Constrained simulation of permafrost thermal changes from 1980 to 2018 on the Qinghai-Tibet Plateau

Hailong Ji[1,2], Xiaobo Wu[3], Shuping Zhao[2], Zhuotong Nan[1,2,4*]

[1] Yangtze River Delta Urban Wetland Ecosystem National Field Scientific Observation and Research Station, School of Environmental and Geographical Sciences, Shanghai Normal University, Shanghai, 200234, China

[2] State Key Laboratory of Climate System Prediction and Risk Management, Nanjing Normal University, Nanjing, 210023, China

[3] College of Resources, Sichuan Agricultural University, Chengdu, 611130, China

[4] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China

*Correspondent author: Zhuotong Nan, nanzt@shnu.edu.cn

Ji H, Wu X, Zhao S, Nan Z*. Constrained simulation of permafrost thermal changes from 1980 to 2018 on the Qinghai-Tibet Plateau. Global and Planetary Change. 2026. https://doi.org/10.1016/j.gloplacha.2026.105542.

Abstract

Land surface models (LSMs) are widely used to assess regional permafrost changes, but suffer from considerable uncertainty, partly due to inappropriate model constraints. In this study, we developed a map-based calibration that utilized a high-fidelity permafrost map to constrain a modified Noah LSM on the Qinghai-Tibet Plateau (QTP), as an alternative to the conventional constraint by in-situ observations (termed site-based calibration). Validation against the permafrost observation network across the QTP demonstrated the new method's superior performance. Specifically, the root mean square error (RMSE) is 0.68 m for active layer thickness (ALT), 0.41 °C for temperature at the top of permafrost (TTOP), and 1.30 °C for ground temperature at the depth of 10 m (GT10m), and 0.08 m/10a, 0.36 °C/10a, 0.64 °C/10a for corresponding change rates. The effect of site-based calibration highly depends on the calibration site, sometimes yielding results inferior to the default parameter. Based on the optimized parameter ensemble from map-based calibration, our simulation results showed a contraction of the permafrost extent on the QTP from $1.262\pm0.048\times10^6$ km^2 in 1980 to $1.085\pm0.049\times10^6$ km^2 in 2018 at a rate of 44.4×10^3 km^2/10a, which falls toward the lower end of previous modeling estimates. Despite the pervasive permafrost thermal degradation, unexpected atmospheric cooling on the northwestern plateau in the late 1990s decelerated the overall trend. This study contributes an innovative calibration method to the permafrost modeling community and provides a more realistic permafrost change on the QTP.

Keywords

Permafrost change, Land surface model, Model calibration, Parametric ensemble, Qinghai-Tibet Plateau

1 Introduction

Permafrost, as a critical component of cold regions such as the pan-Arctic and the Qinghai-Tibet Plateau (QTP) (Brown et al., 1997), is well-recognized as being highly sensitive to climate change (Biskaborn et al., 2019). Over the past decades, global permafrost has undergone pervasive degradation, evidenced by rising permafrost temperature (Smith et al., 2022), lengthening thawing periods, deepening active layers, and the emergence of thermokarst landscapes (Luo et al., 2022). These changes substantially reshape local hydrological regimes (Wang et al., 2023) and ecological systems (Ehlers et al., 2022), potentially threatening adjacent residential areas (Yao et al., 2022) and even disrupting the global carbon budget (Mu et al., 2025).

Despite its significance, knowledge of regional permafrost changes is still limited, largely hindered by the challenges of observing subsurface conditions across vast, remote areas (Zhao et al., 2021). Process-based models, such as land surface models (LSMs), can produce spatiotemporally continuous profiles of soil temperature and moisture, addressing gaps in in-situ and remote-sensing observations. Process-based models are built on Fourier's and Darcy's laws, which are physically consistent with the real world (Riseborough et al., 2008; Smith et al., 2022). Nonetheless, the reliable application of the process-based models is often hampered by several factors, including potentially oversimplified representations of freezing-thawing processes (Aga et al., 2023; Li et al., 2020), inappropriate parameterization schemes (Ji et al., 2024; Zhao et al., 2023), uncertainties in prescribing initial conditions (Elshamy et al., 2020; Ji et al., 2022), and inaccuracies in meteorological forcing data (Abdelhamed et al., 2023; Guo et al., 2017; Tang et al., 2023).

Through decadal refinement, the performance of the process-based model has been improved progressively, but it still does not fully meet the needs of the permafrost study community. The persistent bottleneck can be attributed to the considerable spatial heterogeneity of the

subsurface, the lack of direct observation data, and the inherent complexity of soil thawing-freezing processes. However, the rapidly changing climate and permafrost urge us to improve the accuracy and reliability of the model as much as possible. Parameter calibration offers a practical and effective approach (Beven & Binley, 1992; Tsai et al., 2021), and often requires much less effort than model diagnosis and modification. Conventionally, model calibration utilizes in-situ time series of soil temperature and/or moisture content as the target (termed site-based calibration), with the optimized parameters subsequently applied to regional simulations (Gao et al., 2018; Jafarov et al., 2012; Jiang et al., 2024b). However, this "site calibration-regional simulation" paradigm has been criticized for the poor spatial transferability of parameters (Rosero et al., 2010). Its applicability is questionable and might be worse than the default in some regions, where vegetation, soil texture, and local topography, among other environmental factors, differ greatly from the calibration site. Despite the concern, there is no solid evidence to prevent modelers from improving the model's performance through site-based calibration.

Besides soil temperature and moisture, permafrost distribution map is another important piece of information on permafrost conditions (Biskaborn et al., 2019; Chadburn et al., 2017), which receives much less attention in model calibration. Compared to in-situ observations, the permafrost map has much coarser information granularity (i.e., permafrost presence or absence in a specific region) but is available across the study area. If permafrost map is used as the calibration target, the optimized parameters have to trade off across subregions to achieve the best overall performance across the entire calibration region. Therefore, it can effectively circumvent the parameter transferability issue encountered in the site-specific calibration. However, a permafrost map provides only categorical or binary data (i.e., presence or absence). The binary classification compresses the complex, multi-layered thermal regime into a single bit of information per grid cell. Using a binary map as a calibration target will inevitably lead

to much more severe equifinality than site-based calibration. While previous studies have demonstrated the effectiveness of map-based calibration, these efforts have typically involved semi-empirical models with very few parameters (e.g., Cao et al., 2023; Hu et al., 2020). To our knowledge, map-based constraints have rarely been applied to calibrate highly parameterized, process-based permafrost models, largely because the severe equifinality and computational burden imposed by binary spatial targets remain profound technical challenges (Abdelhamed et al., 2023; Hu et al., 2023). Furthermore, calibrating a process-based model with a permafrost map requires that the map is specific for a year, rather than a climatic average. However, most available permafrost maps are produced from a pool of observations spanning a long period to address the data scarcity. The recently published permafrost map by Cao et al. (2023), hereafter referred to as the Cao map, was produced specifically for the year 2010, and it provides a unique opportunity to explore this calibration strategy with process-based permafrost models.

In this study, we utilized a modified version of the Noah LSM to assess the effectiveness of different calibration methods. We further reconstructed the spatiotemporal patterns of QTP permafrost changes over the period 1980-2018 by optimized parameter-ensemble simulation. This study aims to provide reliable and robust historical permafrost evolution in the world's largest low-latitude permafrost region, and to contribute a novel constraint method for numerical permafrost modeling at the regional scale.

2 Materials and Methods

2.1 Study area and data

The QTP's elevation increases from southeast to northwest, dominating the spatial patterns of air temperature and permafrost. Permafrost is continuous in the plateau's hinterland (Figure 1), covering approximately 1.086×10^6 km^2 excluding glaciers and lakes (Cao et al., 2023). The

southeastern plateau, influenced by the monsoon, is relatively humid, contrasting with the arid northwestern part under the westerlies. This permafrost region supports diverse vegetation, mainly swamp meadow, alpine meadow, alpine steppe, and alpine desert (Wang et al., 2016), reflecting variations in water and energy availability. Despite severe physical weathering, the slow rate of soil formation on the QTP results in high gravel content. Over recent decades, the plateau has experienced significant climate change, with a mean annual air temperature (MAAT) of −1.93 °C, increasing by 0.39 °C per decade, and annual precipitation of about 400 mm, increasing by 34 mm per decade.

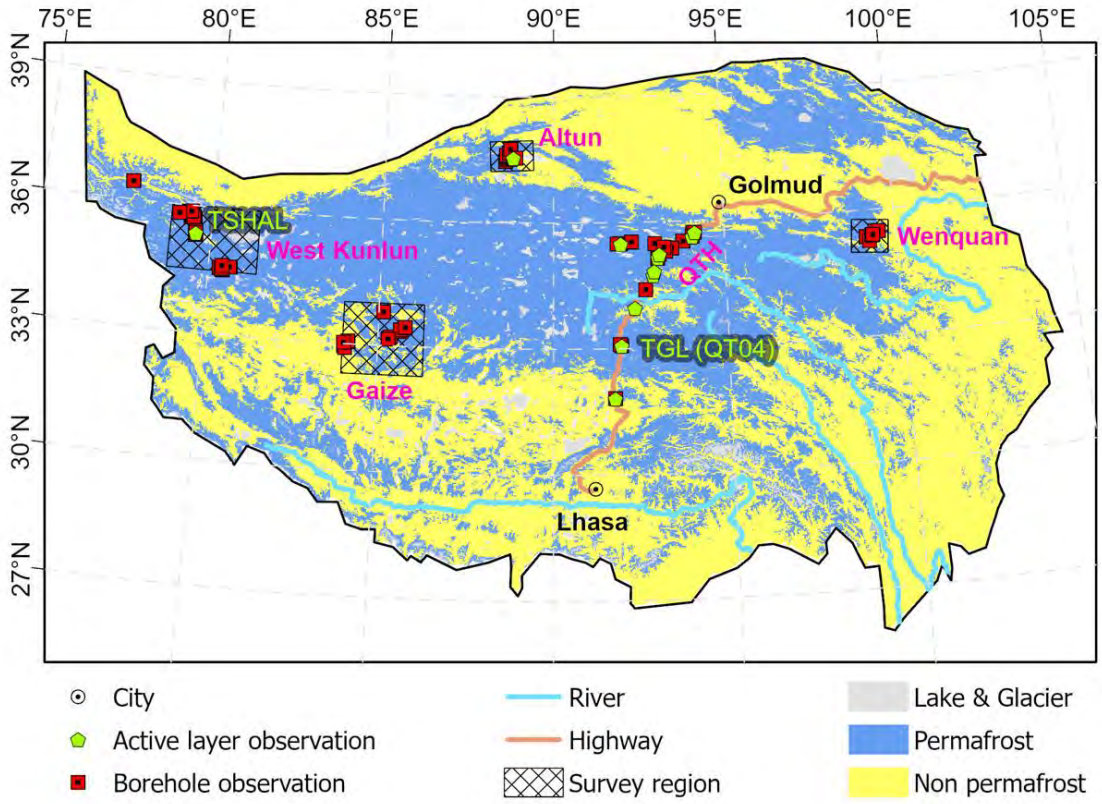

Figure 1 An overview of permafrost distribution and in-situ permafrost observation network on the Qinghai-Tibet Plateau (QTP). The permafrost observation network (Zhao et al., 2021) comprises 12 active-layer observation stations and 84 borehole observation stations, spanning 2004 to 2018; the TGL(QT04) and TSHAL stations were used to conduct two site-based model calibration experiments. The permafrost distribution is specific for the year 2010 (Cao et al., 2023).

The permafrost observation network across the QTP (Zhao et al., 2021) provides daily shallow soil temperature and moisture content at 12 active-layer monitoring stations and annual deep ground temperature at 84 boreholes. The observation record spans from 2004 to 2018, but the data availability at most sites is much shorter due to late establishment or instrument malfunction. The soil temperature is measured by thermocouple probe or thermistor, with an accuracy of ±0.1 °C to ±0.2 °C. The liquid soil moisture is measured by time/frequency domain reflectometer, with an accuracy of ±2.5% (volumetric water content). These monitoring sites are clustered in five survey regions: Wenquan, Qinghai-Tibet Highway (QTH), Altun, Gaize, and West Kunlun. These observations served for the modified Noah LSM's calibration and validation, as detailed in Sections 2.5 and 2.6.

The Cao map (Cao et al., 2023) was produced at a 1 km resolution using a modified surface frost number model, driven by gap-filled satellite-derived land surface temperature, and further constrained by multiple survey-based subregion permafrost maps. Validation against the borehole records showed that the Cao map achieved an overall accuracy of 0.85, higher than the previous maps. Given the highly constrained timing of input and validation data, the Cao map represents permafrost conditions specifically for the year 2010. It offers a snapshot in time rather than a climatological average, a characteristic absent from many other QTP permafrost maps.

The China meteorological forcing dataset from the Institute of Tibetan Plateau Research (ITP-forcing) (He et al., 2020) was used to drive the modified Noah LSM. This dataset includes essential meteorological variables such as near-surface air temperature, downward shortwave and longwave radiation, precipitation rate, specific humidity, surface pressure, and wind speed, provided at 0.1° spatial and 3-hour temporal resolutions for the period 1979-2018. The 1:1,000,000 Vegetation Atlas of China (Hou & Zhang, 2021) was used to identify vegetation type, while the climatic average of 1/12° leaf area index from the Global Inventory Modeling

and Mapping Studies (Zhu et al., 2013) was used to classify vegetation coverage as sparse or dense. The soil column configuration was based on a 1-km Multilayer Soil Texture Dataset (Wu & Nan, 2016), which was developed from soil survey profiles on the QTP. These finer-resolution land environmental datasets, as well as the Cao map, were first aggregated to a 0.1° spatial resolution through a majority aggregation method, followed by nearest neighbor resampling to align with the standard grid of the ITP-forcing data.

2.2 The modified Noah LSM

The parameterization schemes of Noah LSM version 3.4.1 (Ek et al., 2003) were modified to adapt the QTP's alpine environment characterized by sparse vegetation and coarse soil, and to account for the impermeable effect of ground ice on water migration (Chen et al., 2015; Wu et al., 2018) (Figure 2). To more reliably simulate permafrost changes over decadal timescales, the soil column stratification was extended from a homogeneous 4-layer (2.0 m) to a heterogeneous 18-layer (15.2 m) (Figure 2b). This extension exceeds the depth of zero annual amplitude (DZAA), allowing the model to capture the thermal and hydraulic memories of deep soil. Consequently, the lower boundary depth of the thermal field (ZBOT) was moved from 8 m to 40 m to meet the numerical stability requirement of the Crank-Nicolson difference scheme. The fixed ground temperature at ZBOT (TBOT) was estimated linearly from elevation, with coefficients derived from deep borehole observations (Chen et al., 2015).

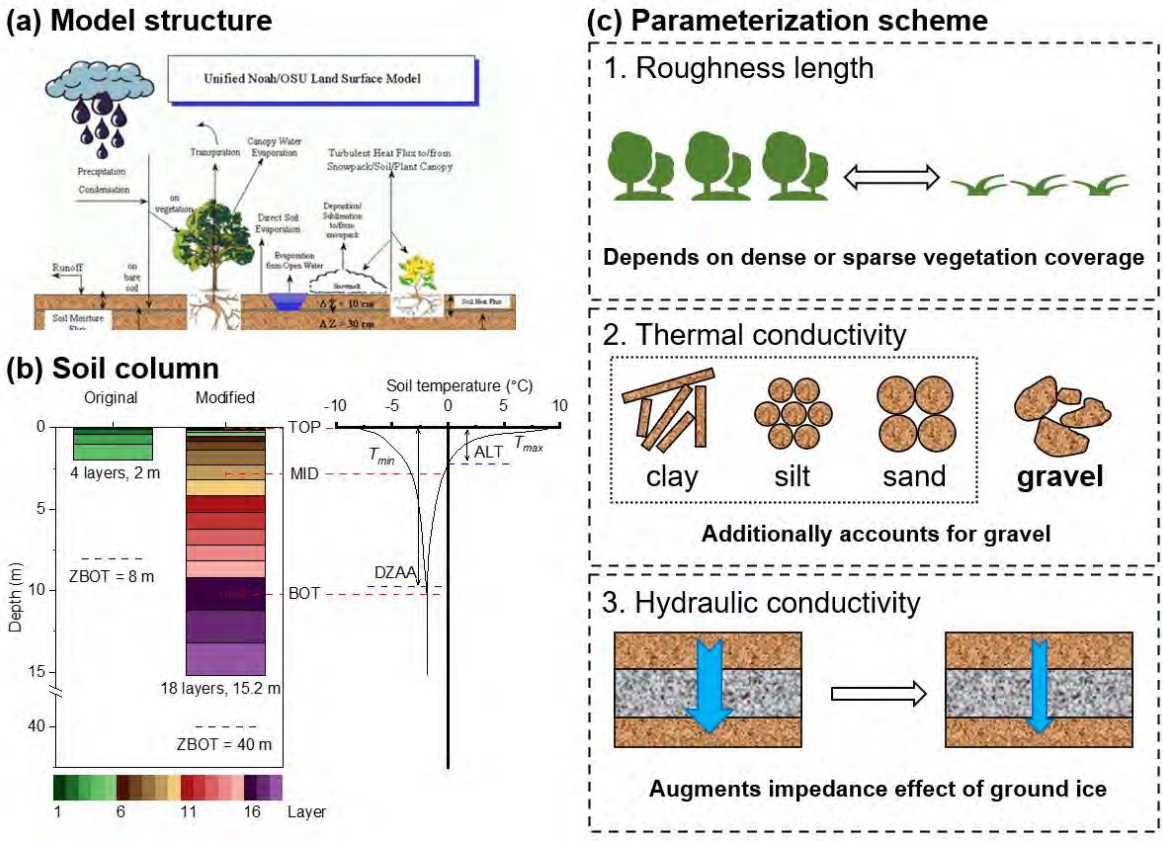

Figure 2 Diagram illustrating the modified Noah LSM for application on the QTP. The model retains the structure of the original Noah LSM (a) (copied from https://ral.ucar.edu/model/unified-noah-lsm) but incorporates adaptations for the alpine permafrost environment, detailed in Wu et al. (2018): (b) extension of the soil column to 15.2 m (copied from Ji et al. (2022)); and (c) modification of three parameterization schemes. ALT: active layer thickness; DZAA: depth of zero annual amplitude; Tmin and Tmax: minimum and maximum annual soil temperature; ZBOT: lower boundary depth of soil thermal field. TOP (0.1m), MID (2.7m) and BOT (10.2m) represent three selected depths used in sensitivity analysis.

Considering the physical processes represented by the modified Noah LSM and the available forcing data, simulations were performed at a spatial resolution of 0.1° (over 25,000 modeling cells covering the QTP) and a temporal resolution of 3 hours for the period 1979-2018. The first five years of the forcing data (1979-1983) were repeated for a 500-year spin-up to establish quasi-equilibrium initial conditions, followed by the main simulation driven by the actual climate data from 1979 to 2018 (Ji et al., 2022).

2.3 Parameter space

Parameters in both the original and modified Noah LSM can be categorized as either hard-coded ones, embedded in the source code and not recommended for adjustment (Cuntz et al., 2016; Ek et al., 2003), or tabulated ones, stored in lookup tables. This study focuses exclusively on the tabulated parameters. Parameters related to urban processes were excluded due to the negligible human activity on the QTP. Additionally, the lower boundary temperature (TBOT) was involved due to potential uncertainty from its initial estimation (Chen et al., 2015).

These tunable parameters comprise homogeneous (or universal) ones, which are spatially uniform values across all modeling cells, and heterogeneous ones, whose values depend on the vegetation type, soil texture, or elevation of the specific modeling cell (Table 1). Permissible ranges for universal parameters were determined using the Noah LSM user guide, parameterization studies (Chen et al., 1997; Schaake et al., 1996; Wetzel & Chang, 1988), and previous sensitivity analyses (Cuntz et al., 2016; Hu et al., 2023). Given the limited prior information, the permissible range of vegetation or soil-dependent heterogeneous parameters was set to not exceed the inter-class differences, specifically ±10% or ±20% around their default values. The vegetation root depth (NROOT), represented by the number of soil layers, was allowed to vary by ±2 around the default value, while ensuring it remained positive. Certain soil parameters (i.e., DRYSMC, REFSMC, SATDW, and WLTSMC) were updated passively based on the well-established relationships (Campbell, 1974; Wetzel & Chang, 1988). As for TBOT, ±1.5 °C was allowed to vary around the default value. In total, there are 30 parameters involved, with details in Table 1.

Table 1 Details of model parameters for calibration.

Group	Code	Default	Range[#]	Unit	Hetero.[&]	Module(s)[§]	Physical meaning	Reference(s)
Universal	SBETA[@]	−2	−3 – −1		No	Thermal	Vegetation canopy effect on ground heat flux, dependent on vegetation fraction.	User guide
	FXEXP [@]	2	0.2 – 4.0		No	Evapotranspiration	Bare soil evaporation exponent; FXEXP=1 yields linear reduction of bare soil evaporation with decreased soil moisture between MAXSMC and DRYSMC, while FXEXP>1 yields greater-than-linear reduction	User guide
	CSOIL	2.00E+06	1.26E+06 – 3.50 E+06	J $m^{-3}K^{-1}$	No	Thermal	Soil heat capacity	User guide
	SALP	2.6	2.5 – 4.0		No	Atmosphere-Land	Shape parameter used in function to infer percent area snow cover from snow depth	User guide
	REFDK	2.00E-06	Passive		No	Hydrology	Parameter used with REFKDT to determine surface runoff.	User guide
	REFKDT	3	0.5 – 5.0		No	Hydrology	Controls surface infiltration and the partitioning of total runoff into surface and subsurface flow; higher values decrease surface runoff.	User guide; Schaake et al. (1996)
	FRZK[@]	0.15	0.10 – 0.20		No	Hydrology	Base reference value for the frozen-soil freeze factor, representing the ice content threshold for soil impermeability.	User guide
	CZIL[@]	0.2	0.0 – 1.0		No	Atmosphere-Land	Zilintikevich parameter, controlling the ratio of thermal to momentum roughness length (Z0).	Chen et al. (1997)
	TOPT	298.0	293.0 – 303.0	K	No	Evapotranspiration	Optimum air temperature for canopy transpiration	User guide
	CMCMAX	5E−04	1E−04 – 2	m	No	Evapotranspiration	Maximum canopy water capacity	User guide
	CFACTER	0.5	0.40 – 0.95		No	Evapotranspiration	Exponent used in function for canopy water evaporation	User guide
	RSMAX	5000	2000 – 10000	s m^{-1}	No	Evapotranspiration	Maximum stomatal resistance	User guide
Vegetation	NROOT	4 – 8	±2		Yes	Evapotranspiration	Number of soil layers from the surface that plant roots reach.	User guide
	RS	40 – 400	±20%	s m^{-1}	Yes	Evapotranspiration	Minimal stomatal resistance	User guide

	RGL	30 – 100	±20%		Yes	Evapotranspiration	Radiation stress parameter	User guide
	HS	36 – 55	±20%		Yes	Evapotranspiration	Vapor pressure deficit coefficient	User guide
	SNUP	0.02 – 0.08	±20%	m	Yes	Atmosphere-Land	Water-equivalent snow depth threshold for 100% snow cover and maximum snow albedo.	User guide
	LAIMIN	0.1 – 3.0	±20%		Yes	Evapotranspiration	Minimum value of leaf vegetation index (LAI) used in actual LAI calculations.	Source code
	LAIMAX	0.75 – 6.00	±20%		Yes	Evapotranspiration	Maximum value of LAI used in actual LAI calculations.	Source code
	EMISSMIN[@]	0.88 – 0.98	±10%		Yes	Atmosphere-Land	Minimum value of surface emissivity used in background surface emissivity calculations.	Source code
	EMISSMAX[@]	0.88 – 0.98	±10%		Yes	Atmosphere-Land	Maximum value of surface emissivity used in background surface emissivity calculations.	Source code
	ALBEDOMIN[@]	0.12 – 0.38	±20%		Yes	Atmosphere-Land	Minimum value of surface emissivity used in background snow-free albedo calculations.	Source code
	ALBEDOMAX[@]	0.12 – 0.38	±20%		Yes	Atmosphere-Land	Maximum value of surface emissivity used in background snow-free albedo calculations.	Source code
	Z0MIN	0.01 – 0.50	±20%	m	Yes	Atmosphere-Land	Minimum value of Z0 used in background Z0 calculations.	Source code
	Z0MAX	0.01 – 0.50	±20%	m	Yes	Atmosphere-Land	Maximum value of Z0 used in background Z0 calculations.	Source code
Soil	BB	2.53 – 11.55	±10%		Yes	Hydrology & Thermal	Exponent in functions for soil hydraulic conductivity and matric potential.	User guide
	MAXSMC[@]	0.200 – 0.476	±10%	$m^3 m^{-3}$	Yes	All	Soil porosity	User guide
	REFSMC	0.170 – 0.412	Passive	$m^3 m^{-3}$	Yes	Evapotranspiration	Field capacity at which transpiration weakens.	Wetzel and Chang (1988)
	WLTSMC	0.006 – 0.138	Passive	$m^3 m^{-3}$	Yes	Evapotranspiration	Wilting point at which transpiration ceases.	Wetzel and Chang (1988)
	DRYSMC	0.006 – 0.138	Passive	$m^3 m^{-3}$	Yes	Evapotranspiration	Threshold of top layer soil moisture at which direct evaporation from soil ceases.	Source code
	SATPSI	0.036 – 0.759	±10%	m	Yes	All	Saturated soil matric potential	User guide
	SATDK	9.8E−07 – 1.4E−04	±20%	$m\ s^{-1}$	Yes	Hydrology	Saturated soil hydraulic conductivity	User guide
	SATDW	6.1E−07 – 1.4E−04	Passive	$m^2 s^{-1}$	Yes	Hydrology	Saturated soil water diffusivity	Campbell (1974)

	QTZ	0.07 – 0.94	±20%		Yes	Thermal	Quartz content	User guide
Site	TBOT[@]	270 – 275	±1.5	K	Yes	Thermal	Soil-temperature bottom-boundary condition at a soil depth of ZBOT, which is the assumed nominal soil depth where the amplitude of soil-temperature annual cycle is near zero.	User guide; Chen et al. (2015)

"±xx" denotes the permissible range centered on the default value, with "xx" being a relative range (percentage) or an absolute range (numerical value). "Passive" indicates that the parameter's value is estimated by its dependency on other parameters.

& homogeneous (spatially uniform, marked as No) or heterogeneous (marked as Yes) parameter.

§ the physical modules affected by the parameter. "All" signifies that the parameter influences all four modules.

@ one of the top 10 most sensitive parameters

2.4 Sensitivity analysis

In the context of this regional modeling framework, the parameter sensitivity analysis serves primarily as a strategic prerequisite to reduce the dimensionality of the parameter space (Figure 3), rather than an exhaustive investigation of spatial parameter variability. Given that parameter behavior likely varies with environmental conditions, three representative modeling cells were selected to conduct the sensitivity analysis. These cells are: TGL (centered at 91.95 °E, 33.05 °N, named after the TGL(QT04) active layer observation station, Figure 1), representing cold permafrost conditions; WQ (centered at 99.45 °E, 35.45 °N, named after the Wenquan survey region), representing warm permafrost; and GZ (centered at 84.25 °E, 33.65 °N, named after the Gaize survey region), representing seasonally frozen ground (SFG). While each modeling cell incorporates vertically heterogeneous soil textures in actual simulations, the sensitivity analysis assumed a homogeneous soil profile for simplicity. In the subsequent calibration process, the sensitive heterogeneous parameters were extended according to the predominant vegetation type(s) and soil texture(s) at/in the calibration site/region.

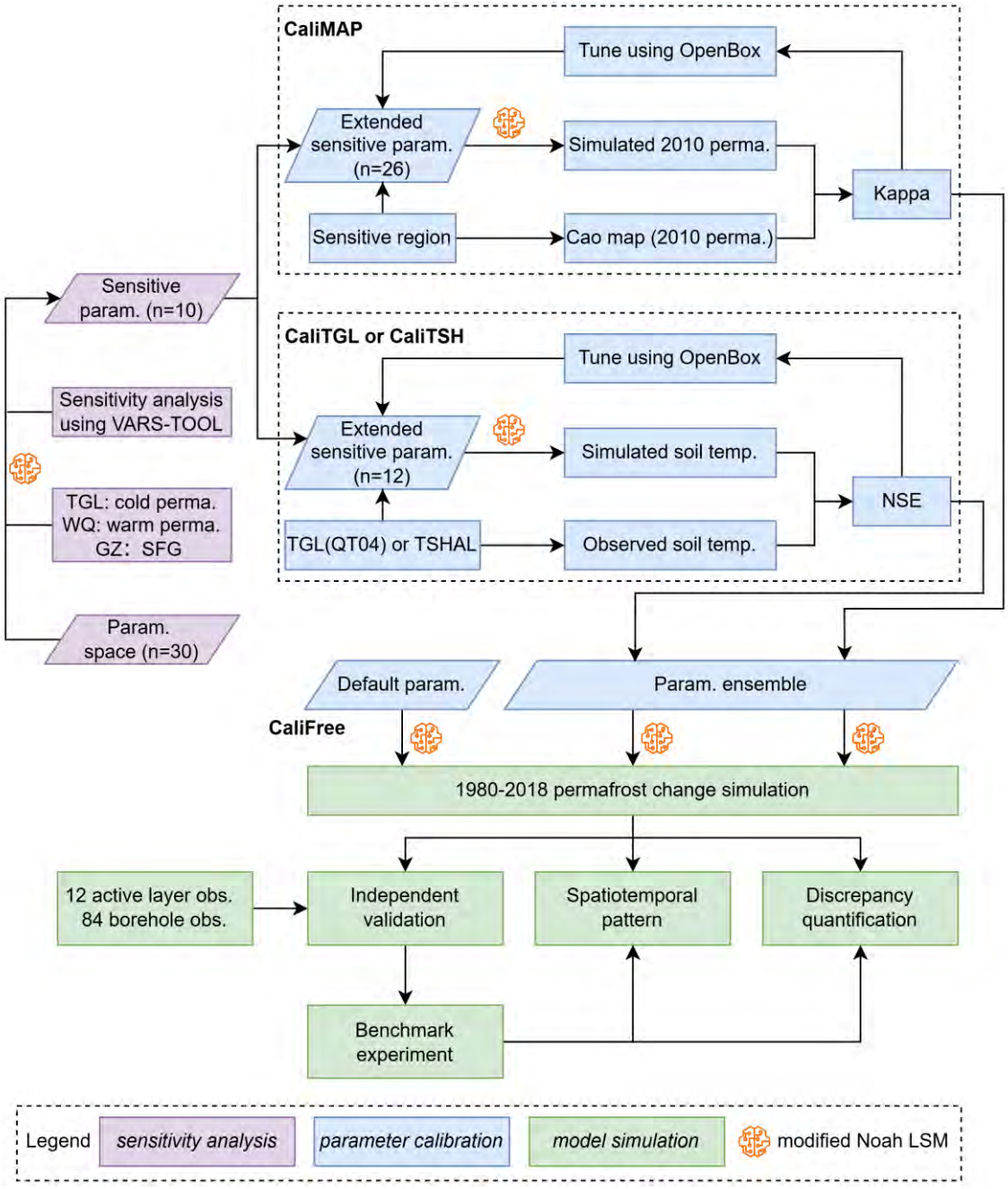

Figure 3 The flowchart of this study. TGL, WQ and GZ are three representative sites where parameter sensitivity analysis was conducted. CaliTGL, CaliTSH, CaliMAP and CaliFree are four calibration experiments. The value in parentheses indicates the number of parameters. Perma.: permafrost, SFG: seasonally frozen ground, param.: parameters, temp.: temperature, obs.: observations.

The global sensitivity analysis was performed using the "star-based variogram analysis of response surfaces" (STAR-VARS) method (Razavi & Gupta, 2016a), chosen for its effectiveness, efficiency, and robustness. The STAR-VARS approach begins with a STAR-

based sampling strategy, where m star centers are randomly generated within an n-dimensional normalized parameter space. For each star center, new sampling points (star points) are generated along each dimension at a predefined sampling step size, Δh, yielding a total of $m[n((1/\Delta h)-1)+1]$ sampling points. Each sampling point represents a unique set of parameter values, which are used to run an external model and generate a response, either a model output or a performance metric. The VARS then analyzes the spatial structure and variability of this response surface across the parameter space using variograms, adopting from geostatistics. The integral of the variogram along a given parameter dimension from 0 to 50% of its range (IVARS50) is used to measure the parameter sensitivity. This sensitivity analysis framework was implemented using the VARS-TOOL Python library (Razavi & Gupta, 2019).

Following the recommendations of Razavi and Gupta (2016b), m was set to 100 and Δh to 0.1 (resulting in a total of 27,100 samples as n is 30) to balance reliability and computational cost. To capture variability in parameter behavior, the model responses were specified as mean annual soil temperature and soil moisture content in the year 2010 (to align with the Cao map) at three depths: TOP (0.1m), MID (2.7m), and BOT (10.2m). These depths were chosen to approximate the depths of ground surface, the average active layer thickness (ALT), and the average DZAA across the QTP (Figure 2b). Consequently, a total of 18 model responses (2 state variables × 3 depths × 3 modeling cells) were generated.

As the IVARS50 value depends on the selected model response, its overall ranking was employed to screen the most sensitive parameters. For each of the 18 model responses, the 30 parameters were first ranked (following the standard competition style) in descending order based on their IVARS50 values. This resulted in each parameter receiving 18 individual ranks. To emphasize the importance of parameters that might be highly influential for specific model responses, rather than universally across all, a final sensitivity score was calculated by

summing the top 30% ranks for each parameter. The top 10 sensitive parameters with the lowest final sensitivity scores were selected for subsequent calibration.

2.5 Experiments of model calibration

This study designed four calibration experiments (Table 2 and Figure 3). The CaliFree adopted default parameters directly from the Noah LSM lookup tables, without any calibration. The purpose of CaliFree was to assess the baseline performance of the modified Noah LSM. The other three experiments involved calibration and can be grouped into two categories: site-based and map-based. The site-based calibration contains two experiments (CaliTGL and CaliTSH) to investigate how the representativeness of calibration site affects regional modeling results. The map-based calibration, referred to as CaliMAP, aims to address the challenges of parameter transferability often encountered in the site-based method.

These calibrations were both implemented using the OpenBox Python library (Jiang et al., 2024a), which is based on Bayesian optimization. OpenBox comprises a surrogate model and an acquisition function. The surrogate model approximates iteratively the model response surface across the parameter space using a pairwise parameter-objective function. For high-dimensional optimization problems exceeding 10 dimensions, OpenBox adopts a probabilistic random forest as the surrogate model to maintain computational efficiency. The acquisition function (e.g., Expected Improvement) balances exploration (probing uncharted regions) and exploitation (refining known promising regions) to propose the next set of the most promising parameter configurations. Compared to widely used heuristic optimization algorithms, the Bayesian optimization demonstrates higher efficiency, greater robustness in locating the global optimum, and better suitability for computationally expensive black-box optimization problems. The number of iterations was set to 500, and early stopping was allowed, with a tuning step size of 1/100 for each normalized parameter. A piecewise linear regression framework (utilizing the Bayesian Information Criterion to penalize overfitting) was applied

to detect structural breakpoints in the objective function trajectory. From the optimal sets defined by these quantitative convergence thresholds, 13 ensemble members were randomly selected to perform the parameter-ensemble simulations (Figure 3).

Table 2 Details of four calibration experiments.

Experiment	Calibration	Simulation	Motivation
CaliFree	No calibration	Default set	To assess the baseline performance of the modified Noah LSM
CaliTGL	Calibration using soil temperature observed at TGL(QT04) station	13 ensembles randomly selected from optimal sets	To revisit the conventional paradigm of "site calibration-regional simulation"
CaliTSH	Same as CaliTGL but at TSHAL station, where the climate is more arid	13 ensembles randomly selected from optimal sets	To investigate the effect of calibration site representativeness, together with CaliTGL
CaliMAP	Calibration using the Cao map, a 2010 permafrost distribution map on the QTP	13 ensembles randomly selected from optimal sets	To improve the poor spatial parameter transferability encountered in site-based calibration

2.5.1 Site-based calibration

The CaliTGL and CaliTSH experiments were conducted at two active-layer monitoring stations (Figure 1 and Figure 3): TGL(QT04) (91.94°E, 33.07°N) and TSHAL (79.55°E, 35.36°N). The TGL(QT04) station has a MAAT of –4.7°C, annual precipitation of 352 mm, and is dominated by alpine meadow. In contrast, the TSHAL station has a MAAT of –6.0°C, annual precipitation of 103 mm, and alpine desert vegetation.

At the TGL(QT04) site, soil temperature and moisture monitoring data are available at depths of 5, 10, 40, 105, and 245 cm from April 1, 2007, to October 31, 2010. The TSHAL site provides similar data at depths of 30, 60, 120, 160, and 180 cm from January 1, 2016, to December 31, 2018. For both experiments, the last year of data was used for validation, and the preceding period for calibration. The capability for spatial modeling was assessed by validating against the permafrost observation network across the QTP, as detailed in Section 2.6.

At the TSHAL site, thick excess ice exists at the top of the permafrost, as a relic of Quaternary permafrost (Jin et al., 2020). Given the Noah LSM's limitations in modeling excess ice, soil temperature was chosen as the sole calibration target for both CaliTGL and CaliTSH for fairness. However, we also present and discuss comparisons between simulated and observed moisture changes to evaluate the model's ability to represent hydrological processes within permafrost.

At both sites, sensitive heterogeneous parameters were extended according to predominant soil texture and vegetation types (Table S1). These parameters were tuned within the permissible ranges (Table 1) to maximize the arithmetic mean of the Nash-Sutcliffe efficiency (NSE) coefficient between observed and simulated soil temperature across all available depths during the calibration period. For the regional permafrost change simulation, parameters associated with vegetation and soil types that could not be calibrated retained their default values, as in the CaliFree experiment.

2.5.2 Map-based calibration

To prevent the calibration from being unduly biased toward correcting errors specific to the default parameter set, the sensitive regions were not restricted solely to the cells with classification discrepancies. Instead, a spatial buffer encompassing the 3×3 neighborhood of each discrepant cell was included, expanding the calibration domain to over 13,000 cells (>50% of the QTP). This spatial buffering exposed the optimization algorithm to a broader range of environmental conditions, thereby mitigating the risk of overfitting. Conversely, the entire QTP was not utilized for calibration to avoid severe equifinality; in deeply stable permafrost regions, parameter perturbations within permissible ranges rarely alter the binary frozen ground classification, which would render the Kappa-based objective function largely insensitive to parameter changes. In the sensitive regions, the heterogeneous sensitive parameters were expanded according to the predominant soil texture and vegetation types (Table S1). The

Kappa coefficient between the simulated 2010 permafrost distribution and the Cao map within the sensitive regions was used as the objective function in the CaliMAP (Figure 3). The identification of frozen ground type from simulated soil temperature was detailed in Section 2.7.

2.6 Independent validation

The performance of each calibration strategy was assessed against the permafrost observation network (Zhao et al., 2021) across the QTP through mean annual values and change rates for ALT, temperature at the top of permafrost (TTOP), and ground temperature at the depth of 10 m (GT10m). To assure strict spatial independence, the specific calibration site was excluded from the regional validation network for both CaliTGL and CaliTSH. The result of independent validation was used to determine the benchmark calibration experiment (Figure 3). The derivation of these thermal variables from observed/simulated soil temperature is detailed in Section 2.7. For each thermal variable, observed values within the same modeling cell were averaged. For CaliTGL, CaliTSH and CaliMAP, the ensemble mean was used in validation.

Spatial modeling performance was quantified using root mean square error (RMSE) and mean bias error (MBE), where values closer to zero indicate higher accuracy. Toward a thorough assessment, the RMSE and MBE for each thermal variable were calculated at two scales: individual sites and the entire plateau. At individual sites, site-specific RMSE was used to rank the performance of four calibration experiments. Given the number of deep boreholes is up to 84, for simplicity, the site-specific RMSEs of GT10m were averaged and then ranked within each survey region. At the regional scale, both RMSE and MBE were calculated across all locations for comparison with previous modeling studies.

2.7 Permafrost variables and statistical methods

Several permafrost variables were derived from observed/simulated soil temperature data: permafrost area, ALT, TTOP, DZAA, mean annual ground temperature (MAGT) and GT10m. Following the definition by van Everdingen (1998), a modeling cell of the modified Noah LSM (18 layers, 15.2 m) was classified as permafrost in a given year if at least one soil layer remained continuously frozen (i.e., maximum temperature ≤ 0 °C) throughout both the preceding and current years. Conversely, if a modeling cell did not meet this permafrost criterion but experienced sub-zero temperatures in any soil layer during the year (i.e., minimum soil temperature ≤ 0 °C), it was classified as SFG. Consequently, the analysis of permafrost change starts in 1980, while the simulation began in 1979. To align with previous studies, the modelling cell covered by glaciers or lakes was excluded. The permafrost area was calculated by summing the areas of all permafrost cells using an equal-area projection. For each permafrost cell, ALT was determined as the maximum thawing depth from the surface downwards (Figure 2b), and the mean annual ground temperature at this depth was defined as TTOP. DZAA was estimated as the depth at which the annual ground temperature amplitude decreased to below 0.1 °C (Figure 2b), and the mean annual ground temperature at this depth was defined as MAGT. Additionally, GT10m was derived using linear interpolation for validation purposes, as DZAA and MAGT are often not available in in-situ observations.

The spread of ensemble simulation was represented by a confidence interval (CI) calculated as:

$$CI = \bar{x} \pm t\frac{s}{\sqrt{n}} \quad (1)$$

where $\bar{x}$ is the ensemble mean, s is the sample standard deviation, and n is the ensemble size. Given the small sample size (n=13) and unknown population standard deviation, we used the critical value from the t-distribution, $t_{\alpha/2,n-1}$, in place of the z-score. The t-critical value depends on the chosen confidence level (α) and the degrees of freedom ($n-1$). For a 95% confidence interval with our 13 ensemble simulations, the t value is 2.179. The Mann-Kendall

test and Sen's slope were employed to estimate the change rate in the time series, using the ensemble mean where applicable.

3 Results

3.1 Sensitive parameters

The sensitivity analysis revealed that simulated soil temperature was generally highly sensitive to parameters governing land-atmosphere energy exchange and evapotranspiration (Figure 4, Table 1), such as the lower and upper bounds of vegetation emissivity (EMISSMIN and EMISSMAX), the lower and upper bounds of vegetation albedo (ALBEDOMIN and ALBEDOMAX), the soil moisture stress exponent (FXEXP), and the canopy suppression effect on surface heat flux (SBETA) (Figures S1-S3). Additionally, the lower boundary condition for the thermal field (TBOT) significantly impacted the deep ground temperature (BOT, 10.2m).

Due to the strong coupling between soil moisture and heat through processes like water phase change, the soil freezing characteristic curve, and thermal conductivity, most of the aforementioned parameters also influenced soil moisture. Furthermore, soil porosity (MAXSMC) and the ice content threshold for frozen ground impermeability (FRZK) particularly affected soil moisture dynamics. Notably, simulated soil moisture at the permafrost sites (TGL and WQ) exhibited greater sensitivity to parameter perturbations compared to the SFG site (GZ), suggesting higher uncertainty in permafrost hydrology modeling (Figures S1-S3).

The parameter representing the ratio of thermal roughness length to momentum roughness length (CZIL) showed significant influence only at the TGL and WQ sites. This site-specific sensitivity arose because these two locations have relatively dense vegetation cover, and the original roughness scheme parameterized by CZIL was adopted (Figure 2c). In contrast, the

sparsely vegetated GZ site used an alternative scheme that does not involve CZIL in the modified Noah LSM.

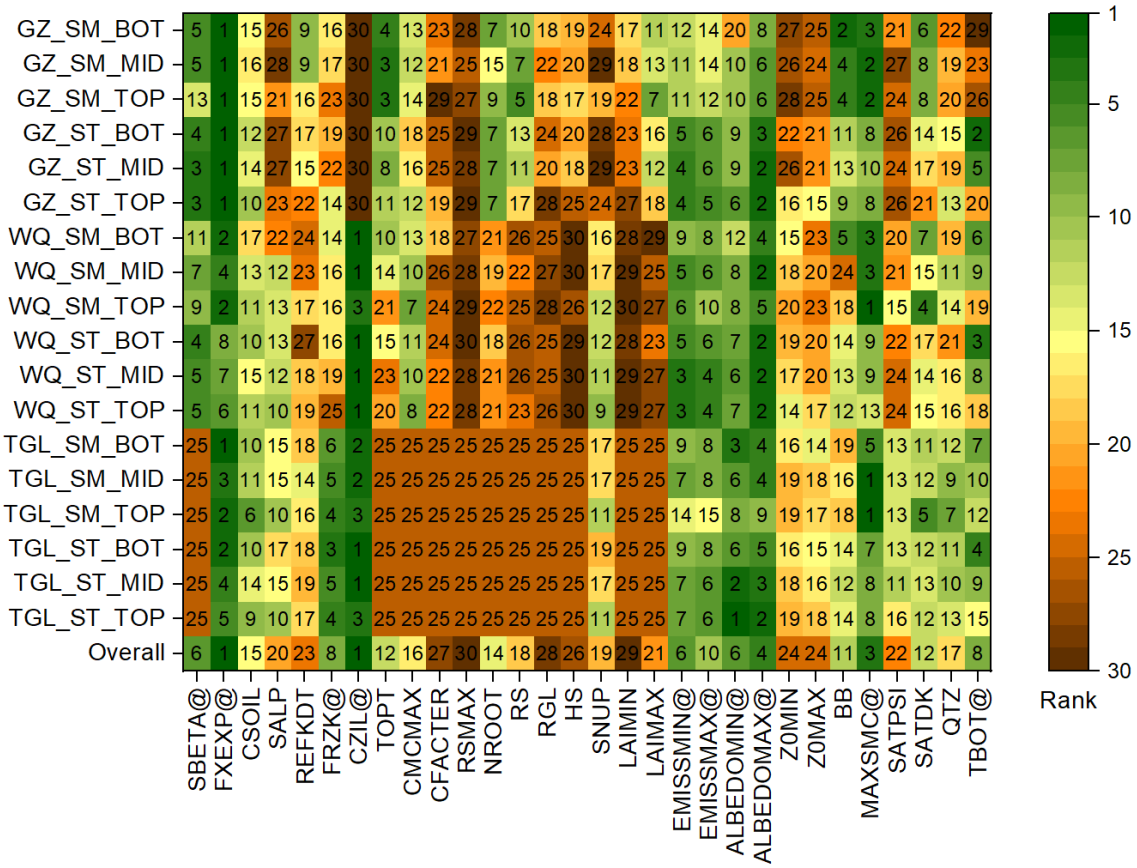

Figure 4 Parameter sensitivity ranking for soil temperature and moisture simulations at three depths for three modeling cells. The x-axis displays parameter names (detailed in Table 1). The y-axis labels follow the format "cell_variable_depth", where "cell" represents the modeling cell (TGL, WQ, or GZ), "variable" represents the state variable (ST for soil temperature, SM for soil moisture), and "depth" indicates the soil depth (TOP: 0.1 m, MID: 2.7 m, BOT: 10.2 m). "Overall" represents the overall sensitivity ranking, with the top 10 (ranked 1st to 10th) most sensitive parameters marked with "@". Raw sensitivity values are shown in Figures S1-S3.

To balance effectiveness and efficiency in the subsequent calibration process, the top 10 sensitive parameters were identified (Figure 4): SBETA, FXEXP, FRZK, CZIL, EMISSMIN, EMISSMAX, ALEBDOMIN, ALBEDOMAX, MAXSMC, and TBOT. These parameters were then extended based on the predominant vegetation types and soil textures at the

calibration sites or sensitive regions (Table S1), yielding 12 parameters for both CaliTGL and CaliTSH, and 26 parameters for CaliMAP (Figure 3).

3.2 Calibration process

Applying the piecewise linear regression to the objective function iteration history (Figure 5), the quantitative convergence thresholds defining the optimal parameter sets were identified as an NSE of 0.88 for CaliTGL, an NSE of 0.95 for CaliTSH, and a Kappa coefficient of 0.49 for CaliMAP. These three experiments exhibited different behaviors during the calibration process. CaliTSH, calibrated at the drier, colder site of TSHAL, achieved a substantial improvement in NSE, increasing from −0.04 with default parameters to 0.95. However, its convergence rate was slower than CaliTGL (Figure 5a, b). This is likely because the default parameters were not well-suited for the environmental conditions at TSHAL, leading to an overestimation of soil temperatures due to surplus heat entering the ground (Figure S4).

In contrast, CaliTGL, calibrated at the TGL(QT04) site, showed a smaller but still significant increase in NSE (from 0.71 with default parameters to 0.88) and a much faster convergence rate. The default parameters at TGL(QT04) already provided satisfactory performance in shallower soil layers, but accuracy deteriorated at depths exceeding 100 cm (Figure S6), a discrepancy effectively addressed by optimized parameters.

Despite achieving high accuracy at their respective calibration sites, the optimized parameter sets from CaliTGL and CaliTSH suffered from poor spatial transferability. When the optimized parameters from CaliTGL were applied to the TSHAL site, the obtained NSE was 0.23, slightly higher than the default parameters (NSE = −0.04), but still greatly lower than the optimal result (NSE = 0.95) achieved by CaliTSH. Similarly, but more seriously, the optimized parameters from CaliTSH achieved an NSE of 0.68 at the TGL(QT04) site, even lower than the default parameters (NSE = 0.71).

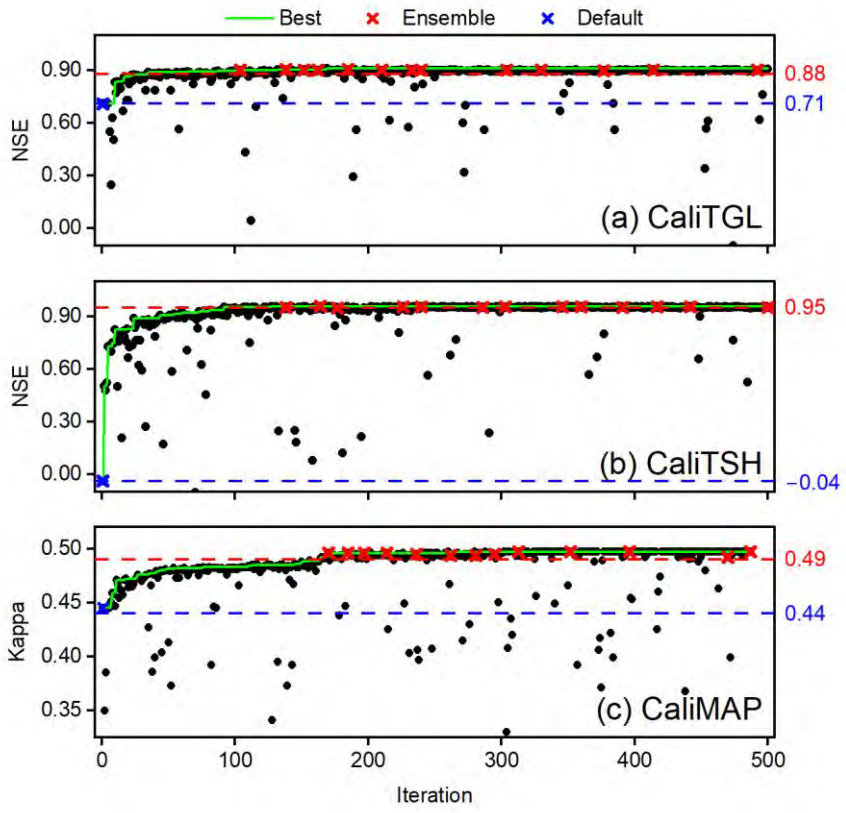

Figure 5 Iteration of the objective function during the optimization process for (a) CaliTGL, (b) CaliTSH, and (c) CaliMAP. In each panel, the blue dashed line indicates the model performance achieved with default parameters (i.e., CaliFree), the red dashed line represents the performance criterion above which the parameter sets were recognized as optimal, and the red crosses show the 13 ensemble members randomly selected from the optimal sets. In (a) and (b), the objective function is the Nash-Sutcliffe Efficiency (NSE) coefficient of soil temperature. For CaliMAP, the objective function is the Kappa coefficient of the 2010 permafrost distribution in sensitive regions.

CaliMAP, which calibrated parameters in the sensitive regions containing over 13,000 cells, improved the Kappa coefficient from 0.44 to 0.49, indicating moderate agreement in the 2010 permafrost distribution between the simulation and the Cao map. While this absolute value may appear modest, it is important to note that it was calculated exclusively within the highly transitional sensitive regions rather than across the entire QTP. In the sensitive regions, the classification of frozen ground is much more challenging. Nevertheless, CaliMAP effectively reduced the prominent false negatives observed with the default parameters along the periphery

of the continuous permafrost region in Northern Tibet (Figure S8). The convergence of the objective function in CaliMAP was considerably slower than that of CaliTGL and CaliTSH (Figure 5), primarily due to the much larger parameter space (26 versus 12).

3.3 Performance comparison of calibration methods

The performance of each calibration strategy was evaluated by the permafrost observation network in terms of the mean annual values and change rates of key permafrost variables: ALT, TTOP and GT10m. The results revealed that CaliMAP consistently outperformed other strategies at most sites, exhibiting the highest number of top rankings irrespective of the permafrost variable considered (Figure 6). CaliFree generally ranked third for ALT and TTOP, but with a comparatively better ability to simulate GT10m. Notably, CaliTSH frequently ranked lowest despite achieving a high NSE (>0.95) at its calibration site (TSHAL), underscoring a significant limitation in its spatial transferability. CaliTGL performed better than CaliTSH, likely due to the broader representativeness of the TGL(QT04) site's climate and permafrost conditions for the existing observation network. Given its highest performance, the CaliMAP was regarded as the benchmark experiment.

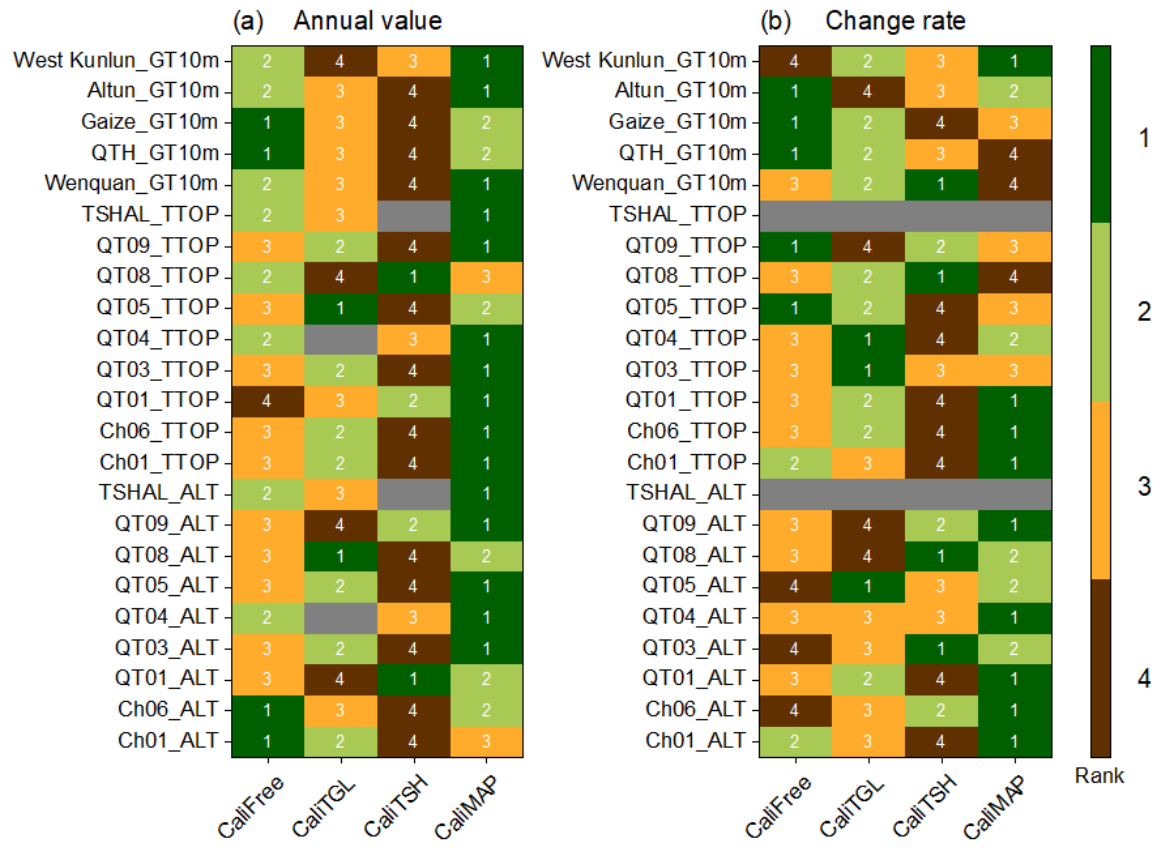

Figure 6 Performance ranking of four experiments based on Root Mean Square Error (RMSE) against independent in-situ observations. The filled color and number indicate the ranking, while the gray box represents validation that was not applicable due to either insufficient observation span or a violation of spatial independence. The y-axis label follows the format "observation site/survey region_permafrost thermal variable".

The regional scale validation of CaliMAP is presented in Figure 7. For mean annual values (Figure 7a-c), the RMSE was 0.68 m for ALT (83 samples), 0.41°C for TTOP (83 samples), and 1.30°C for GT10m (291 samples). As for corresponding change rates (Figure 7d-f), the RMSE was 0.08 m/10a for ALT (8 time series), 0.36°C/10a for TTOP (8 time series), and 0.64°C/10a for GT10m (38 time series). The consistently negative MBE across all variables suggests a systematic underestimation by the modified Noah LSM. This bias may stem from the model's limitations in representing key physical processes contributing to permafrost degradation, such as convective heat transfer within the active layer, or an underestimation of

the energy and moisture exchange efficiency at the land-atmosphere interface, potentially related to parameters like roughness length and soil thermal conductivity.

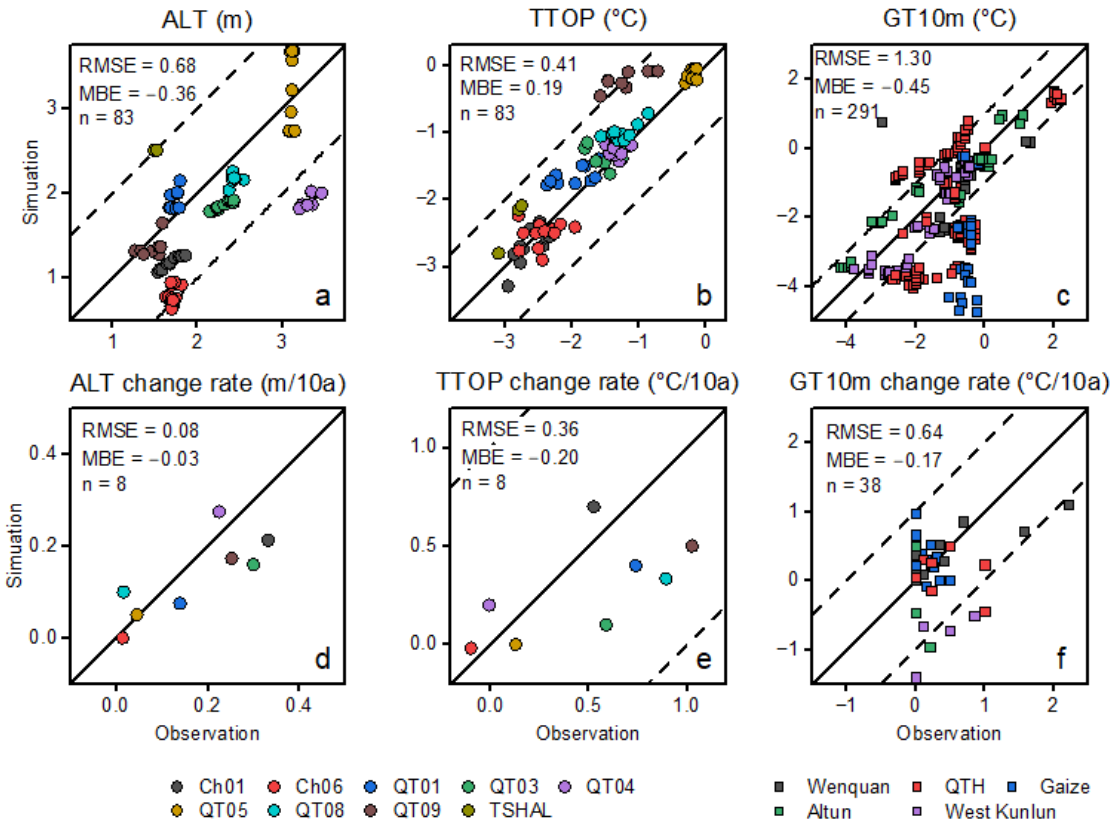

Figure 7 Scatter plots showing the regional scale comparison between in-situ observations and CaliMAP simulations (ensemble mean of 13 members) for mean annual (a) ALT, (b) TTOP, (c) GT10m, and their respective change rates: (d) ALT, (e) TTOP, (f) GT10m. A solid 1:1 line and dashed lines representing a ±1 offset are plotted for visual assessment of the model performance.

3.4 Regional permafrost changes over 1980-2018

The CaliMAP ensemble simulations revealed a considerable reduction in permafrost area across the QTP, decreasing from $1.262\pm0.048\times10^6$ km^2 (95% CI) in 1980 to $1.085\pm0.049\times10^6$ km^2 in 2018 (Figure 8). There was a total reduction of $177\pm2\times10^3$ km^2 over the study area, at an average rate of 44.4×10^3 km^2/10a, in response to a regional climate warming rate of 0.39 °C/10a. Concurrently, the area covered by SFG continuously expanded, surpassing the permafrost area by the mid-1980s (Figure 8b). For comparison, while glacier retreat and lake

expansion have received considerable attention (Xu et al., 2024; Yao et al., 2022), their area changes are relatively minor, estimated at around 10×10^3 km^2 in total (Guo et al., 2015; Zhang et al., 2017). This order-of-magnitude difference emphasizes the potentially far-reaching and currently underestimated ecological and hydrological impacts of permafrost degradation on the QTP.

The most pronounced permafrost loss occurred along the fringes of the continuous permafrost region, particularly in transition zones, as well as in the southwestern (Gangdese Range) and eastern (Three Rivers Headwaters Region) parts of the plateau. Notably, the northwestern plateau experienced a distinct cooling in the late 1990s (circa 1998) (Figure 8a), leading to the formation of new permafrost and a slight increase in permafrost area during the early 2000s (Figure 8b). This cooling event aligns with the independently observed Karakoram anomaly in adjacent glaciated regions (Farinotti et al., 2020), although a lack of in-situ meteorological data in this remote region prevents direct validation.

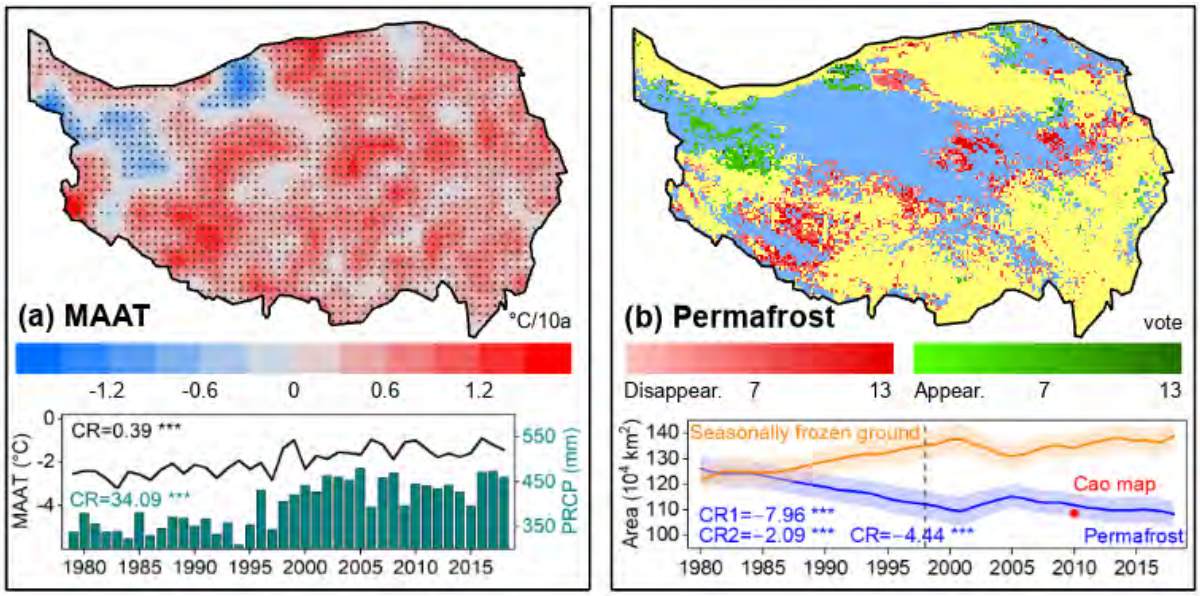

Figure 8 Spatiotemporal patterns of (a) mean annual air temperature (MAAT) trend and (b) simulated permafrost disappearance/appearance (number of ensemble votes for change) across the QTP. Shaded areas represent the 95% confidence interval. Change rates (per decade) are shown for two periods: CR1 (1980-1997) and CR2 (1998-2018), as well as for the entire period CR (1980-2018). Statistical significance of the trends is indicated by the asterisks: ***(p <

519 0.001), **($p < 0.01$), *($p < 0.05$), and NS ($p \geq 0.05$). In the spatial map of MAAT trends (a),
520 grid cells with $p < 0.05$ are marked as black dots. In the time series of permafrost area in (b),
521 the red dot around the year 2010 represents the permafrost area estimated by the Cao map.

522 Beyond the loss in permafrost area, widespread thermal changes were evident across a broader
523 region (Figure 9). The simulated increase in ALT, the upward shift of the DZAA, and the rise
524 in TTOP and MAGT collectively indicated a thinning and warming permafrost regime across
525 the QTP, albeit with considerable spatial heterogeneity. Consistent with the newly formed
526 permafrost in the northwest (Figure 8b), the modeling results of thermal variables also revealed
527 permafrost cooling in this region, extending over a larger area than the observed permafrost
528 gain.

529 The atmospheric cooling event in the late 1990s significantly influenced the overall trends in
530 regionally averaged thermal variables over the 1980-2018 period. Change rates were notably
531 lower when considering the entire period compared to the 1980-1997 warming phase (Figure
532 9): ALT increased by 0.02 m/10a (1980-2018) vs. 0.12 m/10a (1980-1997); TTOP increased
533 by 0.10 °C/10a vs. 0.37 °C/10a; DZAA decreased (indicating shallower depth) by −0.53 m/10a
534 vs. −0.61 m/10a; and MAGT increased by 0.19 °C/10a vs. 0.23 °C/10a. The temporal
535 differences were more pronounced in the shallow-layer thermal variables (ALT and TTOP)
536 compared to the deep-layer variables (DZAA and MAGT), reflecting the attenuation of surface
537 climate change signals with increasing depth. The seemingly anomalous finding of a greater
538 change rate in MAGT (0.20 °C/10a) compared to TTOP (0.10 °C/10a) over the past four
539 decades can be attributed to the anomalous MAAT decrease during the late 1990s. However,
540 during the general warming period of 1980-1997, the change rate of TTOP (0.37 °C/10a) was
541 higher than that of MAGT (0.23 °C/10a), as is expected under a dominant top-down permafrost
542 degradation mode in the QTP (Jin et al., 2006; Wu et al., 2010).

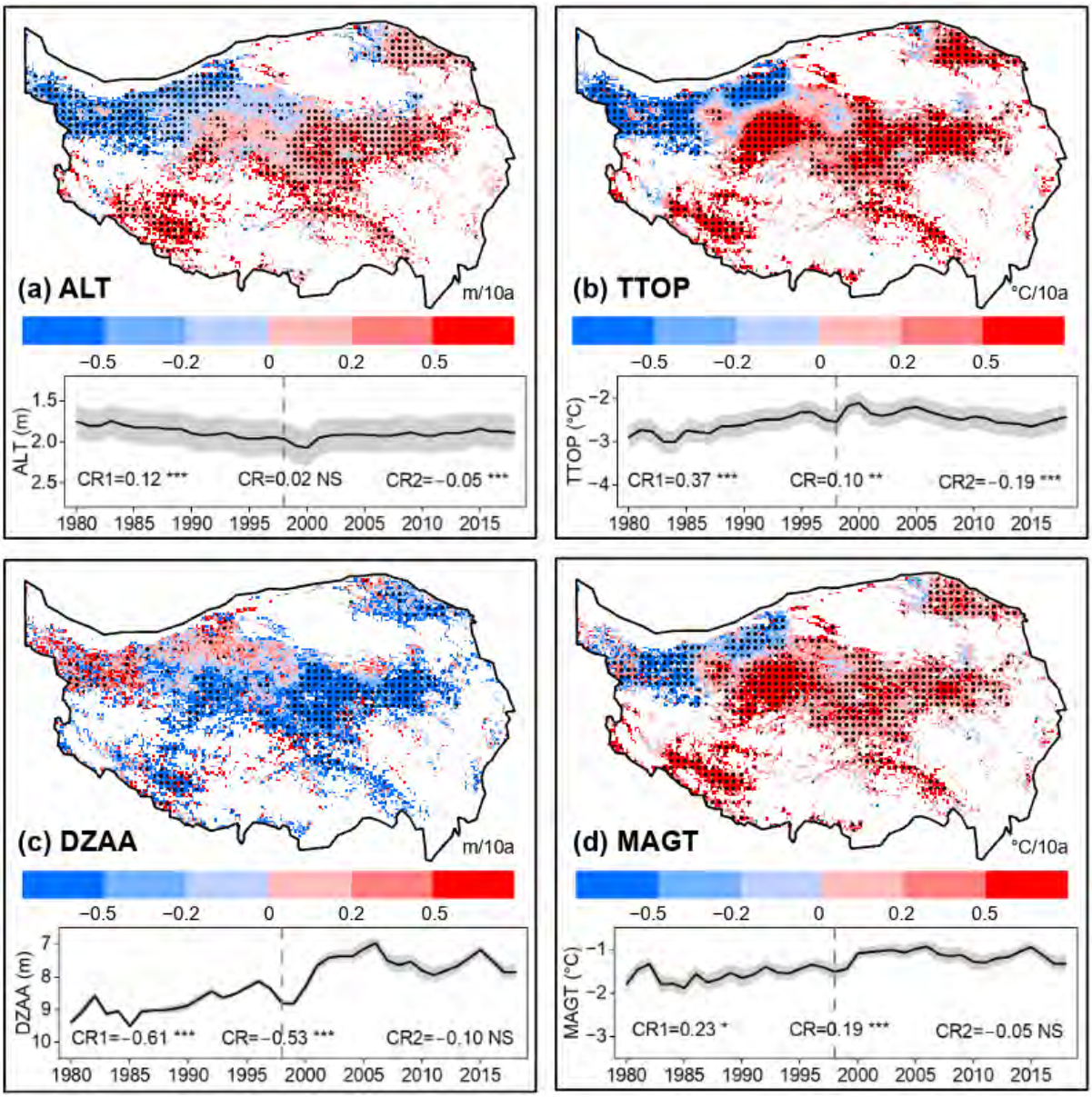

Figure 9 Spatiotemporal patterns of simulated permafrost thermal variables: (a) active layer thickness (ALT), (b) temperature at the top of permafrost (TTOP), (c) depth of zero annual amplitude (DZAA), and (d) mean annual ground temperature (MAGT). The spatial maps show the trend (change rate in m/10a or °C/10a) of each variable across the QTP. The time series below each map illustrates the regionally averaged mean annual values of the respective thermal variable from 1980 to 2018, along with change rates for 1980-1997 (CR1), 1998-2018 (CR2), and 1980-2018 (CR), with statistical significance indicated as in Figure 8.

While the cooling in the late 1990s provides a partial explanation for the observed piecewise trends in permafrost thermal variables, the contribution of increased precipitation (wetting) should also be considered. Unlike the localized cooling in the western plateau, wetting occurred

more broadly across the QTP (Wang et al., 2018; Zhang et al., 2021). In the arid and semi-arid regions prevalent in the middle and western plateau, increased precipitation can induce permafrost cooling through mechanisms such as enhanced evapotranspiration (Shen et al., 2015) and increased soil heat capacity (Zhang et al., 2021). However, the high spatial consistency observed between trends in MAAT and permafrost thermal variables (Figure 8a and Figure 9) suggests that MAAT remains the primary driver of permafrost changes on the QTP.

4 Discussion

4.1 Regional discrepancies caused by calibration methods

Taking CaliMAP as the benchmark (Figure 6), the regional discrepancies in permafrost area and key thermal variables simulated by the other three calibration experiments were quantified. For the permafrost area (Figure 10a), the discrepancy caused by CaliFree or CaliTGL did not exceed 10%, while that related to CaliTSH ranged from 20% to over 50%.

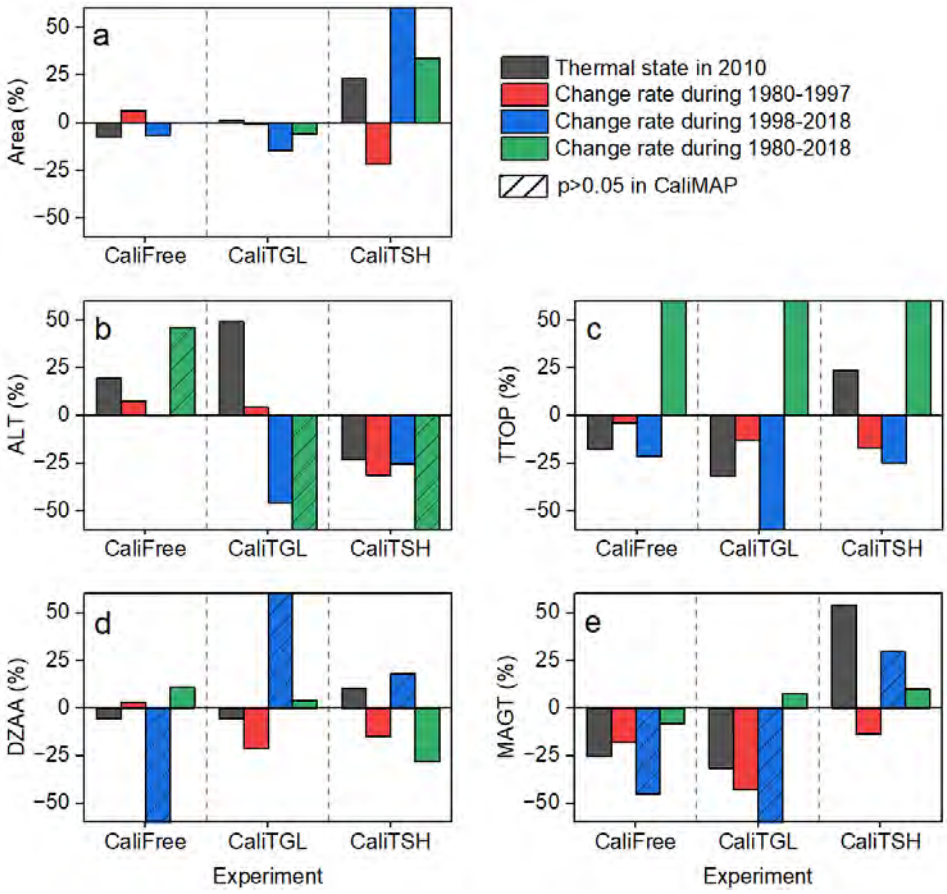

Figure 10 Percentage discrepancies in modeling the permafrost area, regionally averaged thermal variables, and their change rates relative to CaliMAP. For each metric, the discrepancy is calculated as (CaliXXX − CaliMAP) / CaliMAP, where CaliXXX represents one of CaliFree, CaliTGL, and CaliTSH.

As for other permafrost thermal variables (Figure 10b-e), the results showed that all three experiments caused considerable discrepancies compared to the benchmark experiment. For ALT in 2010 (Figure 10b), CaliFree and CaliTGL tended to overestimate, whereas CaliTSH underestimated. A similar discrepancy occurred in the change rates of ALT. Regarding TTOP and MAGT (Figure 10c and 9e), CaliFree and CaliTGL generally simulated colder conditions, whereas CaliTSH tended to be warmer. There were also substantial discrepancies in terms of DZAA (Figure 10d). It is noteworthy that the thicker 2010 ALT coexisted with the colder 2010 TTOP and GT10m in CaliFree and CaliTGL. This contrasting discrepancy highlights the non-dominant role of thermal conditions on ALT at the regional scale, aligning with the previous

findings that observed ALT is strongly modulated by soil moisture (Clayton et al., 2021). Furthermore, the simulated regional ALT is generally thicker in CaliTGL but thinner in CaliTSH. These opposing regional biases are likely linked to the specific thermal conditions at their respective calibration sites: the observed ALT is 3.3 m at the TGL site but only 1.5 m at the TSHAL site (Zhao et al., 2021). This clearly demonstrates that the unique environmental characteristics of a local calibration site can inappropriately propagate into the regional simulation, underscoring the limitations of parameter transferability in conventional site-based calibration.

4.2 Soil moisture modeling challenge

While permafrost is defined by ground temperature, soil moisture plays an implicit yet critical role in modulating both the seasonal variations and long-term trends of ground temperature through latent heat effects (Clayton et al., 2021; Nicolsky & Romanovsky, 2018; Smith et al., 2022). The strength of the coupling between temperature and moisture regimes significantly influences the reliability of permafrost modeling. In CaliTGL and CaliTSH, soil temperature served as the sole calibration target for fairness. However, temperature-targeted calibration had divergent effects on soil moisture simulation at their calibration sites. At the TGL(QT04) station, where CaliTGL was conducted, the simulation fidelity of soil moisture improved concurrently with soil temperature (Figure S7), likely benefiting from the coupling between soil temperature and moisture fields in the model. In contrast, at the TSHAL station, where CaliTSH was conducted, persistent and substantial discrepancies in soil moisture were exhibited regardless of parameter calibration (Figure S5). This suggests that the absence of critical physical processes, such as the excess ice at the TSHAL station, cannot be fully compensated for by parameter calibration alone. However, most current process-based permafrost models lack the capability to simulate excess ice, with only a few recent exceptions (Aga et al., 2023; Westermann et al., 2023).

Even setting aside excess ice, modeling soil moisture in cold regions remains a formidable challenge, given the repeated freezing-thawing cycles associated with unfrozen water content and cryosuction (Kurylyk & Watanabe, 2013). In fact, soil moisture modeling remains a recognized and persistent bottleneck within the broader land surface modeling community (Blöschl et al., 2019; Fisher & Koven, 2020). While the theoretical framework for soil water migration is provided by Richards' law, analogous to heat conduction governed by Fourier's law, accurately estimating critical hydrological parameters (e.g., soil water conductivity and soil water potential) remains a significant challenge. These parameters often exhibit high horizontal and vertical heterogeneity, strongly influenced by the microscale structure of the porous medium. On the other hand, emerging deep learning techniques provide a promising methodology for improving soil moisture modeling in classical LSMs. Their high capability in fitting complex nonlinear relationships can be leveraged to provide more accurate parameter estimations (Tsai et al., 2021) or even to directly learn the underlying governing equations from observational data (Song et al., 2025).

4.3 Strengths

Current regional permafrost modeling usually relied on default parameters or site-based calibration (Riseborough et al., 2008). However, the default parameter often results in compromised simulation accuracy (e.g., CaliFree), given the unique regional environment and inherent complexity of permafrost. While site-based calibration can yield satisfactory performance at the calibration site, it suffers from overfitting and easily results in uncontrolled model behavior (e.g., CaliTSH). The performance of the conventional method depends on the representativeness of the calibration site, which is hard to quantify and can only be judged through the modeler's experience.

Recognizing the practical limitations of existing approaches, this study sought to constrain the regional permafrost simulations by a high-fidelity permafrost map. In contrast to site-based

calibration, map-based calibration aims to find a parameter set that performs well across the entire spatial domain. The theoretical advancement is supported by the validation against independent in-situ observations across the QTP (Figure 6). To overcome the severe equifinality imposed by using a categorical map as a constraint, we designed a targeted, four-step mitigation framework. First, we utilized a global sensitivity analysis tool (VARS-TOOL) to aggressively reduce the parameter space, ensuring only the most dominant parameters were calibrated. Second, we confined the map-based calibration exclusively to sensitive regions, specifically the transition zones where permafrost is thermally marginal. Constraining the model to these boundaries imposes a sharper penalty on parameter sets that would otherwise produce identical permafrost presence outputs. Third, we employed a highly efficient global optimization algorithm (OpenBox) to navigate the remaining complex parameter space. Finally, rather than relying on a single optimal parameter set, we conducted a multi-parameter ensemble simulation, allowing us to provide a robust envelope of regional permafrost dynamics. Besides, similar strategies of incorporating regional-scale constraints have been implemented in hydrology modeling (Chen et al., 2017; Mei et al., 2023), and our previous studies have explored it through a frost number model but with only one tunable parameter (Cao et al., 2023; Hu et al., 2020). This study advanced this calibration strategy to a process-based model with 30 tunable parameters, a much more complex problem.

When evaluated against state-of-the-art permafrost modeling studies on the QTP (Table 3), our study achieved comparable spatial performance while offering a more conservation estimate of long-term permafrost degradation. Our estimated rate of permafrost area loss (44.4×10^3 km^2/10a) falls within the lower end of previous reported ranges (36.2 to 92.0×10^3 km^2/10a). Furthermore, the cooling trend and associated shallower ALT observed in the western plateau since 1998 are consistent with findings from recent independent findings utilizing the Community Land Model (Huang et al., 2024).

In terms of spatial accuracy, our modeling yielded an RMSE of 0.68 m for ALT across 83 samples (compared to a range of 0.22–0.69 m across 18–76 samples in previous studies) and an RMSE of 1.30 °C for deep ground temperature across 291 samples (compared to 0.45–0.72 °C across 46–150 samples), similar to previous studies (Guo & Wang, 2013; Huang et al., 2024; Qin et al., 2017; Wang et al., 2023; Zheng et al., 2020). It is crucial to contextualize these metrics, as this study employed a strictly independent validation dataset that is nearly twice the size of those used in previous efforts. Moreover, our study extends the validation to mean annual values of TTOP and the temporal trends of all three key thermal variables (ALT, TTOP and GT10m), which are often absent in prior literature yet crucial for establishing confidence in regional change simulations.

Table 3 Synthesis of published modeling studies on the QTP showing the reported change rates of permafrost area and active layer thickness (ALT) over the past decades, compared with the results of this study.

Study	**Model**	**Period**	**Change rate in permafrost area** (km^2/10a)	**Change rate in ALT** (m/10a)
Guo and Wang (2013)	Community Land Model version 4	1981-2010	-92.0×10^3	0.15
Qin et al. (2017)	Geophysical Institute Permafrost Lab version 2	1980-2013		0.31
Wang et al. (2019)	Logistic regression model	1980s-2000s	-66.0×10^3	
Zheng et al. (2020)	Geomorphology Based EcoHydrological Model	2002-2016	-83.3×10^3	0.32
Wang et al. (2023)	Geomorphology Based EcoHydrological Model	1980-2019	-36.8×10^3	
Shen et al. (2024)	eXtreme Gradient Boosting	1970-2020	-36.2×10^3	0.05
Huang et al. (2024)	Community Land Model version 5	1980-2020		0.01 (1980-1999) −0.01(2000-2020)
This study	Modified Noah LSM	1980-2018	-44.4×10^3	0.12(1980-1997) −0.05(1998-2018)

It is important to note that this study was conducted at a spatial resolution of 0.1°, which is coarser than the resolution (e.g., 0.01°) used in some other studies. While finer spatial resolution meteorological and environmental input data might potentially lead to simulations aligning more closely with in-situ observations, we caution that most numerical permafrost models are built with a one-dimensional vertical structure (e.g., Figure 2). Such one-dimensional models are inherently limited in their ability to adequately resolve fine-scale processes such as lateral heat transfer (Aas et al., 2019; Painter et al., 2016), which plays a critical role in the evolution of warm, spatially discontinuous permafrost (island permafrost) within the transition zone (Jin et al., 2006; Wu et al., 2010). Therefore, operating a one-dimensional model at a much finer-scale resolution may be inappropriate, despite the appeal of higher accuracy. The accuracy achieved by our simulation at 0.1° resolution is encouraging, highlighting both the suitability of the physical processes incorporated in the modified Noah LSM and the effectiveness of the map-based calibration approach for regional permafrost modeling on the QTP.

4.4 Limitations and outlook

Despite the promising results, our modeling approach still has room for improvement. The parameter sensitivity analysis, which used mean annual soil temperature and moisture content as model responses, might have obscured important variations occurring at finer temporal scales (Hu et al., 2023). Seasonal biases are a well-documented challenge in numerical permafrost simulation (e.g., Smith et al., 2022; Zhao et al., 2023). In the parameter sensitivity analysis, the soil column was assumed to be vertically homogeneous (configured as Loam, the dominant regional soil type, Table S1) to establish a generalized baseline. We acknowledge that parameters, particularly soil hydraulic parameters, may exhibit different sensitivities within a realistically stratified soil column due to complex interfacial dynamics. While site-specific heterogeneous stratifications could be tested, such localized sensitivity rankings would

likely introduce representativeness biases when extrapolated across the highly diverse and complex regional stratigraphy of the QTP. Though the three sites (TGL, WQ, GZ) were selected to represent distinct thermal regimes across a broad geographic transect, they cannot fully capture the full spectrum of parameter sensitivities across the highly heterogeneous QTP. Furthermore, the selection of 10 sensitive parameters for calibration was a pragmatic threshold driven by computational constraints. Because the overall parameter sensitivity scores exhibited a gradual continuum without a distinct mathematical breakpoint, expanding the parameter set to include more variables (particularly soil-texture-dependent parameters) would have exponentially increased the dimensionality of the map-based optimization beyond feasible limits. Future work should incorporate sensitivity analysis across diverse environmental gradients, varying temporal scales, and realistic soil columns to better diagnose and improve the physical model structure. Nevertheless, this spatial limitation does not undermine the core findings of this study, as the sensitivity analysis was designed to pragmatically reduce the parameter space for the CaliMAP optimization (Figure 3), rather than to serve as a standalone spatial investigation.

In the CaliMAP, the spatially uniform ±1.5 °C perturbation for TBOT may not fully compensate for existing spatial biases in the default estimation. It is also important to contextualize this limitation against conventional methods. Specifically, site-based calibrations implicitly apply a single, point-calibrated TBOT perturbation uniformly across the entire regional domain. However, the uniform perturbation in CaliMAP represents a pragmatic compromise constrained by current data scarcity, a limitation that can be readily resolved as high-resolution spatial geothermal datasets become available. In contrast, the spatial projection of a local parameter inherent to site-based calibration is a fundamental methodological flaw that cannot be mitigated by future knowledge updates. Furthermore, while the Kappa coefficient utilized in CaliMAP successfully constrained the spatial distribution of permafrost,

this categorical objective function inherently lacks the capacity to constrain the vertical thermal structure or the absolute magnitude of internal temperature biases. As a result, the map-based optimization may overlook compensating errors within the subsurface thermal regime. These limitations are evidenced by the lower accuracy in simulating the GT10m trend compared to the uncalibrated CaliFree baseline (though CaliMAP still outperformed the site-based CaliTGL and CaliTSH). Moving forward, adopting a multi-objective optimization framework that simultaneously integrates horizontal spatial constraints (regional distribution maps) with vertical thermal constraints (local borehole temperature profiles) represents a critical frontier for advancing regional permafrost modeling.

The map-based calibration, involving a significantly larger number of grid cells (> 13,000 in this study), demands considerably more computational resources compared to traditional site-based calibration. However, it is justified given the improved regional accuracy. In fact, there are a few numerical models specifically designed for permafrost (Jafarov et al., 2012; Painter et al., 2016; Westermann et al., 2023), and the model operations largely inherit from the land surface and hydrological modeling community (Elshamy et al., 2020; Ji et al., 2022). However, permafrost possesses characteristics distinct from the shallow land surface, such as substantial thermal inertia (a relic from the Last Glacial Maximum) (Aga et al., 2023; Jin et al., 2020) and significant spatial variability even at local scales (Riseborough et al., 2008), which require specific adaptation in the model and model operation. The computationally intensive simulation should not be regarded as a burden, but rather as a rare opportunity to advance permafrost modeling.

Furthermore, our study acknowledges the limitation of relying on a single atmospheric forcing dataset (ITP-forcing) to drive change simulation (Guo et al., 2017). Although ITP-forcing exhibits leading accuracy compared to other datasets when evaluated against meteorological stations on the QTP (He et al., 2020), the sparse and uneven distribution of weather stations

introduces uncertainties in the forcing data, particularly across such a vast and heterogeneous region (Miao et al., 2024). Most existing meteorological datasets are deterministic and do not account for the uncertainties associated with data acquisition and processing, especially in remote regions (Tang et al., 2023). Consequently, an integrated simulation framework incorporating parameter ensembles, up-to-date forcing ensembles, and potentially model ensembles is essential for simulating and projecting permafrost change to support adaptive management.

5 Conclusions

This study has advanced our capacity to accurately and robustly model permafrost changes by innovatively applying a permafrost map to constrain a modified Noah LSM. Through validation against extensive in-situ observations, the efficacy of the map-based calibration was demonstrated, and its optimal parameter ensemble was used to simulate permafrost changes on the QTP. Our key conclusions are as follows:

1. The effectiveness of the conventional site-based calibration strategy highly depends on the environmental representativeness of the local calibration site, often resulting in poor spatial transferability. Despite a higher computational cost, the map-based calibration effectively circumvents the parameter transferability issue, achieving the best performance overall. It yielded RMSEs of 0.68 m, 0.41 °C and 1.30 °C for annual ALT, TTOP and GT10m, respectively, alongside errors of 0.08 m/10a, 0.36°C/10a and 0.64°C/10a for their respective change rates.

2. Over the past four decades (1980-2018), under an atmospheric warming rate of 0.39 °C/10a, permafrost on the QTP has undergone pervasive thermal degradation. This is evidenced by rising ground temperature (0.10 °C/10a for TTOP, 0.20 °C/10a for MAGT), deepening active layers (0.04 m/10a for ALT), and an upward shift in the depth of zero annual amplitude (−0.52

m/10a for DZAA). The permafrost area decreased from $1.262 \pm 0.048 \times 10^6$ km^2 (95% CI) in 1980 to $1.085 \pm 0.049 \times 10^6$ km^2 in 2018, with a total reduction of $177 \pm 2 \times 10^3$ km^2 and a rate of 44.4×10^3 km^2/10a. This estimated rate of thawing permafrost falls within the lower range of previous estimates.

3. The permafrost change exhibited substantial spatial and temporal heterogeneity across the QTP. The anomalous air cooling in the northwestern plateau, specifically the Karakoram region, during the late 1990s led to localized permafrost development and a temporary deceleration of the overall regional permafrost degradation trend.

Data Availability

The datasets generated and/or analysed during the current study are available in the figshare repository, https://doi.org/10.6084/m9.figshare.28721504

Author contributions

HJ: Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; XW: Investigation, Methodology, Software, Writing – original draft; SZ: Resources, Writing – review & editing, Funding acquisition; ZN: Conceptualization, Formal Analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Acknowledgments

This study was funded by the National Key Research and Development Program of China (No. 2022YFF0711703) and the National Natural Science Foundation of China (No. 42571149, 42171125).

Competing Interests

All authors declare no financial or non-financial competing interests.

References

Aas, K. S., Martin, L., Nitzbon, J., et al. (2019). Thaw processes in ice-rich permafrost landscapes represented with laterally coupled tiles in a land surface model. *The Cryosphere*, *13*(2), 591-609. https://doi.org/10.5194/tc-13-591-2019

Abdelhamed, M. S., Elshamy, M. E., Razavi, S., et al. (2023). Challenges in hydrologic-land surface modeling of permafrost signatures—A Canadian perspective. *Journal of Advances in Modeling Earth Systems*, *15*(3), e2022MS003013. https://doi.org/10.1029/2022MS003013

Aga, J., Boike, J., Langer, M., et al. (2023). Simulating ice segregation and thaw consolidation in permafrost environments with the CryoGrid community model. *The Cryosphere*, *17*(10), 4179-4206. https://doi.org/10.5194/tc-17-4179-2023

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279-298. https://doi.org/10.1002/hyp.3360060305

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

Blöschl, G., P., B. M. F., Antonio, C., et al. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal*, *64*(10), 1141-1158. https://doi.org/10.1080/02626667.2019.1620507

Brown, J., Ferrians Jr, O. J., Heginbottom, J. A., et al. (1997). *Circum-Arctic map of permafrost and ground-ice conditions*. https://doi.org/10.3133/cp45

Campbell, G. S. (1974). A simple method for determining unsaturated conductivity from moisture retention data. *Soil Science*, *117*(6), 311-314. https://doi.org/10.1097/00010694-197406000-00001

Cao, Z., Nan, Z., Hu, J., et al. (2023). A new 2010 permafrost distribution map over the Qinghai-Tibet Plateau based on subregion survey maps: a benchmark for regional permafrost modeling. *Earth System Science Data*, *15*(9), 3905-3930. https://doi.org/10.5194/essd-2022-206

Chadburn, S. E., Burke, E. J., Cox, P. M., et al. (2017). An observation-based constraint on permafrost loss as a function of global warming. *Nature Climate Change*, *7*(5), 340-344. https://doi.org/10.1038/nclimate3262

Chen, F., Janjić, Z., & Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP Mesoscale Eta Model. *Boundary-Layer Meteorology*, *85*(3), 391-421. https://doi.org/10.1023/a:1000531001463

Chen, H., Nan, Z., Zhao, L., et al. (2015). Noah modelling of the permafrost distribution and characteristics in the West Kunlun area, Qinghai-Tibet Plateau, China. *Permafrost and Periglacial Processes*, *26*(2), 160-174. https://doi.org/10.1002/ppp.1841

Chen, X., Long, D., Hong, Y., et al. (2017). Improved modeling of snow and glacier melting by a progressive two-stage calibration strategy with GRACE and multisource data: How snow and glacier meltwater contributes to the runoff of the Upper Brahmaputra River basin? *Water Resources Research*, *53*(3), 2431-2466. https://doi.org/10.1002/2016WR019656

Clayton, L. K., Schaefer, K., Battaglia, M. J., et al. (2021). Active layer thickness as a function of soil water content. *Environmental Research Letters*, *16*(5), 055028. https://doi.org/10.1088/1748-9326/abfa4c

Cuntz, M., Mai, J., Samaniego, L., et al. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *Journal of Geophysical Research: Atmospheres*, *121*(18), 10,676-610,700. https://doi.org/10.1002/2016JD025097

Ehlers, T. A., Chen, D., Appel, E., et al. (2022). Past, present, and future geo-biosphere interactions on the Tibetan Plateau and implications for permafrost. *Earth-Science Reviews*, *234*, 104197. https://doi.org/10.1016/j.earscirev.2022.104197

Ek, M. B., Mitchell, K. E., Lin, Y., et al. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(D22), 8851. https://doi.org/10.1029/2002jd003296

Elshamy, M. E., Princz, D., Sapriza-Azuri, G., et al. (2020). On the configuration and initialization of a large-scale hydrological land surface model to represent permafrost. *Hydrology and Earth System Sciences*, *24*(1), 349-379. https://doi.org/10.5194/hess-24-349-2020

Farinotti, D., Immerzeel, W. W., de Kok, R. J., et al. (2020). Manifestations and mechanisms of the Karakoram glacier Anomaly. *Nature Geoscience*, *13*(1), 8-16. https://doi.org/10.1038/s41561-019-0513-5

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2018MS001453. https://doi.org/10.1029/2018MS001453

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*(2), 657-673. https://doi.org/10.5194/tc-12-657-2018

Guo, D., & Wang, H. (2013). Simulation of permafrost and seasonally frozen ground conditions on the Tibetan Plateau, 1981-2010. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5216-5230. https://doi.org/10.1002/jgrd.50457

Guo, D., Wang, H., & Wang, A. (2017). Sensitivity of historical simulation of the permafrost to different atmospheric forcing data sets from 1979 to 2009. *Journal of Geophysical Research: Atmospheres*, *122*(22), 12269-12284. https://doi.org/10.1002/2017jd027477

Guo, W., Liu, S., Xu, J., et al. (2015). The second Chinese glacier inventory: data, methods and results. *Journal of Glaciology*, *61*(226), 357-372. https://doi.org/10.3189/2015JoG14J209

He, J., Yang, K., Tang, W., et al. (2020). The first high-resolution meteorological forcing dataset for land process studies over China. *Scientific Data*, *7*, 25. https://doi.org/10.1038/s41597-020-0369-y

Hou, X., & Zhang, X. (2021). *Vegetation Map of the People's Republic of China (1:1,000,000)* Plant Data Center of Chinese Academy of Sciences. https://doi.org/10.12282/plantdata.0155

Hu, J., Zhao, S., Nan, Z., et al. (2020). An effective approach for mapping permafrost in a large area using subregion maps and satellite data. *Permafrost and Periglacial Processes*, *31*(4), 548-560. https://doi.org/10.1002/ppp.2068

Hu, W., Ma, W., Yang, Z.-L., et al. (2023). Sensitivity analysis of the Noah-MP land surface model for soil hydrothermal simulations over the Tibetan Plateau. *Journal of Advances in Modeling Earth Systems*, *15*(3), e2022MS003136. https://doi.org/10.1029/2022MS003136

Huang, J., Li, C., Jia, B., et al. (2024). Variation in the permafrost active layer over the Tibetan Plateau during 1980–2020. *Atmospheric and Oceanic Science Letters*, 100536. https://doi.org/10.1016/j.aosl.2024.100536

Jafarov, E. E., Marchenko, S. S., & Romanovsky, V. E. (2012). Numerical modeling of permafrost dynamics in Alaska using a high spatial resolution dataset. *The Cryosphere*, *6*(3), 613-624. https://doi.org/10.5194/tc-6-613-2012

Ji, H., Fu, X., Nan, Z., et al. (2024). An effective medium theory-based unified model for estimating thermal conductivity of unfrozen and frozen soils. *CATENA*, *239*, 107942. https://doi.org/10.1016/j.catena.2024.107942

Ji, H., Nan, Z., Hu, J., et al. (2022). On the spin-up strategy for spatial modeling of permafrost dynamics: A case study on the Qinghai-Tibet Plateau. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002750. https://doi.org/10.1029/2021MS002750

Jiang, H., Shen, Y., Li, Y., et al. (2024a). Openbox: A Python toolkit for generalized black-box optimization. *Journal of Machine Learning Research*, *25*(120), 1-11. https://jmlr.org/papers/v25/23-0537.html

Jiang, H., Yi, Y., Yang, K., et al. (2024b). Soil freeze/thaw dynamics strongly influences runoff regime in a Tibetan permafrost watershed: Insights from a process-based model. *CATENA*, *243*, 108182. https://doi.org/10.1016/j.catena.2024.108182

Jin, H., Vandenberghe, J., Luo, D., et al. (2020). Quaternary permafrost in China: Framework and discussions. *Quaternary*, *3*(4). https://doi.org/10.3390/quat3040032

Jin, H., Zhao, L., Wang, S., et al. (2006). Thermal regimes and degradation modes of permafrost along the Qinghai-Tibet Highway. *Science in China Series D-Earth Sciences*, *49*(11), 1170-1183. https://doi.org/10.1007/s11430-006-2003-z

Kurylyk, B. L., & Watanabe, K. (2013). The mathematical representation of freezing and thawing processes in variably-saturated, non-deformable soils. *Advances in Water Resources*, *60*, 160-177. https://doi.org/10.1016/j.advwatres.2013.07.016

Li, X., Wu, T., Zhu, X., et al. (2020). Improving the Noah-MP model for simulating hydrothermal regime of the active layer in the permafrost regions of the Qinghai-Tibet Plateau. *Journal of Geophysical Research: Atmospheres*, *125*(16), 1-20. https://doi.org/10.1029/2020JD032588

Luo, J., Niu, F., Lin, Z., et al. (2022). Inventory and Frequency of Retrogressive Thaw Slumps in Permafrost Region of the Qinghai–Tibet Plateau. *Geophysical Research Letters*, *49*(23), e2022GL099829. https://doi.org/10.1029/2022GL099829

Mei, Y., Mai, J., Do, H. X., et al. (2023). Can hydrological models benefit from using global soil moisture, evapotranspiration, and runoff products as calibration targets? *Water Resources Research*, *59*(2), e2022WR032064. https://doi.org/10.1029/2022WR032064

Miao, C., Immerzeel, W. W., Xu, B., et al. (2024). Understanding the Asian water tower requires a redesigned precipitation observation strategy. *Proceedings of the National Academy of Sciences*, *121*(23), e2403557121. https://doi.org/10.1073/pnas.2403557121

Mu, C., Li, K., Liu, S., et al. (2025). Recent intensified riverine CO2 emission across the Northern Hemisphere permafrost region. *Nature Communications*, *16*(1), 3616. https://doi.org/10.1038/s41467-025-58716-3

Nicolsky, D. J., & Romanovsky, V. E. (2018). Modeling Long-Term Permafrost Degradation. *Journal of Geophysical Research: Earth Surface*, *123*(8), 1756-1771. https://doi.org/10.1029/2018JF004655

Painter, S. L., Coon, E. T., Atchley, A. L., et al. (2016). Integrated surface/subsurface permafrost thermal hydrology: Model formulation and proof-of-concept simulations. *Water Resources Research*, *52*(8), 6062-6077. https://doi.org/10.1002/2015WR018427

Qin, Y., Wu, T., Zhao, L., et al. (2017). Numerical modeling of the active layer thickness and permafrost thermal state across Qinghai-Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *122*(21), 11604-11620. https://doi.org/10.1002/2017jd026858

Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, *52*(1), 423-439. https://doi.org/10.1002/2015WR017558

Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, *52*(1), 440-455. https://doi.org/10.1002/2015WR017559

Razavi, S., & Gupta, H. V. (2019). A multi-method Generalized Global Sensitivity Matrix approach to accounting for the dynamical nature of earth and environmental systems models. *Environmental Modelling & Software*, *114*, 1-11. https://doi.org/10.1016/j.envsoft.2018.12.002

Riseborough, D., Shiklomanov, N., Etzelmuller, B., et al. (2008). Recent advances in permafrost modelling. *Permafrost and Periglacial Processes*, *19*(2), 137-156. https://doi.org/10.1002/ppp.615

Rosero, E., Yang, Z.-L., Wagener, T., et al. (2010). Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. *Journal of Geophysical Research: Atmospheres*, *115*(D3). https://doi.org/10.1029/2009JD012035

Schaake, J. C., Koren, V. I., Duan, Q.-Y., et al. (1996). Simple water balance model for estimating runoff at different spatial and temporal scales. *Journal of Geophysical Research: Atmospheres*, *101*(D3), 7461-7475. https://doi.org/10.1029/95jd02892

Shen, M. G., Piao, S. L., Jeong, S. J., et al. (2015). Evaporative cooling over the Tibetan Plateau induced by vegetation growth. *Proceedings of the National Academy of Sciences*, *112*(30), 9299-9304. https://doi.org/10.1073/pnas.1504418112

Shen, T., Jiang, P., Ju, Q., et al. (2024). Permafrost on the Tibetan Plateau is degrading: Historical and projected trends. *Journal of Hydrology*, *628*, 130501. https://doi.org/10.1016/j.jhydrol.2023.130501

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

Song, W., Shi, L., He, L., et al. (2025). Deep Learning Identification of the Governing Equation for Water Flow in Heterogeneous Soils From Data. *Water Resources Research*, *61*(3), e2024WR037786. https://doi.org/10.1029/2024WR037786

Tang, G., Clark, M. P., Knoben, W. J. M., et al. (2023). The Impact of Meteorological Forcing Uncertainty on Hydrological Modeling: A Global Analysis of Cryosphere Basins. *Water Resources Research*, *59*(6), e2022WR033767. https://doi.org/10.1029/2022WR033767

Tsai, W.-P., Feng, D., Pan, M., et al. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications*, *12*(1), 5988. https://doi.org/10.1038/s41467-021-26107-z

van Everdingen, R. O. (1998). *Multi-language glossary of permafrost and related ground-ice terms*. http://globalcryospherewatch.org/reference/glossary_docs/Glossary_of_Permafrost_and_Ground-Ice_IPA_2005.pdf

Wang, T., Wu, T., Wang, P., et al. (2019). Spatial distribution and changes of permafrost on the Qinghai-Tibet Plateau revealed by statistical models during the period of 1980 to 2010. *Science of The Total Environment*, *650*, 661-670. https://doi.org/10.1016/j.scitotenv.2018.08.398

Wang, T., Yang, D., Yang, Y., et al. (2023). Pervasive permafrost thaw exacerbates future risk of water shortage across the Tibetan Plateau. *Earth's Future*, e2022EF003463. https://doi.org/10.1029/2022EF003463

Wang, X., Pang, G., & Yang, M. (2018). Precipitation over the Tibetan Plateau during recent decades: a review based on observations and simulations. *International Journal of Climatology*, *38*(3), 1116-1131. https://doi.org/10.1002/joc.5246

Wang, Z., Wang, Q., Zhao, L., et al. (2016). Mapping the vegetation distribution of the permafrost zone on the Qinghai-Tibet Plateau. *Journal of Mountain Science*, *13*(6), 1035-1046. https://doi.org/10.1007/s11629-015-3485-y

Westermann, S., Ingeman-Nielsen, T., Scheer, J., et al. (2023). The CryoGrid community model (version 1.0) – a multi-physics toolbox for climate-driven simulations in the terrestrial cryosphere. *Geoscientific Model Development*, *16*(9), 2607-2647. https://doi.org/10.5194/gmd-16-2607-2023

Wetzel, P. J., & Chang, J.-T. (1988). Evapotranspiration from nonuniform surfaces: A first approach for short-term numerical weather prediction. *Monthly Weather Review*, *116*(3), 600-621. https://doi.org/10.1175/1520-0493(1988)116<0600:EFNSAF>2.0.CO;2

Wu, J. C., Sheng, Y., Wu, Q. B., et al. (2010). Processes and modes of permafrost degradation on the Qinghai-Tibet Plateau. *Science China-Earth Sciences*, *53*(1), 150-158. https://doi.org/10.1007/s11430-009-0198-5

Wu, X., & Nan, Z. (2016, 10-15 July 2016). *A multilayer soil texture dataset for permafrost modeling over Qinghai-Tibetan Plateau* 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China. https://doi.org/10.1109/igarss.2016.7730283

Wu, X., Nan, Z., Zhao, S., et al. (2018). Spatial modeling of permafrost distribution and properties on the Qinghai-Tibet Plateau. *Permafrost and Periglacial Processes*, *29*(2), 86-99. https://doi.org/10.1002/ppp.1971

Xu, F., Zhang, G., Woolway, R. I., et al. (2024). Widespread societal and ecological impacts from projected Tibetan Plateau lake expansion. *Nature Geoscience*. https://doi.org/10.1038/s41561-024-01446-w

Yao, T., Bolch, T., Chen, D., et al. (2022). The imbalance of the Asian water tower. *Nature Reviews Earth & Environment*. https://doi.org/10.1038/s43017-022-00299-4

Zhang, G., Nan, Z., Zhao, L., et al. (2021). Qinghai-Tibet Plateau wetting reduces permafrost thermal responses to climate warming. *Earth and Planetary Science Letters*, *562*, 116858. https://doi.org/10.1016/j.epsl.2021.116858

Zhang, G., Yao, T., Shum, C. K., et al. (2017). Lake volume and groundwater storage variations in Tibetan Plateau's endorheic basin. *Geophysical Research Letters*, *44*(11), 5550-5560. https://doi.org/10.1002/2017GL073773

Zhao, L., Zou, D., Hu, G., et al. (2021). A synthesis dataset of permafrost thermal state for the Qinghai–Tibet (Xizang) Plateau, China. *Earth System Science Data*, *13*(8), 4207-4218. https://doi.org/10.5194/essd-13-4207-2021

Zhao, Y., Nan, Z., Cao, Z., et al. (2023). Evaluation of parameterization schemes for matric potential in frozen soil in land surface models: A modeling perspective. *Water Resources Research*, *59*(6), e2023WR034644. https://doi.org/10.1029/2023WR034644

Zheng, G., Yang, Y., Yang, D., et al. (2020). Remote sensing spatiotemporal patterns of frozen soil and the environmental controls over the Tibetan Plateau during 2002–2016. *Remote Sensing of Environment*, *247*, 111927. https://doi.org/10.1016/j.rse.2020.111927

Zhu, Z., Bi, J., Pan, Y., et al. (2013). Global data sets of vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference

Vegetation Index (NDVI3g) for the period 1981 to 2011. *Remote Sensing*, *5*(2), 927-948. https://doi.org/10.3390/rs5020927

Supporting Information for

Constrained simulation of permafrost thermal changes from 1980 to 2018 on the Qinghai-Tibet Plateau

Hailong Ji[1,2], Xiaobo Wu[3], Shuping Zhao[2], Zhuotong Nan[1,2,4*]

[1] Yangtze River Delta Urban Wetland Ecosystem National Field Scientific Observation and Research Station, School of Environmental and Geographical Sciences, Shanghai Normal University, Shanghai, 200234, China

[2] State Key Laboratory of Climate System Prediction and Risk Management, Nanjing Normal University, Nanjing, 210023, China

[3] College of Resources, Sichuan Agricultural University, Chengdu, 611130, China

[4] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China

*Correspondent author: Zhuotong Nan, nanzt@shnu.edu.cn

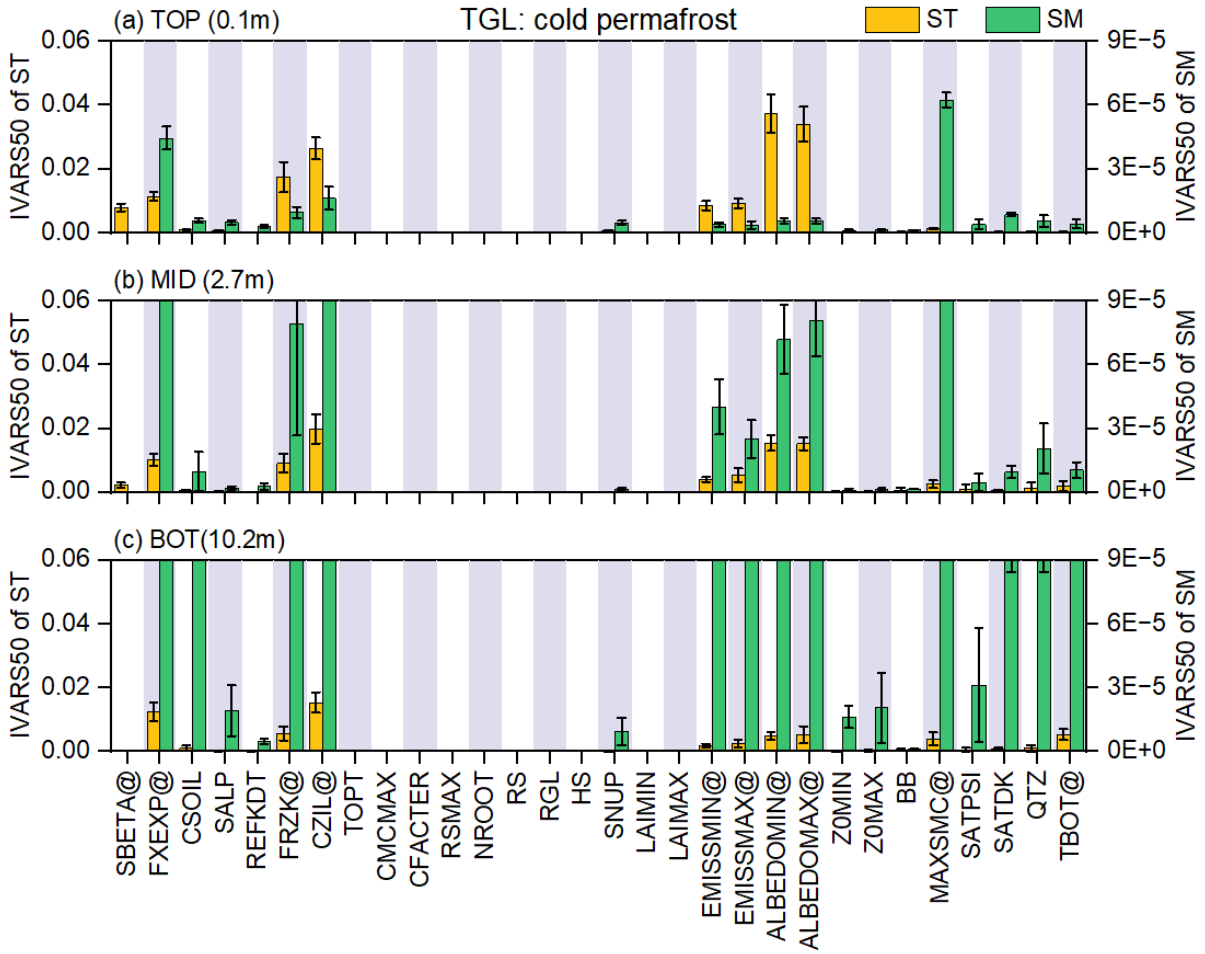

Figure S1 IVARS50-based sensitivity of simulated soil temperature (ST) and moisture (SM) to model parameters at varying depths (TOP: 0.1 m, MID: 2.7 m, BOT: 10.2 m) for the TGL modeling cell, which represents cold permafrost. Parameters makred with "@" are those identified as among the top 10 most sensitive parameters in the overall ranking shown in Figure 3.

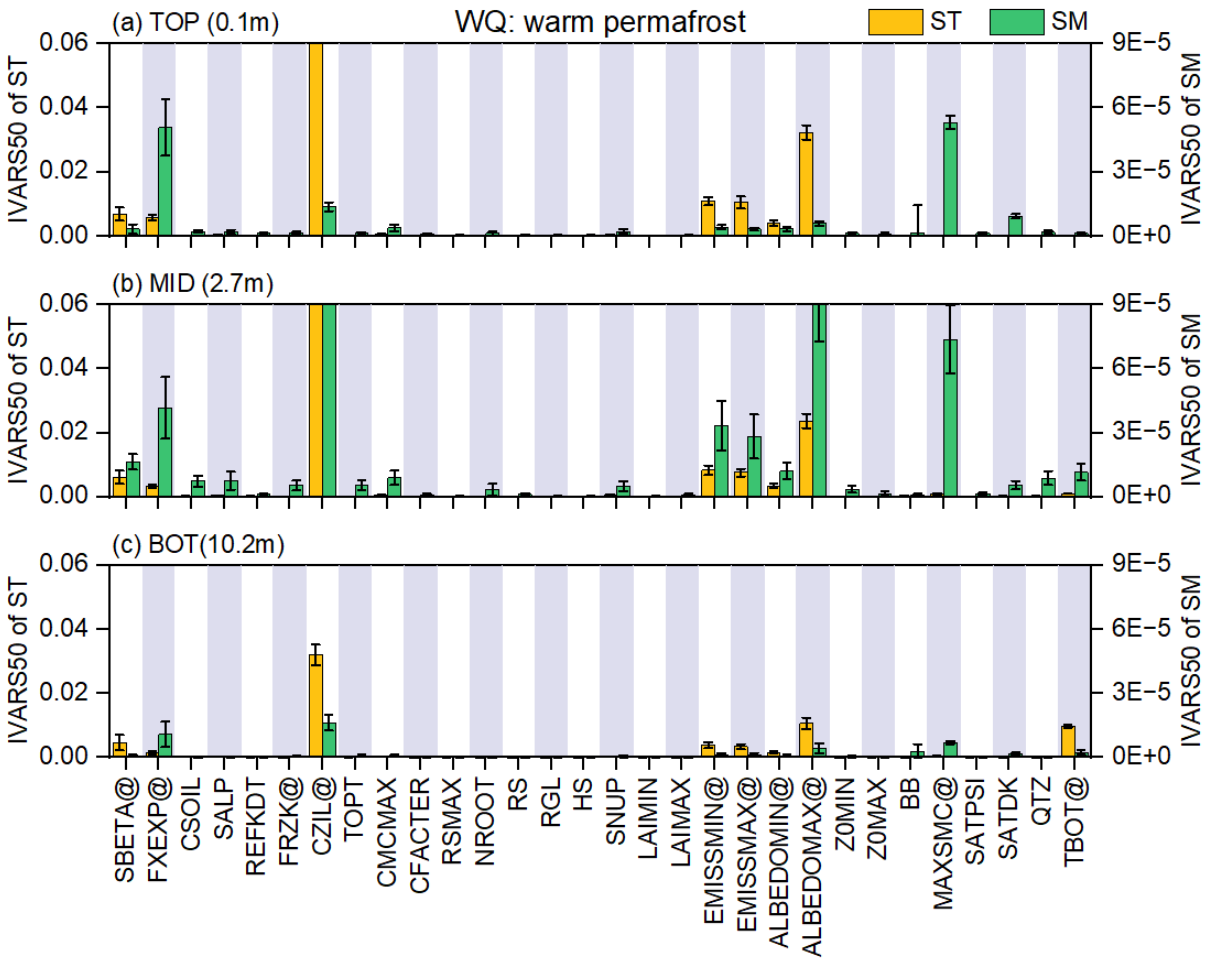

Figure S2 Same as Figure S1, but for the WQ cell, representing warm permafrost conditions.

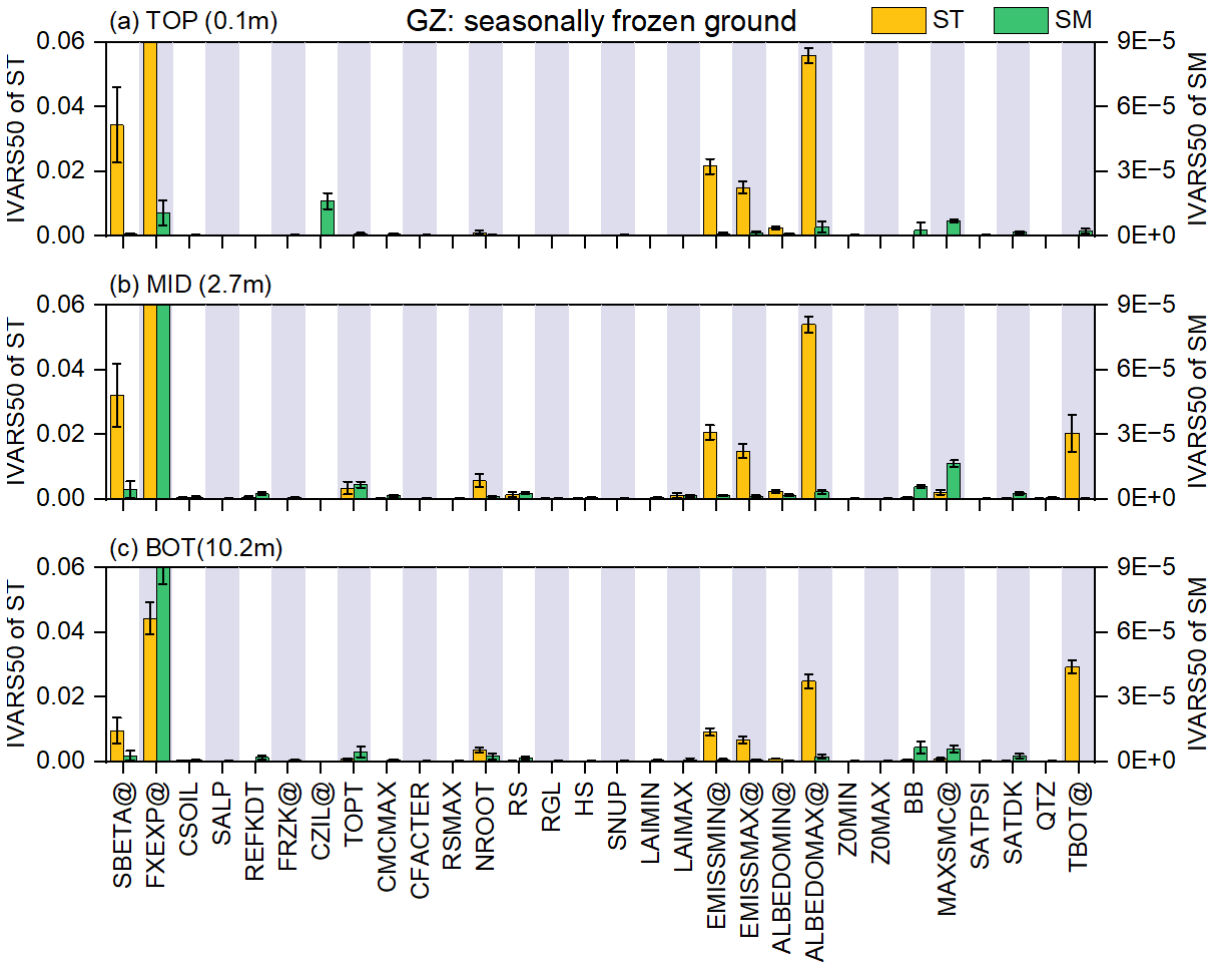

Figure S3 Same as Figure S1, but for the GZ cell, representing seasonally frozen ground conditions.

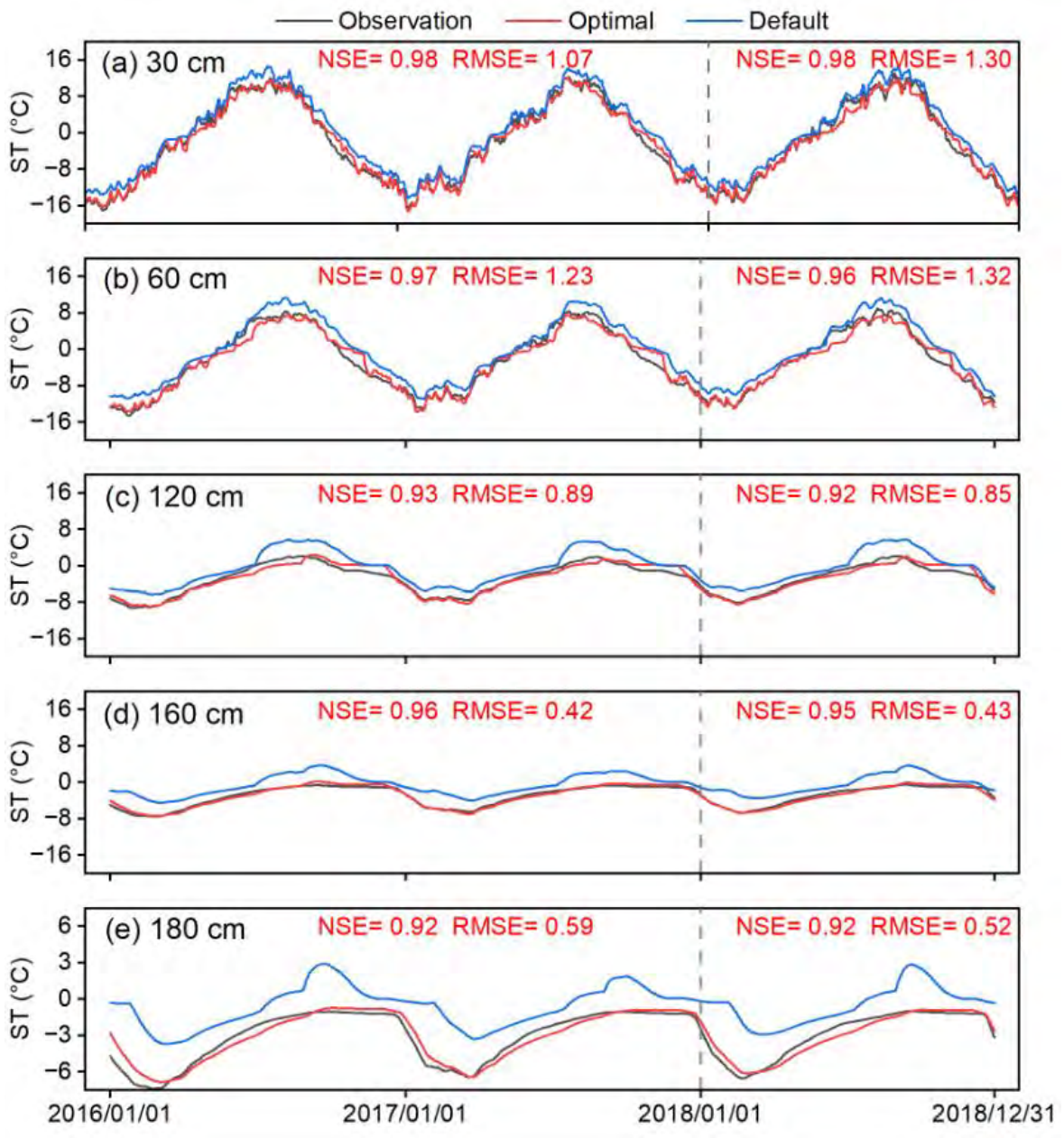

Figure S4 Calibration (2016/01/01-2017/12/31) and validation (2018/01/01-2018/12/31) of soil temperature (ST) in CaliTSH at the TSHAL site. The figure shows observed soil temperature along with simulations using the optimal parameter set (CaliTSH) and default parameters (CaliFree) at five depths (30 cm, 60 cm, 120 cm, 160 cm, and 180 cm). The Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) values for CaliTSH are indicated for each depth and both periods.

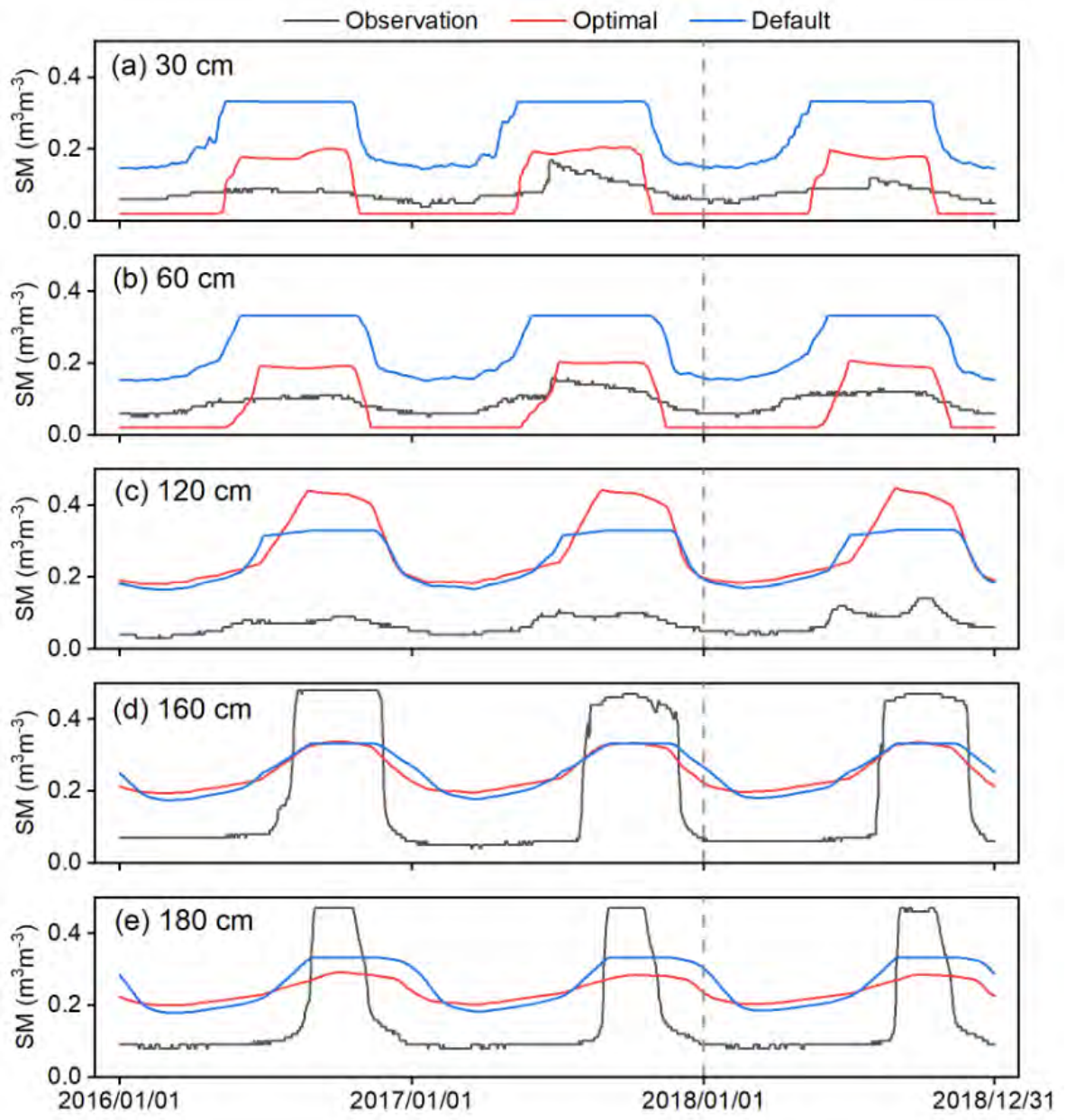

Figure S5 Soil moisture dynamics at the TSHAL site simulated using the optimal parameters from the soil temperature-calibrated CaliTSH experiment. The figure displays observed soil moisture contents alongside simulations from the optimal and default parameter sets at five depths (30, 60, 120, 160, and 180 cm) for the period 2016-2018. Soil moisture was not targeted during calibration in CaliTSH.

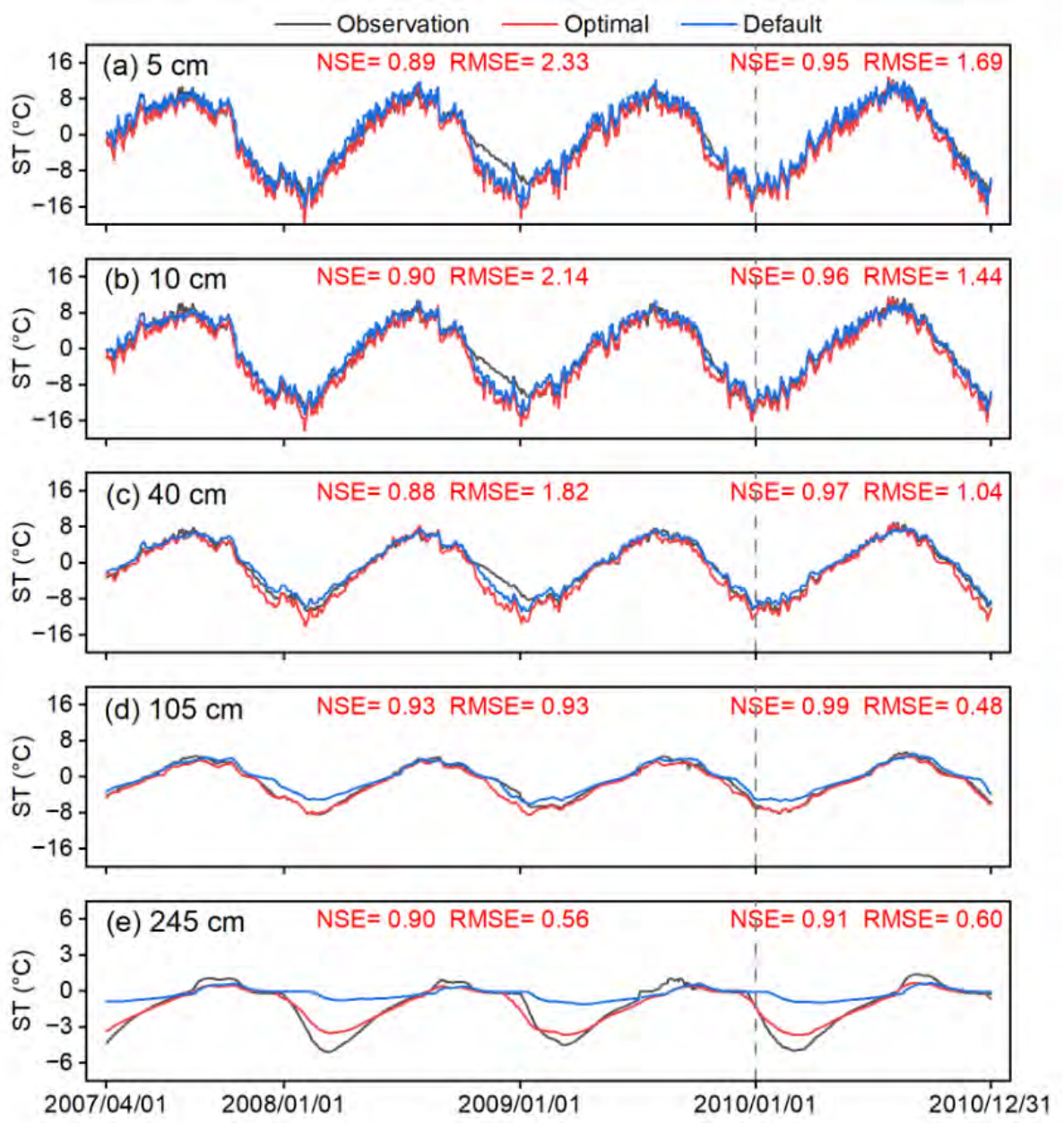

Figure S6 Same as Figure S4, but for CaliTGL at the TGL(QT04) site, with a calibration period of 2007/04/01-2009/12/31 and a validation period of 2010/01/01-2010/12/31.

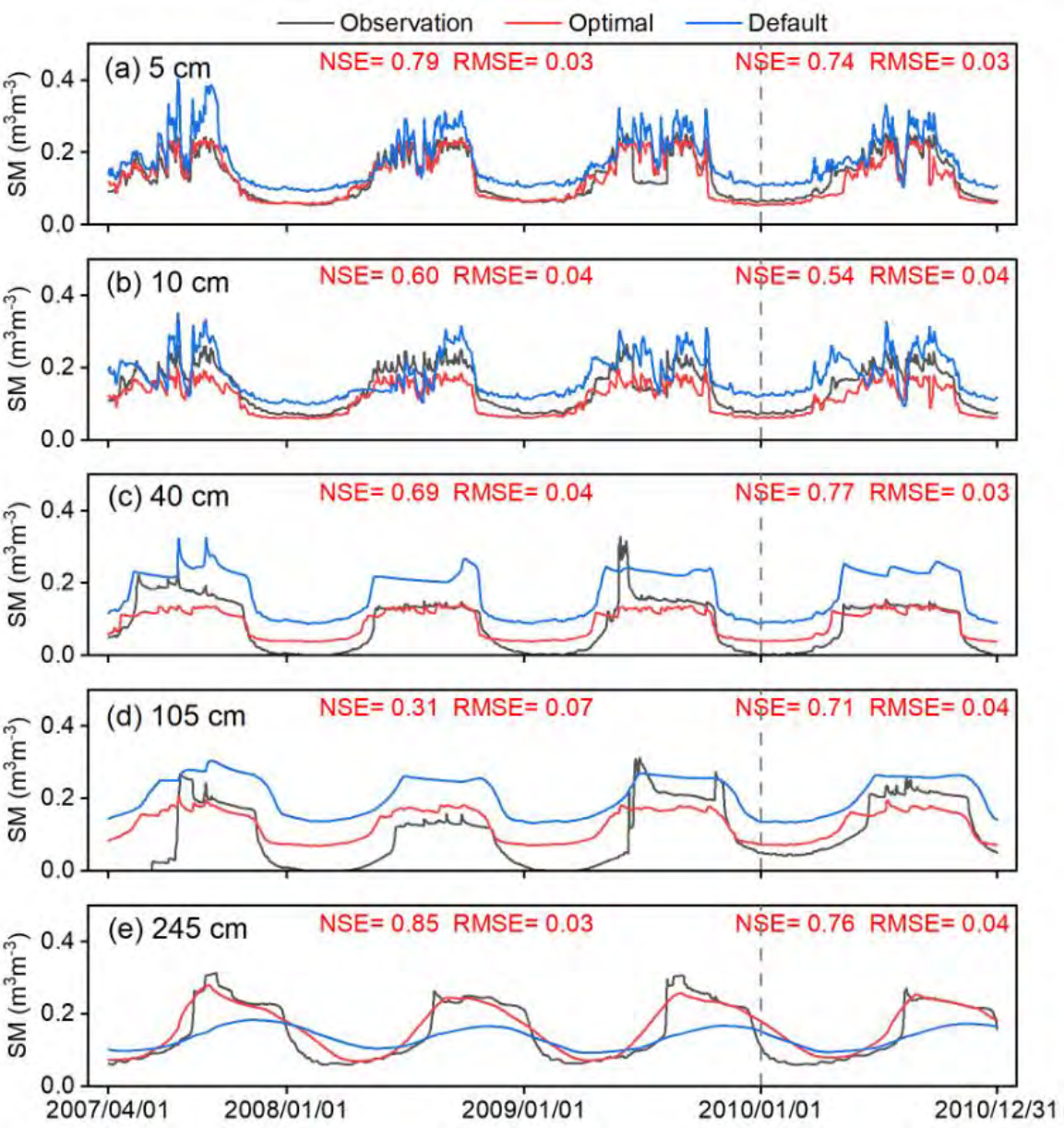

Figure S7 Same as Figure S5, but for CaliTGL at the TGL(QT04) site.

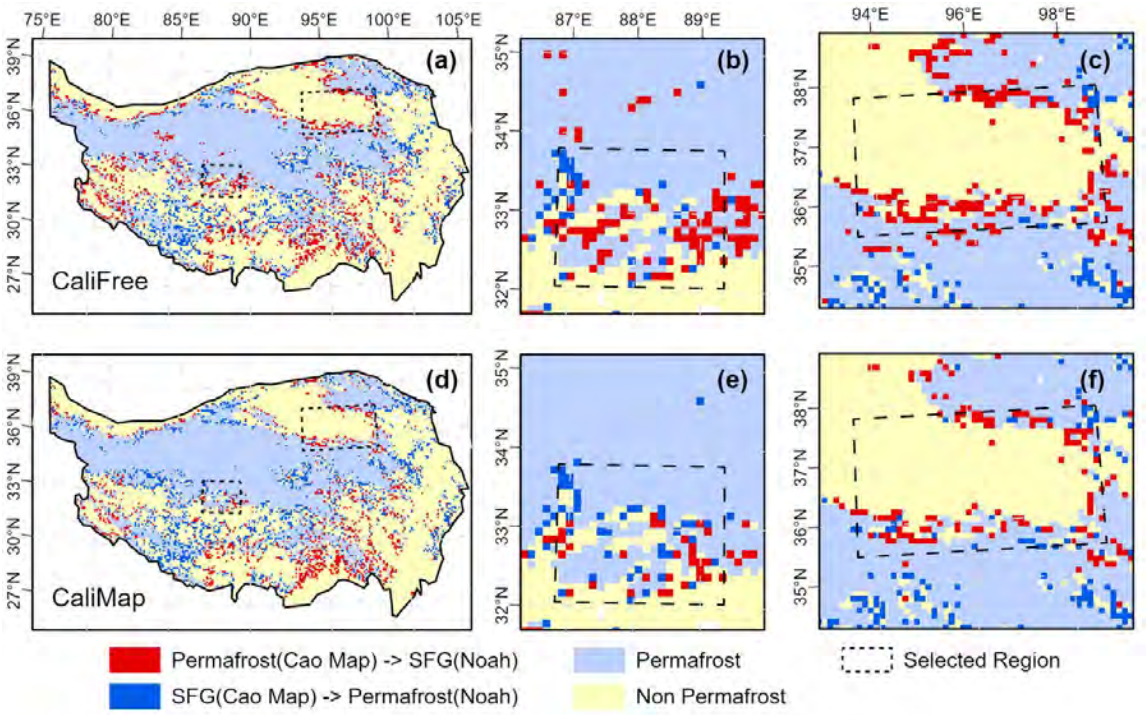

Figure S8 Spatial differences between simulated permafrost distribution and the Cao map. The first row (a-c) presents results from CaliFree (default parameters), and the second row (d-f) presents results from CaliMAP (optimized parameters). The second and third columns show zoomed-in sections of (a) and (d). SFG: seasonally frozen ground.

Table S1 Statistics on vegetation types and soil textures at sites and sensitive cells.

Site/Area	Type	Code	Name	Number[#]	Percentage
TGL(QT04)	Vegetation	7	Grassland	1	100%
	Soil	7	Loam	7	38.9%
		12	Sand	3	16.7%
		13	Gravel	7	38.9%
		--	Other	1	5.5%
TSHAL	Vegetation	19	Barren or sparsely vegetated	1	100%
	Soil	11	Loamy sand	3	16.7%
		12	Sand	3	16.7%
		14	Bedrock	12	66.6%
Sensitive cells	Vegetation	7	Grassland	5435	40.2%
		8	Shrubland	1049	7.8%
		9	Mixed shrubland/grassland	3762	27.8%

		19	Barren or sparsely vegetated	1672	12.4%
		--	Other	1156	11.8%
	Soil	7	Loam	108852	44.7%
		9	Sandy loam	19246	7.9%
		12	Sand	33705	13.9%
		13	Gravel	13839	5.7%
		14	Bedrock	58902	24.2%
		--	Other	8744	3.6%

In our configuration of the modified Noah LSM, each modeling cell was assigned a single vegetation type and a vertically heterogeneous 18-layer soil texture. The set of sensitive cells identified for calibration comprised more than 13,000 cells.