100	319	19	20	699
	testing	42	58	40	148	8	8	304
Sep	training	96	132	72	285	14	17	616
	testing	38	57	30	123	9	8	265
Oct	training	75	131	67	300	20	11	604
	testing	41	58	27	123	11	6	266

the sample dataset were divided into two groups: 70 % were randomly selected as the training dataset, and the remaining 30 % were used as the testing dataset. To prevent overfitting, the stratified sampling was used to ensure that the number of samples for each land cover type was approximately maintained at a certain ratio between the training and testing datasets. The number of samples for each month is detailed in Table 2.

3.2. Modelling

3.2.1. Model description

The random forest algorithm was used to establish models that integrate LST and multiple environment factor data. It is an ensemble learning algorithm that aggregates multiple decision trees. First, the bootstrap sampling is used to extract multiple sample subsets from the sample dataset. Decision trees are constructed for each subset, and the average of prediction results from these decision trees is taken as the final prediction result (Belgiu and Drăguţ, 2016). The 36.8 % of the training samples not selected in the bootstrap sampling are employed to calculate out-of-bag errors, which are used to evaluate the importance of features (Tassi et al., 2021). The average importance of the feature data related to each environment factor was taken as the importance of that environment factor in this study. The RF model is characterized by effective prevention overfitting, robust universality, and insensitivity to missing data and multicollinearity. RF has demonstrated commendable performances in fitting nonlinear relationships (Lembrechts et al., 2022; Ni et al., 2021). The RF models build in this study can be abstractly represented by Formula 1.

$$\text{SST} = f_1(\text{LST}) + f_2(\text{Veg}) + f_3(\text{Snow}) + f_4(\text{Soil}) + f_5(\text{Topography}) + f_6(\text{SR}) \tag{1}$$

where LST represents the land surface temperature data derived from MODIS satellite. Veg, Snow Soil, Topography and SR represent the vegetation data, snow data, soil composition data, topography data and solar radiation data, respectively.

3.2.2. Models training and prediction

The random forest models for each month were trained based on the data from the sample dataset, and then the pan-Arctic SST from 2003 to 2022 was predicted. Before training models, all of the LST and environment factor data were normalized to a range of 0 to 1 by dividing their maximum values. All processes were implemented based on the scikit-learn package in Python. The model training procedure includes two steps: First, determine the optimal hyperparameters (number of trees and max features) for the models. The number of trees was set from 10 to 300 with a step size of 5, and the maximum features was set from 1 to 20. The 10-fold cross-validation was used to evaluate the performances of models based on various hyperparameters (details can be seen in Table A. 1) (Ni et al., 2021). The hyperparameters corresponding to the smallest cross-validation error were chosen (Table 3). Second, models for each month were trained based on training dataset, and the testing dataset was used for accuracy assessment. Finally, the trained models were used to predict the February to October monthly SST data from 2003 to 2022. The missing values in SST results were filled with the average of the adjacent months.

3.3. Accuracy evaluation

To independently assess the contributions of incorporating multiple environment factors and monthly-based modeling to SST estimation accuracy, comparisons were made with the results calculated by three other methods: MLR modeling by season (season-based MLR) (Eq. (2)) (Cao et al., 2021), MLR modeling by month (monthly-based MLR), RF modelling by season (season-based RF). The coefficient of determination (R^2) (Eq. (3)) and root mean square error (RMSE) (Eq. (4)) were used as the metrics for quantitative accuracy assessment. It is worth noting that the widely used multiple linear regression method establishes models solely between SST and LST (Eq. (2)), neglecting the influence of multiple environment factors.

$$\text{SST} = a \times \text{Terra}_{\text{day}} + b \times \text{Aqua}_{\text{day}} + c \times \text{Terra}_{\text{night}} + d \times \text{Aqua}_{\text{night}} + e \tag{2}$$

where, $\text{Terra}_{\text{day}}$ and Aqua_{day} represent the LST data observed by MODIS sensors on Terra and Aqua satellites during the daytime, respectively. $\text{Terra}_{\text{night}}$ and $\text{Aqua}_{\text{night}}$ represent the LST data observed by MODIS sensors on Terra and Aqua satellites during the nighttime, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\text{SST}_{\text{obs},i} - \text{SST}_{\text{mod},i}\right)^2}{\sum_{i=1}^{n}\left(\text{SST}_{\text{obs},i} - \overline{\text{SST}_{\text{obs}}}\right)^2} \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\text{SST}_{\text{obs},i} - \text{SST}_{\text{mod},i}\right)^2} \tag{4}$$

where SST_{obs} represents site observed SST data, SST_{mod} represents modeled SST data, n is the number of samples, $\overline{\text{SST}_{\text{obs}}}$ represents average of all site observed SST data.

4. Results

4.1. Importance of environment factors for monthly-based modeling

Fig. 4 shows the importance of environment factors in SST estimation modeling in different months. It can be seen that the all five selected environment factors contribute to the monthly-based modeling, although importance vary among different months. The vegetation and snow are the most important factors for modeling. The importance of vegetation to the model is lower than 2.5 % from February to April, and it fluctuates, increasing and then decreasing, from May to October, peaking at a maximum of 12.47 % in June. The importance of snow to models is lower from May to August (1.18 – 3.28 %), while higher in the cold season. The highest values are observed in March and April, which are 4.97 % and 4.23 %, respectively.

4.2. Overall accuracy evaluation

Monthly SST data in February to October were produced from 2003 to 2022 based on monthly-based RF modeling (Fig. 5). The SST trends in pan-Arctic can be seen in Figure A. 1 and Figure A. 2. The overall accuracies were evaluated based on site observed data in the sample dataset, and the accuracies were compared with the results calculated by other methods.

Comparison of overall accuracy by different models shows that the RF models fed with LST and multiple environment factor data have significantly improved in accuracy compared with MLR models based

Table 3
Hyperparameters of Random Forest models.

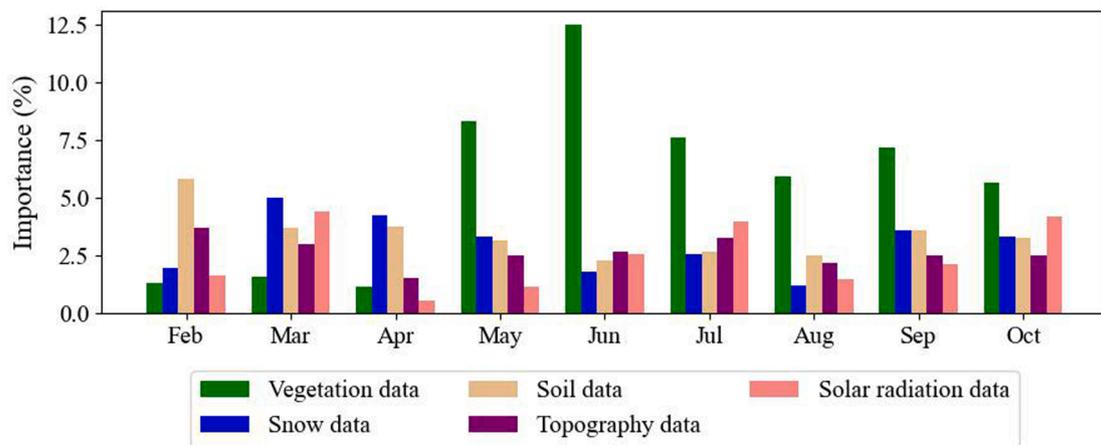| Model parameters	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
Number of trees	231	86	191	176	246	86	106	71	246
Max features	12	4	3	2	3	5	8	11	13

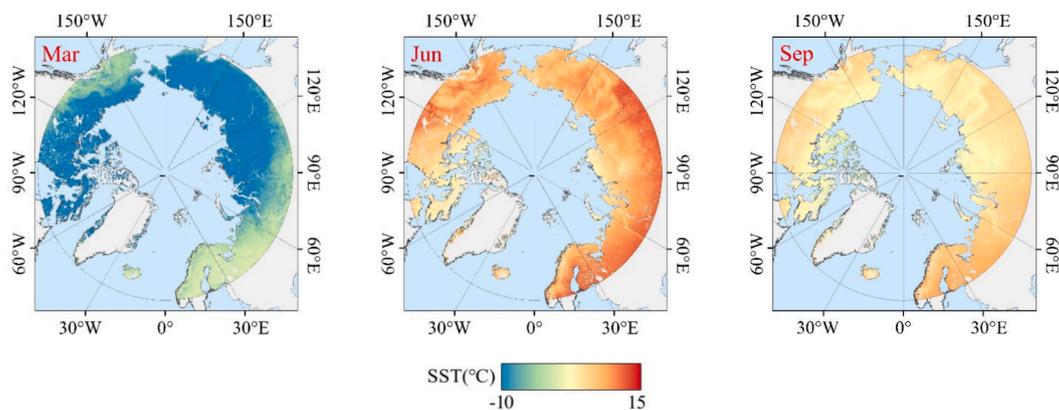**Fig. 4.** Importance of environment factors for month-based modeling.

Fig. 5. Average SST for the period 2003——2022 in Mar, Jun, and Sep.

solely on LST both in training (Table 4) and testing (Table 5). From the comparison of training accuracy, the RMSE of the seasonal and monthly RF models are 0.68——0.69 °C and 0.49——0.85 °C, respectively, which are far better than the 2.57——3.13 °C and 1.7——4.18 °C of the seasonal and monthly MLR models. The accuracy validation results based on the test set are similar, with the random forest model significantly outperforming the MLR results. According to results of season-based modeling, the RMSEs of RF models compared to MLR models reduce by 0.68——1.3 °C and the R^2 increase by 0.24——0.6 in warm season, and the RMSE decrease by 0.59——2.14 °C and the R^2 increase by 0.17——0.37 in cold season. According to the results of monthly-based models, the RMSEs for RF models range from 0.89 to 1.88 °C, which is lower than the 1.46——4.03 °C of the MLR model based solely on LST. Among them, the average RMSEs reduce by 0.64 °C, the R^2 increase by 0.23 in warm season, with the most significant improvement in accuracy

in May. The average RMSEs reduce by 1.27 °C, the R^2 increase by 0.266 in the cold season, with the most significant improvement in accuracy in March.

Table 5 also shows that the monthly-based models have higher accuracy compared with the season-based models. According to testing accuracies, the RMSEs of monthly-based MLR models for different months ranges from 1.41 to 4.03 °C, which is 0.02 – 1.01 °C lower than that of season-based models, and R^2 greater than 0.008——0.45. The most significant improvement in accuracy is observed in August. Monthly-based RF models are 0.89——1.88 °C, which are 0.01——0.25 °C lower than that of season-based models in all months except February, and R^2 increased by 0.005——0.081 °C. The greatest improvement in accuracy is observed in June. The season-based RF model performed better than monthly-based RF in February.

Table 4
Overall training accuracies.

Season	Month	Season-based MLR		Season-based RF		Monthly-based MLR		Monthly-based RF	
		RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Warm	May	2.57	0.749	0.68	0.982	2.56	0.67	0.64	0.98
	Jun					2.63	0.427	0.80	0.926
	Jul					2.16	0.397	0.71	0.935
	Aug					1.71	0.599	0.529	0.962
Cold	Feb	3.13	0.733	0.69	0.97	3.77	0.465	0.81	0.974
	Mar					4.18	0.567	0.85	0.982
	Apr					2.66	0.688	0.55	0.987
	Sep					1.70	0.702	0.49	0.975
	Oct					1.82	0.525	0.5	0.964

Table 5

Overall testing accuracies.

Season	Month	Season-based MLR		Season-based RF		Monthly-based MLR		Monthly-based RF	
		RMSE	R²	RMSE	R²	RMSE	R²	RMSE	R²
Warm	May	2.73	0.631	1.51	0.869	2.35	0.68	**1.421**	**0.883**
	Jun	2.86	0.325	1.84	0.689	2.46	0.44	**1.761**	**0.714**
	Jul	2.28	0.331	1.6	0.73	2.03	0.448	**1.504**	**0.755**
	Aug	2.42	0.198	1.12	0.802	1.41	0.649	**1.021**	**0.815**
	Feb	3.42	0.517	**1.66**	**0.886**	3.32	0.544	1.81	0.862
	Mar	4.03	0.583	1.89	0.907	4.03	0.591	**1.88**	**0.91**
Cold	Apr	2.82	0.621	1.29	0.921	2.69	0.656	**1.25**	**0.926**
	Sep	1.56	0.724	0.95	0.898	1.46	0.759	**0.89**	**0.91**
	Oct	1.84	0.513	1.25	0.774	1.7	0.585	**1**	**0.855**

4.3. Accuracy evaluation in different land covers

Fig. 6 shows the accuracies comparison in different land covers on testing datasets (detailed information is shown in Table A. 2——6). Comparison of different models results show that RF fed with LST and multiple environment factors perform better than MLR based solely on LST, with the varying RMSE reductions in different land covers. From the comparison results of the monthly-based models (orange and green bars), the RMSEs by monthly-based RF models results are 0.41——2.33 °C in all land covers, which are much lower than MLR models (RMSE between 0.62 and 9.22 °C). Notably, the largest reduction in RMSE is observed in barrens, grassy woodlands areas and permanent wetlands, with average reductions in RMSEs of 2.65 °C, 1.13 °C and 1.07 °C, respectively. The RMSEs in these areas decrease more

significantly in the cold season, with average reductions in RMSEs of 3.72 °C, 1.36 °C and 1.11 °C, respectively. The smallest reduction in RMSE is observed in the shrublands area, with an average decrease of 0.84 °C. In forests areas, the decreases in RMSEs are more pronounced in warm season compared to cold season, especially in June and August.

The results of monthly-based models are better than those of season-based models, with RMSEs reduction varying among different land cover. For the MLR models, the RMSEs of monthly-based models (shown in orange bars) are 0.62——9.22 °C in all land covers, which are 1.15——9.88 °C lower than those of season-based models (shown in red bars). The areas with the most significant decrease are permanent wetlands, barrens and shrublands, with average RMSEs decrease of 0.36 °C, 0.28 °C and 0.277 °C, respectively. For the RF models, the accuracy of monthly-based models (shown in green bars) also shows improvement
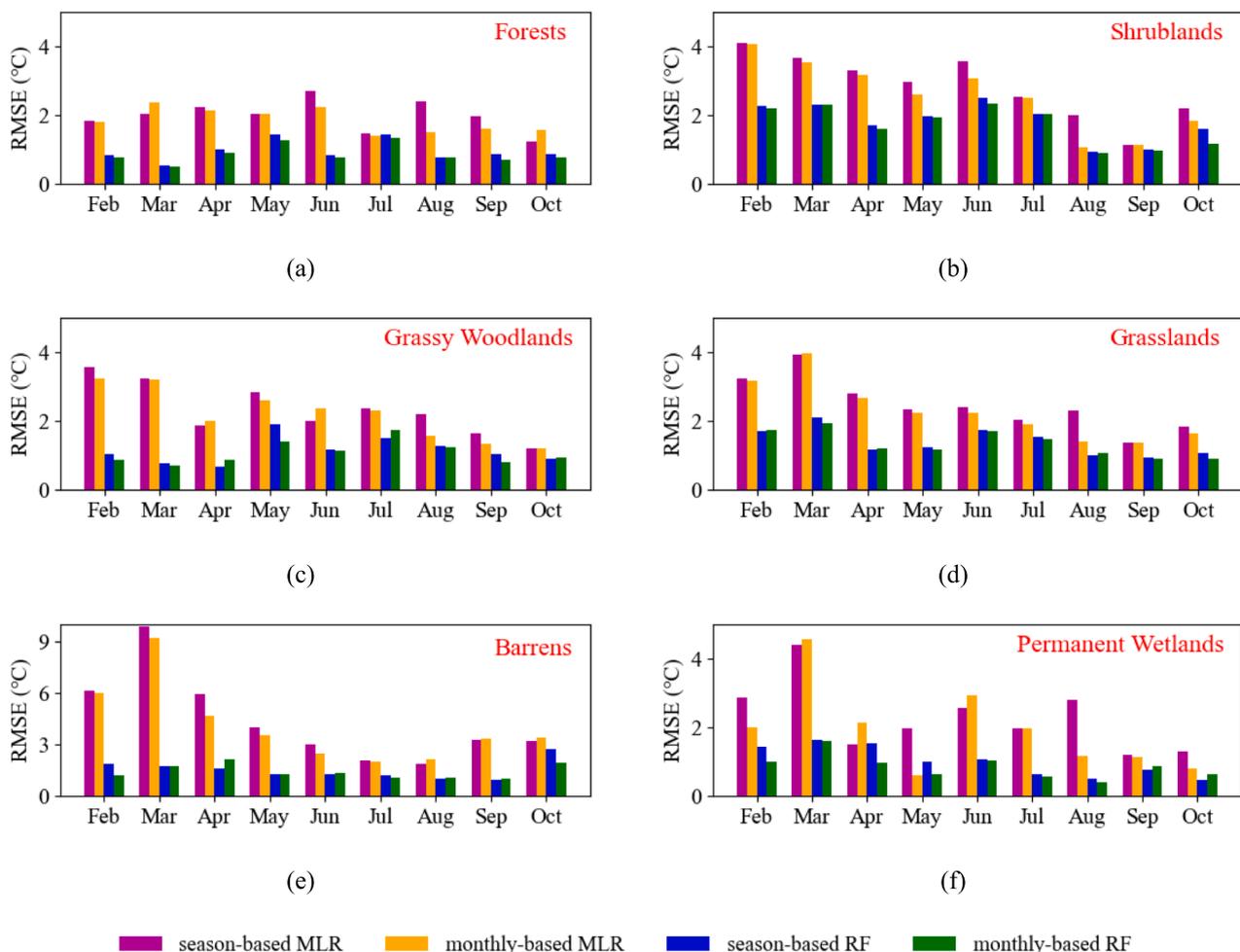

Fig. 6. Accuracies in different land covers.

compared with season-based models (shown in blue bars). The RMSEs of monthly-based models in all land covers are 0.41 – 2.33 °C, which are lower than those of season-based models (0.49——2.78 °C). The most significant improvement in accuracy is observed in the shrublands area, with an average RMSE reduction of 0.102 °C.

5. Discussion

5.1. Advantages of incorporating multiple environment factors to SST estimation

Estimating the mean SST over a certain period from MODIS LST four instantaneous data has always been a challenge (Ran and Li, 2019). The widely used method, which establishes MLR models between SST and four instantaneous LST data, neglects the influence of multiple environment factors. The method proposed in this study significantly improves pan-Arctic monthly SST estimation accuracy by incorporating multiple environment factors. The reason is that MODIS LST is usually the temperature at the top of vegetation or snow, which leads to the biases between LST and SST. The environment factors selected in the study (vegetation, snow, soil composition, topography, and solar radiation) all contribute to SST estimation modeling (Fig. 4). Incorporating these environment factor data into models helps improve accuracy.

Among the selected environment factors, vegetation and snow data are identified as the most important environment factors to the models (Fig. 4), indicating that vegetation and snow are critical environment factors in shaping biases between SST and LST. In warm season, vegetation cover is prevalent across many regions. A large part of solar shortwave radiation is intercepted by canopy and is used for transpiration. Only a small part of solar shortwave radiation reaches the ground, which results in lower soil temperatures (Heijmans et al., 2022). The height and coverage of vegetation can exert a substantial influence on the temperature biases (Song et al., 2021; Zou et al., 2017). Satellite observations measure the temperature at the top of the canopy, and the influences of vegetation cause significant bias between the satellite observed LST and the SST, especially in forests and shrublands with taller vegetation. Establishing the relationship between LST and SST by MLR without considering the vegetation influences can lead to significant errors. Incorporating vegetation data into the monthly-based RF models significantly contributes to SST estimation in warm season, indicating the ability of RF models to assimilate vegetation's impact on SST, thereby achieving an accuracy improvement in the warm season. In the cold season, persistent snowpack is common due to the lower temperatures in pan-Arctic. SST is simultaneously affected by vegetation and snow. The snow can insulate against heat loss acting as a blanket, resulting in SST warmer than temperature at the top of the snow. Differences in snow depth and density cause spatial variations in insulating effects (Peng et al., 2024), leading to higher RMSE in SST estimation in the cold season when using MLR based solely on LST. Different vegetation types can change the density and depth of surface snow, leading varying impacts on SST (Heijmans et al., 2022). In forest areas, snow is intercepted by the canopy, which reduces the snow impacts on SST. This may explain why RMSEs in forest areas are lower in cold season (Fig. 6). Incorporating snow data into RF models significantly contributes to the SST estimation in the cold season, demonstrating the models' capability to incorporate snow's influence on SST, which leads to a substantial improvement in accuracy in the cold season.

5.2. Advantages of monthly-based modeling

Monthly-based modeling demonstrates superior accuracy over season-based modeling across all months and various land covers. This is attributed to the monthly variability of environment factors like vegetation and snow, which can significantly influence SST. Season-based modeling may obscure these monthly variations, which can lead to inaccuracies.

In the warm season, vegetation data are the most important for modeling. The vegetation height and coverage change between different months, which may introduce various impacts on SST. Data such as LAI and NDVI exhibit monthly variations with vegetation growing, causing their importance to the models varying among different months. Vegetation is commonly vigorous in June and July, which may explain the higher importance of vegetation data to models than in other months. In the cold season, snow data are more contributive to the models. The NDSI snow cover, snow depth, and snow density data change between different months with the snowfall increasing, leading to significant variations in their contribution to the models. This could explain why the importance of snow to the model in March is higher than in February. The widely used method of season-based modeling only takes into account the impacts of environment factors change between seasons. In contrast, monthly-based modeling takes into account the impacts of environment factor change within months, leading to improved SST estimation accuracy in all months and land covers.

5.3. Limits and future improvements

Although incorporating these environment factor data by RF models is helpful to improve accuracy, the models are prone to overfitting. In the study, the close accuracies between training and testing for the majority months indicate that the majority models have not been overfitted (Table 5 and Table A6). However, for a few months, such as June and July, the larger discrepancies between the training and testing accuracies imply slight overfitting. The biases between training and testing datasets may be caused by the imbalanced distribution of samples across different land cover types (Table 2). In the training datasets, the proportion of grasslands samples is significantly higher than that of barrens and permanent wetlands. The models trained based on the training datasets may be more suited for grasslands. Consequently, when these models are used to predict on testing datasets, the accuracies for barrens and permanent wetland are limited, resulting in higher RMSEs. Similarly, the imbalanced distribution of samples within the same land cover type, such as differences in vegetation height and more detailed vegetation types between training and testing datasets, can also lead to biases between training and testing accuracies. Actually, it is quite normal and inevitable that the training accuracies are better than testing accuracies for machine learning algorithms. Models' accuracies are determined by testing accuracies. Improvements in testing accuracies are sufficient to demonstrate the advantages of the proposed method. In the future, with a more balanced sample distribution and more environment factor data (such as vegetation height data) integrated, the biases between training and testing accuracies are expected to be reduced.

The MODIS LST products used in the study (MOD11A1 and MYD11A1) contain missing data and unreliable measurements due to the influences of clouds and the atmosphere (Ahmed et al., 2023). Although compositing data by monthly averaging can mitigate these impacts, it still leads to deviations from actual monthly mean values, especially in pixels with a large amount of missing data within the month. This may limit the accuracy. In addition, the missing data in SST are filled with the average values of adjacent months, which can also introduce discrepancies compared with the actual results. Using high-quality MODIS LST data obtained by atmospheric correction and missing value filling algorithms in future research will help to further improve the accuracy of SST estimation models.

The accuracy of SST estimation from MODIS LST can be improved by incorporating multiple environment factors, and the method can still be improved. The influences of multiple above-ground environment factors on SST are taken into account. However, the subsurface heat conduction process can also affect the SST. The heat from deeper soil layers can be transmitted to the surface soil, and this process is influenced by factors such as the soil's thermal conductivity and heat capacity. Studying the subsurface heat conduction process will likely lead to additional

refinements in SST estimation accuracy.

In addition, there is a scale mismatch between site observed data and environment factor data values from 1 km resolution. Soil surface temperature exhibits spatial heterogeneity, especially in areas like permanent wetlands, where a single pixel contains both water bodies and land. Representing the SST of a pixel with only a few sites located in it can introduce biases. However, with the improvement of the spatial resolution of remote sensing data, the densification of ground observation stations, and the development of new data interpolation algorithms, the scale mismatch is expected to be improved in the future (Xu et al., 2020).

6. Conclusions

To improve the soil surface temperature (SST) estimation accuracy in the pan-Arctic, the contributions of five environment factors (vegetation, snow, surface soil composition, topography, and solar radiation) to SST estimation are analyzed, and an SST estimation method from MODIS land surface temperature (LST) by incorporating environment factor data and monthly-based modeling based on the RF machine learning algorithm is proposed. The results show that vegetation exerts the most substantial influence on SST estimation from May to October, and snow has the greatest influence in March and April. The proposed method achieves RMSEs of 0.89——1.88 °C for SST estimation from 2003 to 2022, significantly outperforming the widely used MLR method based solely on LST and season-based modeling (RMSE between 1.56 and 4.03 °C). For areas with high vegetation, such as forests, the proposed method significantly improves the accuracy in the warm season, with RMSE reduced from 1.4 − 2.25 °C to 0.78——1.33 °C. For areas with sparse or no vegetation, such as grassy woodlands, grasslands, permanent wetlands, and barren, the proposed method significantly improves the accuracy in the cold season, with RMSE reduced from 0.8 − 9.22 °C to 0.65——2.17 °C. The proposed method has potentials for producing high-precision SST data in pan-Arctic, supporting the estimation of permafrost extent and active layer thickness, and deepening understanding of the stability and dynamic changes of cryosphere environment elements, including climate and ecological hydrology in pan-Arctic.

CRediT authorship contribution statement

Hongxiang Guo: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Wenquan Zhu:** Supervision, Funding acquisition. **Cunde Xiao:** Writing – original draft, Supervision, Conceptualization. **Cenliang Zhao:** Writing – review & editing, Validation. **Liyuan Chen:** Writing – review & editing, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The GTN-P dataset is available from https://gtnp.arcticportal.org/. SoilTemp site observed data are available from https://zenodo.org/records/4558663. The MODIS production, ERA5-Land data and ArcticDEM are available from google earth engine (https://earthengine.google.com/). The SoilGrid250 data are available from https://soilgrids.org/.

Acknowledgment

Appendix A

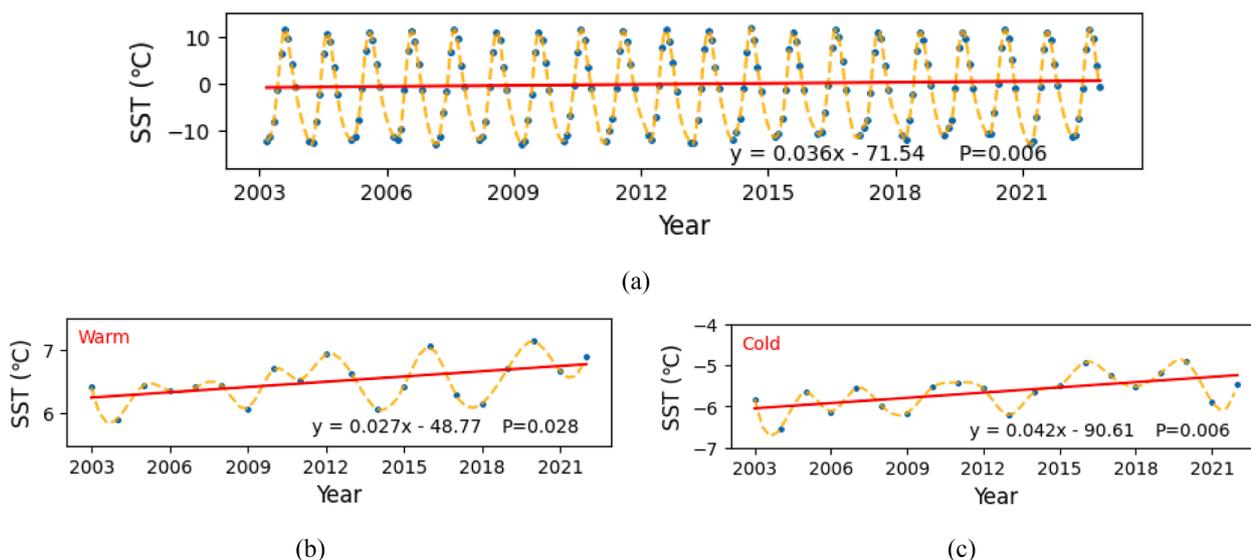
(a)

(b)   (c)

Fig. A1. SST trends in pan-Arctic by linear regression (a is the monthly SST trends, b is SST trends for the warm season, c is SST trends for the cold season).
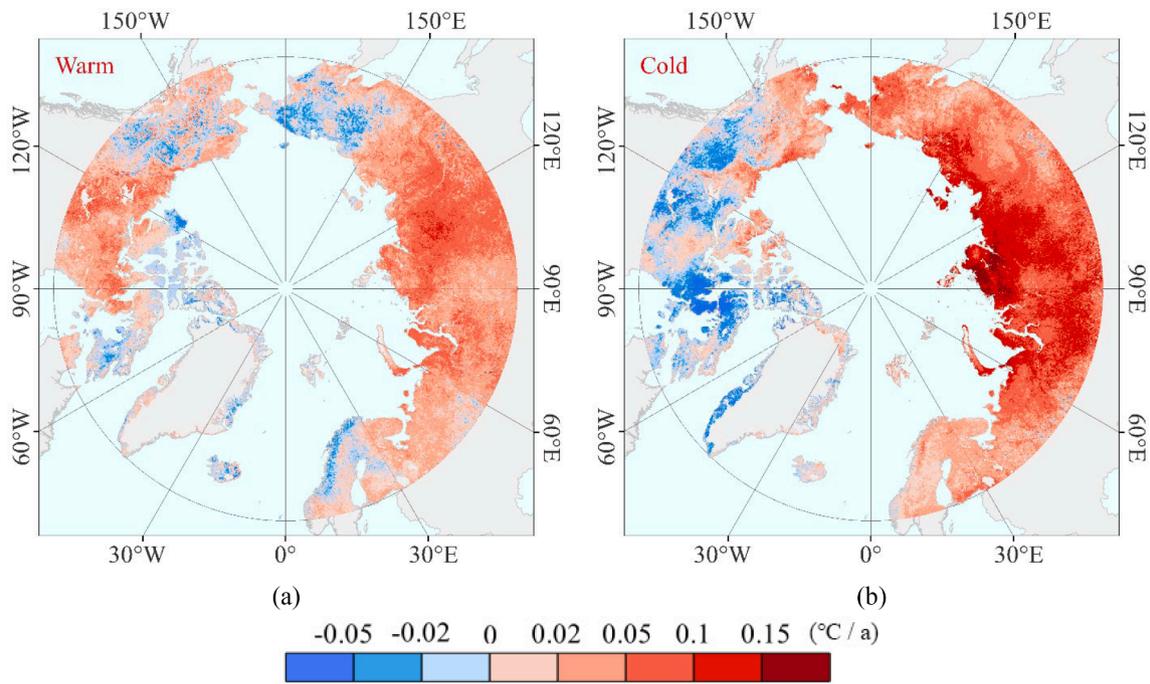
Fig. A2. SST trends in pan-Arctic by Mann-Kendall analysis (a is the SST trend for warm season, b is the SST trend for cold season).

Table A1
Details of the hyperparameter and training process for monthly-based RF models.

Month	Sample numbers	n_estimators	max_features	10-fold CrossValidate RMSE
Feb	929	231	12	2.29
Mar	916	86	4	2.38
Apr	894	191	3	1.96
May	910	176	2	1.73
Jun	984	246	3	2.09
Jul	976	86	5	1.92
Aug	1003	106	8	1.35
Sep	881	71	11	1.37
Oct	870	246	13	1.32
Cold	4490	246	11	1.98
Warm	3873	101	6	2.12

Table A2
Test accuracies of the season-based MLR models in different land covers.

	Forests		Shrublands		Grassy Woodlands		Grasslands		Barrens		Permanent Wetlands	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
May	2.05	0.338	2.98	0.278	2.83	−0.23	2.34	0.55	3.98	0.558	1.97	−3.299
Jun	2.69	0.235	3.58	−1.335	1.99	0.331	2.42	0.243	2.99	0.065	2.58	−0.364
Jul	1.46	0.45	2.53	0.269	2.36	−0.1	2.04	0.344	2.09	0.262	1.98	0.308
Aug	2.39	−2.18	1.99	−0.36	2.2	−0.268	2.29	−0.218	1.88	0.292	2.81	−5.151
Feb	1.85	−1.42	4.09	0.488	3.57	−9.093	3.25	0.451	6.15	−0.6	2.87	0.649
Mar	2.03	−5.34	3.68	0.594	3.25	−5.615	3.95	0.482	9.88	0.197	4.41	−22.34
Apr	2.23	−0.62	3.31	0.569	1.88	−0.484	2.8	0.589	5.95	−2.315	1.51	−9.285
Sep	1.98	−1.14	1.15	0.756	1.64	0.151	1.39	0.659	3.29	−0.199	1.21	0.324
Oct	1.23	0.4	2.21	−0.255	1.22	0.593	1.83	0.004	3.22	0.387	1.31	−1.83

Table A3

Test accuracies of the monthly-based MLR models in different land covers.

	Forests		Shrublands		Grassy Woodlands		Grasslands		Barrens		Permanent Wetlands	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
May	2.04	0.344	2.62	0.442	2.6	−0.044	2.23	0.592	3.54	0.651	0.62	0.614
Jun	2.25	−0.359	3.07	−0.714	2.37	0.047	2.24	0.351	2.48	0.355	2.94	−1.089
Jul	1.4	0.504	2.49	0.295	2.31	−0.052	1.92	0.424	2.01	0.318	1.97	0.33
Aug	1.5	−0.25	1.06	0.612	1.58	0.347	1.4	0.544	2.14	0.077	1.19	−4.757
Feb	1.8	−1.314	4.08	0.491	3.22	−7.218	3.16	0.479	6.01	−0.551	2.01	0.829
Mar	2.38	−7.73	3.55	0.624	3.19	−5.385	3.97	0.476	9.22	0.302	4.58	−24.18
Apr	2.13	−0.48	3.16	0.605	2.01	−0.689	2.66	0.628	4.71	−1.071	2.13	−19.44
Sep	1.61	−0.42	1.14	0.757	1.35	0.424	1.37	0.668	3.36	−0.247	1.14	0.401
Oct	1.56	0.028	1.85	0.117	1.22	0.59	1.63	0.21	3.41	0.313	0.8	−0.062

Table A4

Accuracies of the season-based RF models in different land covers.

	Forests		Shrublands		Grassy Woodlands		Grasslands		Barrens		Permanent Wetlands	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
May	1.43	0.678	1.98	0.682	1.89	0.452	1.24	0.869	1.32	0.951	1.02	0.887
Jun	0.83	0.815	2.51	−0.144	1.18	0.763	1.75	0.606	1.27	0.832	1.09	0.748
Jul	1.44	0.456	2.05	0.521	1.49	0.562	1.53	0.624	1.23	0.744	0.63	0.947
Aug	0.79	0.655	0.95	0.692	1.26	0.584	1.01	0.766	1	0.799	0.51	0.901
Feb	0.85	0.485	2.26	0.844	1.05	0.128	1.71	0.847	1.89	0.85	1.44	0.912
Mar	0.53	0.607	2.31	0.84	0.78	0.614	2.1	0.85	1.78	0.974	1.65	0.902
Apr	1	0.672	1.7	0.887	0.69	0.8	1.17	0.925	1.64	0.749	1.53	0.891
Sep	0.88	0.579	1.01	0.809	1.05	0.656	0.93	0.847	0.94	0.903	0.77	0.886
Oct	0.89	0.688	1.62	0.329	0.9	0.776	1.07	0.66	2.78	0.543	0.49	0.846

Table A5

Accuracies of the monthly-based RF models in different land covers.

	Forests		Shrublands		Grassy Woodlands		Grasslands		Barrens		Permanent Wetlands	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
May	1.28	0.743	1.94	0.693	1.41	0.695	1.16	0.886	1.28	0.954	0.64	0.936
Jun	0.79	0.831	2.33	0.012	1.13	0.783	1.72	0.622	1.37	0.804	1.03	0.772
Jul	1.33	0.53	2.03	0.53	1.73	0.415	1.48	0.648	1.12	0.79	0.59	0.956
Aug	0.78	0.663	0.9	0.723	1.25	0.58	1.09	0.726	1.12	0.75	0.41	0.911
Feb	0.79	0.552	2.22	0.849	0.86	0.417	1.75	0.84	1.19	0.941	1.02	0.956
Mar	0.5	0.618	2.29	0.843	0.71	0.621	1.95	0.87	1.78	0.974	1.59	0.919
Apr	0.92	0.725	1.61	0.898	0.88	0.678	1.2	0.922	2.17	0.561	0.98	0.955
Sep	0.7	0.734	0.98	0.821	0.8	0.801	0.92	0.849	1	0.89	0.87	0.856
Oct	0.78	0.758	1.17	0.646	0.94	0.756	0.91	0.754	1.98	0.77	0.65	0.721

References

Ahmed, M.R., Ghaderpour, E., Gupta, A., et al., 2023. Opportunities and Challenges of Spaceborne Sensors in Delineating Land Surface Temperature Trends: A Review. A Review. IEEE Sens. J. 23 (7), 6460–6472. https://doi.org/10.1109/JSEN.2023.3246842.

AMAP, 2017. Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017. Arctic Monitoring and Assessment Programme (AMAP), Oslo, Norway, p. xiv + 269 pp.

AMAP, 2021. Arctic Climate Change Update 2021: Key Trends and Impacts. Summary for Policy-makers. Arctic Monitoring and Assessment Programme (AMAP), Tromsø, Norway, p. 16 pp.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Biskaborn, B.K., Lanckman, J.P., Lantuit, H., et al., 2015. The new database of the Global Terrestrial Network for Permafrost (GTN-P). Earth Syst. Sci. Data. 7 (2), 245–259. https://doi.org/10.5194/essd-7-245-2015.

Biskaborn, B.K., Smith, S.L., Noetzli, J., et al., 2019. Permafrost is warming at a global scale. Nat. Commun. 10 (1), 264. https://doi.org/10.1038/s41467-018-08240-4.

Brown, J., Ferrians, O.J., Heginbottom, J.A., Melnikov, E.S., 2002. Circum-Arctic map of permafrost and ground-ice conditions. Version 2.

Cai, H.Y., Han, D.R., Yang, L.S., et al., 2020. Spatiotemporal change in permafrost active layer thickness in the Pan-Arctic region. Natl. Remote Sens. Bull. 24 (08), 1045–1057.

Cao, H., Gao, B., Gong, T., Wang, B., 2021. Analyzing changes in frozen soil in the source region of the yellow river using the MODIS land surface temperature products. Remote Sens. 13 (2), 180. https://doi.org/10.3390/rs13020180.

Friedl, M., Sulla-Menashe, D., 2022. MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V061. V061.

Hachem, S., Duguay, C., Allard, M., 2012. Comparison of MODIS-derived land surface temperatures with ground surface and air temperature measurements in continuous permafrost terrain. The Cryosphere. 6 (1), 51–69. https://doi.org/10.5194/tc-6-51-2012.

Heijmans, M.M.P.D., Magnússon, R.Í., Lara, M.J., et al., 2022. Tundra vegetation change and impacts on permafrost. Nat. Rev. Earth Environ. 3 (1), 68–84. https://doi.org/10.1038/s43017-021-00233-0.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., et al., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One. 12 (2), e0169748.

Hersbach, H., Bell, B., Berrisford, P., et al., 2017. Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate. . doi: 10.24381/cds.143582cf.

Hori, M., Sugiura, K., Kobayashi, K., et al., 2017. A 38-year (1978–2015) Northern Hemisphere daily snow cover extent product derived using consistent objective criteria from satellite-borne optical sensors. Remote Sens. Environ. 191, 402–418. https://doi.org/10.1016/j.rse.2017.01.023.

Ipcc, 2022. Climate Change 2022: Impacts, Adaptation, and Vulnerability. In: Pörtner, H.-.-O., Roberts, D.C., Tignor, M. (Eds.), Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Kurylyk, B.L., Hayashi, M., 2016. Improved Stefan equation correction factors to accommodate sensible heat storage during soil freezing or thawing. Permafr. Periglac. 27 (2), 189–203. https://doi.org/10.1002/ppp.1865.

Lembrechts, J.J., Aalto, J., Ashcroft, M.B., et al., 2020. SoilTemp: A global database of near-surface temperature. Glob. Change Biol. 26 (11), 6616–6629. https://doi.org/10.1111/gcb.15123.

Lembrechts, J.J., van den Hoogen, J., Aalto, J., et al., 2022. Global maps of soil temperature. Glob. Change Biol. 28 (9), 3110–3144. https://doi.org/10.1111/gcb.16060.

Luo, D.L., Wu, Q.B., Jin, H.J., et al., 2016. Recent changes in the active layer thickness across the northern hemisphere. Environ. Earth Sci. 75 (7), 555. https://doi.org/10.1007/s12665-015-5229-2.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., et al., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Environ. Earth Sci. Data. 13 (9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.

Myhra, K., Westermann, S., Etzelmüller, B., 2017. Modelled distribution and temporal evolution of permafrost in steep rock walls along a latitudinal transect in Norway by CryoGrid 2D. Permafr. Periglac. Process. 28 (1), 172–182. https://doi.org/10.1002/ppp.1884.

Ni, J., Wu, T.H., Zhu, X.F., 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 (2) https://doi.org/10.1029/2020JD033402 e2020JD033402.

Niclòs, R., Perelló, M., Puchades, J., et al., 2023. Evaluating Landsat-9 TIRS-2 calibrations and land surface temperature retrievals against ground measurements using multi-instrument spatial and temporal sampling along transects. Int. J. Appl. Earth Obs. Geoinf. 125, 103576 https://doi.org/10.1016/j.jag.2023.103576.

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. https://doi.org/10.1016/j.earscirev.2019.04.023.

Park, H., Kim, Y., Kimball, J.S., 2016. Widespread permafrost vulnerability and soil active layer increases over the high northern latitudes inferred from satellite remote sensing and process model assessments. Remote Sens. Environ. 175, 349–358. https://doi.org/10.1016/j.rse.2015.12.046.

Peng, X., Frauenfeld, O.W., Huang, Y., et al., 2024. The thermal effect of snow cover on ground surface temperature in the Northern Hemisphere. Environ. Res. Lett. 19 (4), 044015 https://doi.org/10.1088/1748-9326/ad30a5.

Ran, Y.H., Li, X., 2019. Progress, Challenges and Opportunities of Permafrost Mapping in China. Adv. Earth Sci. 34 (10), 1015–1027.

Reiners, P., Sobrino, J., Kuenzer, C., 2023. Satellite-derived land surface temperature dynamics in the context of global change—a review. Remote Sens. 15 (7), 1857. https://doi.org/10.3390/rs15071857.

Setianto, A., Triandini, T., 2013. Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. J. Appl. Geol. 5 (1), 2502–2822. https://doi.org/10.22146/jag.7204.

Smith, S.L., O'Neill, H.B., Isaksen, K., et al., 2022. The changing thermal state of permafrost. Nat. Rev. Earth Environ. 3 (1), 10–23. https://doi.org/10.1038/s43017-021-00240-1.

Song, Z., Yang, H., Huang, X., et al., 2021. The spatiotemporal pattern and influencing factors of land surface temperature change in China from 2003 to 2019. Int. J. Appl. Earth Obs. Geoinf. 104, 102537 https://doi.org/10.1016/j.jag.2021.102537.

Tassi, A., Gigante, D., Modica, G., et al., 2021. Pixel- vs. Object-Based Landsat 8 data classification in Google Earth Engine using random forest: The case study of Maiella National Park. Remote Sens. 13 (12), 2299. https://doi.org/10.3390/rs13122299.

Wan, Z., Hook, S., Hulley, G., 2021a. MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 1km SIN. Grid V061. V61.

Wan, Z., Hook, S., Hulley, G., 2021b. MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN. Grid V061. V61.

Wang, D., 2024. MODIS/Terra+Aqua Surface Radiation Daily/3-Hour L3 Global 1km SIN Grid V062. V62.

Westermann, S., Duguay, C.R., Grosse, G., Kääb, A., 2015. Remote sensing of permafrost and frozen ground. Remote Sensing of the Cryosphere 307–344. https://doi.org/10.1002/9781118368909.ch13.

Xu, C., Qu, J.J., Hao, X., et al., 2020. Surface soil temperature seasonal variation estimation in a forested area using combined satellite observations and in-situ measurements. Int. J. Appl. Earth Obs. Geoinf. 91, 102156 https://doi.org/10.1016/j.jag.2020.102156.

Yuan, H., Dai, Y., Xiao, Z., et al., 2011. Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling. Remote Sens. Environ. 115 (5), 1171–1187. https://doi.org/10.1016/j.rse.2011.01.001.

Zou, D.F., Zhao, L., Wu, T.H., et al., 2014. Modeling ground surface temperature by means of remote sensing data in high-altitude areas: test in the central Tibetan Plateau with application of moderate-resolution imaging spectroradiometer Terra/Aqua land surface temperature and ground-based infrared radiometer. J. Appl. Remote Sens. 8 (1), 083516. https://doi.org/10.1117/1.JRS.8.083516.

Zou, D.F., Zhao, L., Sheng, Y., et al., 2017. A new map of permafrost distribution on the Tibetan Plateau. The Cryosphere. 11 (6), 2527–2542. https://doi.org/10.5194/tc-11-2527-2017.