# Challenges in hydrologic-land surface modelling of permafrost signatures - Impacts of parameterization on model fidelity under the uncertainty of forcing

Mohamed S. Abdelhamed[1], Mohamed Elshamy[2], Saman Razavi[1], and Howard Wheater[1]

[1]University of Saskatchewan
[2]Global Institute for Water Security

November 21, 2022

**Abstract**

Permafrost plays an important role in the hydrology of arctic/subarctic regions. However, permafrost thaw/degradation has been observed over recent decades in the Northern Hemisphere and is projected to accelerate. Hence, understanding the evolution of permafrost areas is urgently needed. Land surface models (LSMs) are well-suited for predicting permafrost dynamics due to their physical basis and large-scale applicability. However, LSM application is challenging because of the large number of model parameters and the complex memory of state variables. Significant interactions among the underlying processes and the paucity of observations of thermal/hydraulic regimes add further difficulty. This study addresses the challenges of LSM application by evaluating the uncertainty due to meteorological forcing, assessing the sensitivity of simulated permafrost dynamics to LSM parameters, and highlighting issues of parameter identifiability. Modelling experiments are implemented using the MESH-CLASS framework. The VARS sensitivity analysis and traditional threshold-based identifiability analysis are used to assess various aspects of permafrost dynamics for three regions within the Mackenzie River Basin. The study shows that the modeller may face significant trade-offs when choosing a forcing dataset as some datasets enable the representation of some aspects of permafrost dynamics, while being inadequate for others. The results also emphasize the high sensitivity of various aspects of permafrost simulation to parameters controlling surface insulation and soil texture; a detailed list of influential parameters is presented. Identifiability analysis reveals that many of the most influential parameters for permafrost simulation are unidentifiable. These conclusions will hopefully inform future efforts in data collection and model parametrization.

# Challenges in hydrologic-land surface modelling of permafrost signatures – Impacts of parameterization on model fidelity under the uncertainty of forcing

**Mohamed S. Abdelhamed[1,2,3], Mohamed E. Elshamy[1,4], Saman Razavi[1,2,5], Howard S. Wheater[1,4,5,6]**

[1]Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

[2]Department of Civil and Geological Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

[3]Department of Irrigation and Hydraulics Engineering, Cairo University, Giza, Egypt.

[4]Centre for Hydrology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

[5]School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

[6]Department of Civil and Environmental Engineering, Imperial College London, London, United Kingdom.

**Corresponding author:** Mohamed Abdelhamed (Mohamed.abdelhamed@usask.ca)

**Running Title**: Parameterization of LSMs for Permafrost

**Keywords:** Permafrost, Land surface models, sensitivity, uncertainty, identifiability, Mackenzie River Basin.

**Key Points:**

- There are significant uncertainties in climate forcing datasets that affect the fidelity of permafrost simulations using Land Surface Models.
- The quality of simulated permafrost signatures is primarily controlled by heat insulation and runoff generation parameters.
- Various highly influential model parameters are non-identifiable, leading to significant uncertainty in simulated permafrost characteristics.

## Abstract

Permafrost plays an important role in the hydrology of arctic/subarctic regions. However, permafrost thaw/degradation has been observed over recent decades in the Northern Hemisphere and is projected to accelerate. Hence, understanding the evolution of permafrost areas is urgently needed. Land surface models (LSMs) are well-suited for predicting permafrost dynamics due to their physical basis and large-scale applicability. However, LSM application is challenging because of the large number of model parameters and the complex memory of state variables. Significant interactions among the underlying processes and the paucity of observations of thermal/hydraulic regimes add further difficulty. This study addresses the challenges of LSM application by evaluating the uncertainty due to meteorological forcing, assessing the sensitivity of simulated permafrost dynamics to LSM parameters, and highlighting issues of parameter identifiability. Modelling experiments are implemented using the MESH-CLASS framework. The VARS sensitivity analysis and traditional threshold-based identifiability analysis are used to assess various aspects of permafrost dynamics for three regions within the Mackenzie River Basin. The study shows that the modeller may face significant trade-offs when choosing a forcing dataset as some datasets enable the representation of some aspects of permafrost dynamics, while being inadequate for others. The results also emphasize the high sensitivity of various aspects of permafrost simulation to parameters controlling surface insulation and soil texture; a detailed list of influential parameters is presented. Identifiability analysis reveals that many of the most influential parameters for permafrost simulation are unidentifiable. These conclusions will hopefully inform future efforts in data collection and model parametrization.

## Plain Language Summary

Permafrost (frozen ground for at least two years) is one of several elements that control the rate and magnitude of current global warming. Permafrost plays a critical role in the dynamics of water, heat, and carbon over vast areas globally. For more credible climate/hydrology modelling, it is necessary to assess the ability of available models to reliably reproduce observed permafrost characteristics before using them to evaluate future scenarios. Using a land surface model for different permafrost regions in Canada this study examined three challenges: 1) quantifying the impact of uncertainty in climate forcing data on permafrost simulation, 2) identifying the key parameters that control the quality of permafrost simulation, and 3) assessing the appropriateness of current model structures to reproduce observed permafrost characteristics in the context of parameter uncertainty. In selecting a forcing dataset, permafrost

characteristics exhibited significant trade-offs. The parameters with a large influence on permafrost simulation were identified for the different study areas, but due to model complexity, finding unique values for them was difficult. Several findings were presented to guide further land surface model development, and hence reduce errors in weather/climate modelling.

## 1 Introduction

Permafrost, defined as ground that stays at or below 0°C for at least two years (Everdingen, 1998), plays a central role in the hydrology of arctic and subarctic regions (Dobinski, 2011; Walvoord & Kurylyk, 2016). Permafrost underlies around one-quarter of land in the Northern Hemisphere and one-half of Canada (Obu et al., 2019; Yinsuo Zhang et al., 2008). Several studies have reported increased permafrost temperature over recent decades (*e.g.* Barros et al., 2014; Harris et al., 2009; Meredith et al., 2020; Pan et al., 2016) and projected accelerated temperature rises by 2100 (Burke et al., 2020; Lawrence et al., 2012; McGuire et al., 2018). Such significant change has major implications for hydrological and biogeochemical cycles (Schuur et al., 2015; Walvoord & Kurylyk, 2016). For instance, permafrost thaw can affect the partitioning of water fluxes and stores, thermokarst formation and land subsidence, wildfire occurrence and other ecosystem changes, and streamflow seasonality (Andresen et al., 2020; Dobinski, 2011; Gibson et al., 2018; Hjort et al., 2018; Kokelj & Jorgenson, 2013; Nelson et al., 2002; Schuur et al., 2015; Walvoord & Kurylyk, 2016). Moreover, permafrost stores twice the amount of carbon in the atmosphere, and its release (in the form of carbon dioxide and methane) is likely to have a positive feedback to the global climate and the pace of warming (Burke et al., 2020; McGuire et al., 2018; Schuur et al., 2015).

Earth system models (ESMs) are valuable tools for investigating the potential impacts of climate change on hydrologic and atmospheric conditions. They typically represent land surface processes using a land surface model (LSM), which provides lower boundary conditions to the atmospheric processes modelled within an ESM framework. LSMs have advanced significantly over recent decades through extensive improvements in process representation and enhanced resolution (Prentice et al., 2015; Sellers et al., 1997). The coupled simulation of heat and water across the soil-vegetation-atmosphere interface is a critical feature for permafrost as it accounts for heat transfer with phase change (Jafarov et al., 2012; Riseborough et al., 2008). Efforts to improve permafrost representation have included (but are not limited to) deeper soil configurations, to eliminate the impact of uncertain lower boundary conditions and provide larger thermal memory (Alexeev et al., 2007; Nicolsky et al., 2007), enhanced representation of snow, canopy, and organic soil (including peat and moss) processes that control/regulate thermal insulation of permafrost (Chadburn et al., 2015; Lawrence & Slater, 2008; Wu et al., 2016; Yokohata et al., 2020), and inclusion of vegetation dynamics and carbon-pool processes (Chadburn et al., 2015; Melton et al., 2019). These have reduced the biases in climate projections, as shown by Burke et al. (2020). Therefore, LSMs are well-suited for simulating the major hydrological processes in permafrost regions, having an appropriate physical basis and being applicable at different assessment scales.

Despite these advances, building a high-fidelity model is challenging. The paucity of observational data on permafrost hydraulic and thermal regimes limits the representation of permafrost spatial heterogeneity over large domains (Chadburn et al., 2015; Lamontagne-Hallé et al., 2020; Obu et al., 2019). Further, initializing the model prognostic states is problematic. This is commonly achieved by spinning up the models to reach a set of states that are consistent with the 'transient' climate (Abdelhamed et al., 2021; Elshamy et al., 2020; Sapriza-Azuri et al., 2018), or based on realistic field observations (*e.g.* soil moisture and temperature), if available. The spin-up involves forcing the model with a single (actual or synthetic) year or multiple years of meteorological data repeated in a loop many times, or running the model for a long-enough transient period. Chen and Dudhia (2001) and Rodell et al. (2005) highlighted the biases in surface energy/water flux partitioning that can be introduced due to the improper initialization of state variables.

The structural inadequacy/complexity of LSMs introduces an additional simulation burden. Structural inadequacies of LSMs associated with neglect or oversimplification of permafrost processes (*e.g.* taliks, thermokarst, and aggradation/degradation) complicates model development for permafrost regions, especially those characterized by high heterogeneity (Aas et al., 2019; Devoie et al., 2019; Elshamy et al., 2020). Furthermore, current LSMs have many significant process interactions and contain a large number of free parameters (Prentice et al., 2015). While most of these parameters have a physical meaning, they are usually interpreted and measured at point-scale. In the model they serve as "effective parameters" intended to represent the spatial heterogeneity of the system, and therefore, their feasible ranges can be wide and lead to unrealistic model simulations (Haghnegahdar et al., 2017). In conjunction with the improved realism of process representation in LSMs/ESMs, model complexity and dimensionality have increased considerably, making such models more prone to issues of parameter non-uniqueness (Beven, 2006; Guillaume et al., 2019; Prentice et al., 2015). In general, the development and testing of LSMs, including in permafrost regions, have historically focused on streamflows (more frequently), ET and soil moisture (less frequently) (Yassin et al., 2017). The credible representation of the thermal dynamics of the soil column in cold regions has received significantly less attention, while it directly controls those other variables.

Complex LSMs require a wider spectrum of meteorological variables at finer spatial/temporal scales than simple hydrologic models, which often require limited forcing variables (*e.g.* precipitation and temperature/evapotranspiration) at daily and basin-averaged scales. The input forcing (*e.g.* hydro-meteorological data) uncertainty is typically in the range of 10%-40% (McMillan et al., 2018), and failure to consider such critical uncertainty can lead to unrealistic/biased parameter estimation and misleading water/energy balance calculations. Cold regions are characterized by sparse observational networks, especially in higher latitudes and altitudes, and suffer from inaccuracies related to cold-climate processes (Asong et al., 2020; Wong et al., 2017), which limits the applicability of ground-based observations. On the other hand, remote sensing and model-based forcing products are prone to different sources of uncertainty triggered by data acquisition, processing, rescaling, and imprecisions. Bias-correction (*e.g.* mean-shifting/scaling or quantile mapping) and the number of considered meteorological variables collectively play a pivotal role in the quality (and maintaining cross-correlation structure) of the candidate grid-based forcing dataset.

Here, we examine the capability of different gridded climate datasets to reproduce observed permafrost dynamics under model parameter uncertainty, to aid in selection of the best candidate dataset and highlight the uncertainty propagated due to external forcing. We endorse the formal application of Global Sensitivity Analysis (GSA) (Razavi et al., 2021) as a cornerstone tool for model development by presenting its value for understanding model behaviour, identifying important parameters and characterizing parameter uncertainty. Further, we investigate the identifiability of model parameters, emphasizing parameters with high sensitivity, to ensure model fidelity and underscore the associated structural issues in such models regarding permafrost simulations. **Fig. 1** provides a graphical presentation of the employed methods and expected outcomes of the current study. Our research addresses three specific objectives:

1. Which forcing datasets enable the model to represent important signatures of permafrost dynamics given the uncertainty in model parameters? And do different signatures impose trade-offs in the choice of forcing datasets?
2. Which model parameters are primarily responsible for uncertainty in predicting different signatures of permafrost dynamics? And to what extent will reducing uncertainty in those parameters lead to a reduction of uncertainty in those predictions?
3. Can existing data reduce the uncertainty in those model parameters of primary importance? And which parameters should be the target of future research to reduce uncertainty in predicting permafrost dynamics?

The remainder of the article is organized as follows: **Section 2** presents our methods, case study and model implementation. **Section 3** reports the results of the experiments, and the article ends with a summary and conclusions in **Section 4**.

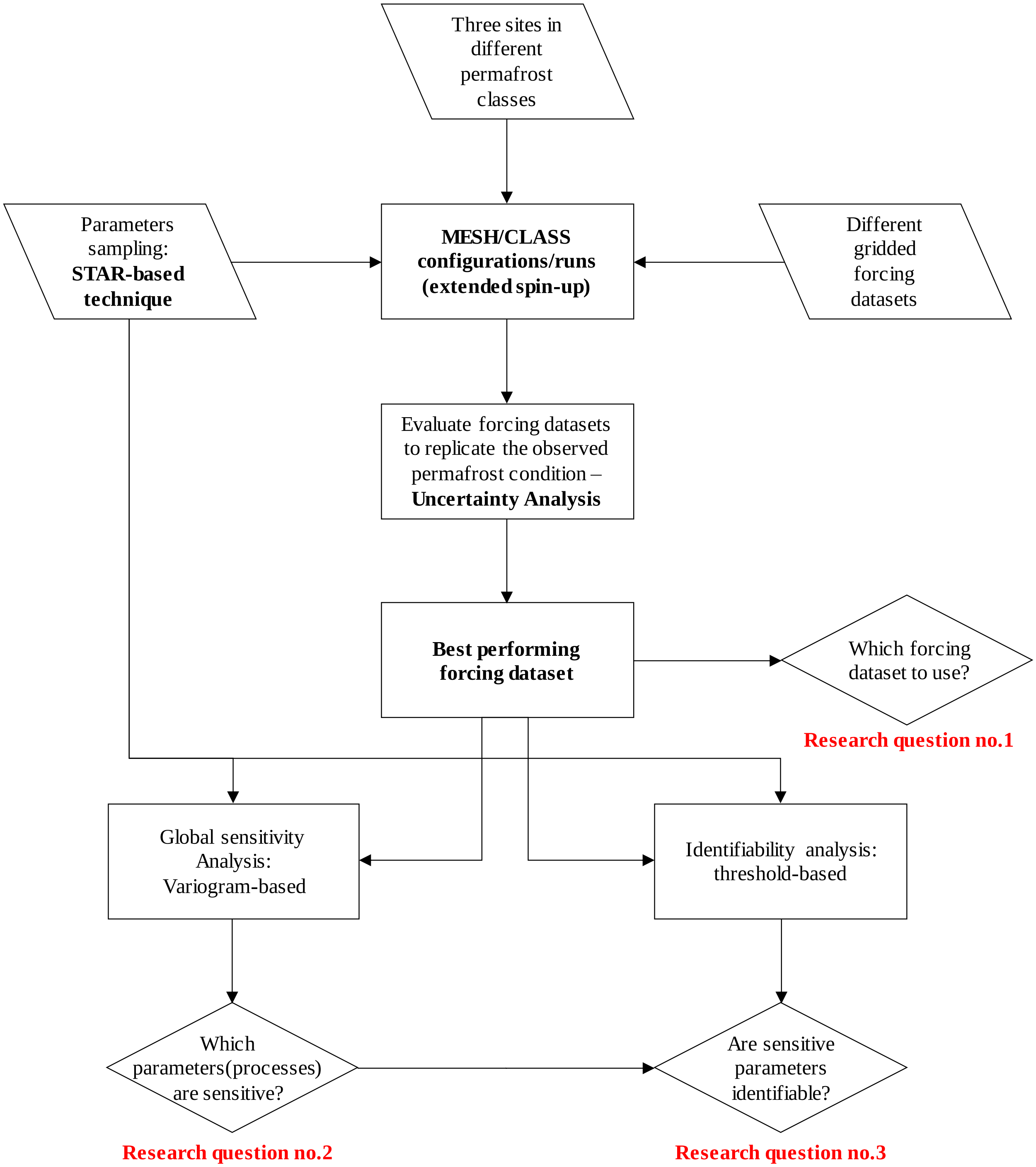



**Fig. 1.** A flowchart of the employed methodology and the expected outcomes.

# 2 Models, datasets, and methods

## 2.1 Model description

The **M**odélisation **E**nvironmenntale communautaire – **S**urface et **H**ydrology (MESH: Pietroniro et al., 2007) is selected for this study because it is physically based, suited for large-scale studies, and includes state-of-the-art representation of the dominant cold regions' processes (Pomeroy et al., 2016). MESH is a semi-distributed, grid-based modelling framework consisting of a land surface component that quantifies vertical energy/water fluxes (CLASS: Verseghy, 1991, 2000; SVS: Husain et al., 2016), algorithms for lateral movement of surface/subsurface flow (WATROF: Soulis et al., 2000; PDMROF: Mekonnen et al., 2014) and a grid-to-grid hydrologic river routing module (WATFLOOD: Kouwen et al., 1993b). The spatial heterogeneity within each cell is represented by subdividing it into tiles based on land cover, soil type, or slope and aspects. Using the Grouped Response Unit concept (GRU: Kouwen et al., 1993a), tiles with the same characteristics (*e.g.* needleleaf forest on sandy soil) in different grid cells share the same physiographic attributes, which reduces the parameterization burden and facilitates parameter transferability across space (Pietroniro & Soulis, 2003). Fluxes are typically calculated at a half-hourly time step at the tile-level and aggregated for each cell based on a weighted average of GRU fractions. MESH is driven by seven meteorological forcing variables: precipitation, air temperature, specific humidity, barometric pressure, incoming shortwave radiation, incoming longwave radiation, and wind speed. Interested readers are referred to Wheater et al. (2021) for a recent account on model developments and applications.

For the current study, CLASS version 3.6 is used as the LSM and WATROF as the runoff generation algorithm. CLASS solves the coupled water and energy balances for a user-specified soil column (the default is a three-layer with thicknesses of 0.1m, 0.25m and 3.75m) generalized across the modelled watershed. We used a deeper soil column with a power-function-based discretization of layers (see **Section 2.4.1**). In CLASS, soil parameters, which determine the thermal and hydraulic regimes, are typically tied to soil texture using pedotransfer functions. Each soil layer's temperature and moisture content evolve at each time step based on the solution of coupled water and energy balance equations. The upper boundary condition of CLASS is determined through solving the surface energy and water balance considering overlaying vegetation and snowcover, and the lower boundary condition as either a zero heat flux or a user-specified geothermal flux at the bottom of the soil column, with free drainage. No lateral migration of heat or moisture between adjacent cells is currently implemented except through surface routing. CLASS requires 17 prognostic variables for each tile covering different model initial states above the ground (*i.e.* snowpack, canopy) and underground (*i.e.* soil moisture (liquid and frozen) and temperature for each layer). Further details are provided in the CLASS manual (Verseghy, 2012).

The following permafrost characteristics were extracted from the continuously simulated soil temperature profiles to describe permafrost dynamics (**Fig. 2**);

1) Temperature envelopes (Tmax and Tmin), calculated as the maximum and minimum soil temperature profiles over the year.
2) **A**ctive **L**ayer **T**hickness (ALT), which is the maximum depth of the 0°C isotherm over the year, taken from the Tmax envelope (*i.e.* thaw only).
3) **M**ean **A**nnual **G**round **T**emperature (MAGTp) at the top of **p**ermafrost (permafrost table).
4) The **D**epth of the **Z**ero **A**nnual **A**mplitude (DZAA), where the Tmax and Tmin envelopes meet within a tolerance of 0.1°C.
5) The depth to **P**ermafrost **B**ase (PB), where the Tmax and Tmin envelopes intersect with the 0°C isotherm, noting that the model does not simulate the freezing point depression.
6) Thermal offset, which is the difference between the mean annual temperature at the ground surface and the permafrost table.
7) Surface offset, which is the difference between the mean annual temperature at the ground surface and mean annual air temperature (MAAT) – divided into a winter offset (Dec to Feb) and a summer offset (June to August).
8) Date of maximum thaw, which is calculated from the evolution of the daily temperature profile of each year and is used to indicate the inter-annual variability of thawing/freezing cycles.

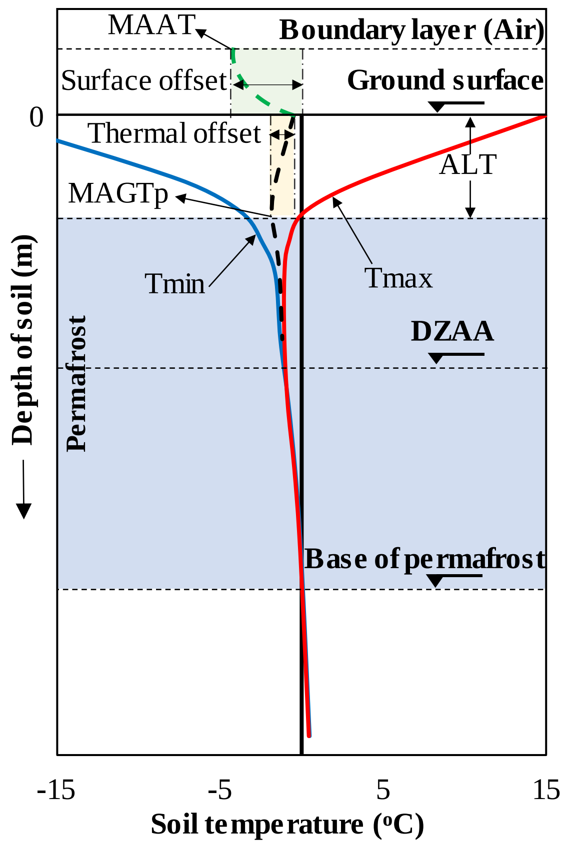



**Fig. 2.** Schematic of the soil column showing variables used to represent permafrost dynamics, modified after Abdelhamed et al. (2021).

## 2.2 Area of study

Three representative permafrost sites with distinctive hydroclimatic conditions in Canada are used in this study (**Fig. 3**): Jean Marie Creek (JMC) underlain by sporadic permafrost, Bosworth Creek (BWC) underlain by discontinuous permafrost, and Havikpak Creek (HPC) underlain by continuous permafrost. The sites are located along the main-stem of the Mackenzie River, Northwest Territories, Canada (**Fig. 3**). The Mackenzie River Basin (MRB) has a drainage area of 1.78 million km$^2$ and partially covers the Yukon, British Columbia, Alberta, Saskatchewan, and the Northwest Territories. More than 75% of the basin is underlain by permafrost based on the permafrost Map of Canada (Hegginbottom et al., 1995), with 16% continuous permafrost in the far north and northwest, 27% discontinuous permafrost covering the central east-to-west part of the basin, 26% sporadic permafrost to the south of discontinuous permafrost regions, and 10% isolated patches of alpine permafrost in the southwest of the basin in Alberta and British Columbia. The current climate of the basin is characterized as subarctic (*i.e.* cold, no dry season, cold summer) according to the Köppen-Geiger classification (Peel et al., 2007), with warmer summers projected in the south under the RCP8.5 climate change scenario (2071-2100) (Beck et al., 2018). The discontinuous and sporadic permafrost regions are characterized by warm ground temperatures (-2 to 0 °C) and the limit of permafrost is expected to shift northward under climate change (DeBeer et al., 2016; Yu Zhang et al., 2008).

The JMC site is dominated by boreal forest (needleleaf) and scattered shrubs on peat plateaux where the permafrost is relatively warm (MAGTp of -0.1°C) with a limited thickness (~ 4m) and relatively shallow active layer (~ 1.5m thick). The available data from the 85-12B borehole (to 9.7m depth) spans the period 1986 to 2000, with no records available in the 21st century. The BWC site is mainly covered by boreal forest (needleleaf and broadleaf) with a thickness of 10-50m (MAGTp of -1.5°C) and an active layer thickness of about 2m on average. The available observations (Norman Wells pump station (84-1) to 13.6m depth) cover 1985 to 2001, 2012, and 2015-2016. The HPC site is covered by taiga forest and shrubs where permafrost is cold (MAGTp of -4°C) with a considerable thickness (> 300m) and an active layer less than 1m thick. Temperatures at Inuvik Airport (site 01TC02) borehole (data to 10m depth) were used for HPC, with data available from 2008 to 2016. **Fig. 4** shows the temperature profiles used here for model evaluation. The sites have different climate conditions with an average annual daily air temperature between -2°C and -9°C and average annual precipitation between 250 and 400 mm $yr^{-1}$ among the three sites over the 1979-2016 period (**Table 1**). Thermal and geological data are available from various Geological Survey of Canada (GSC) reports (Ednie et al., 2013; Smith et al., 2004, 2009, 2010; Smith, Chartrand, Duchesne, & Ednie, 2016; Smith, Chartrand, Duchesne, Ednie, et al., 2016). Further information on the selected experimental sites is available in Elshamy et al. (2020).

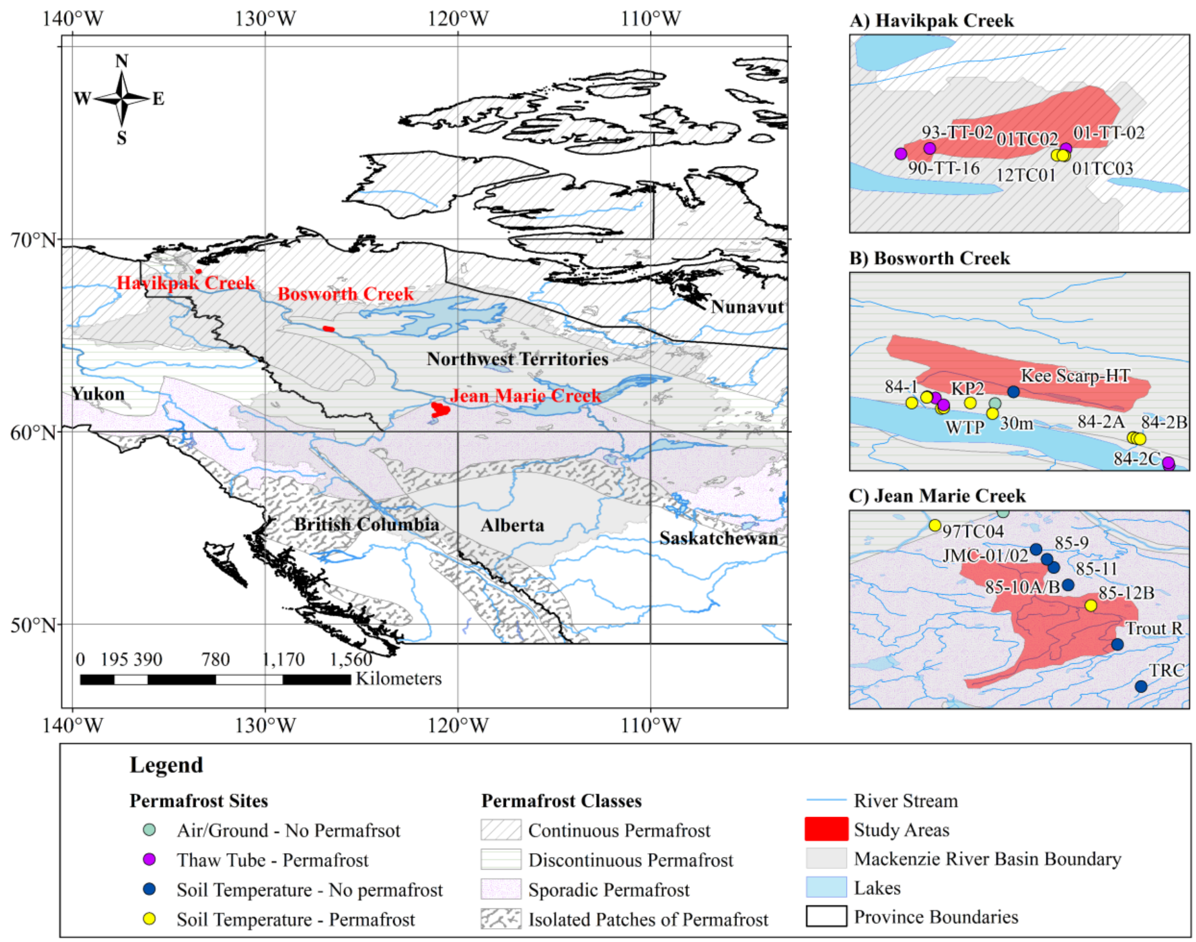



**Fig. 3.** Location of the study area, temperature boreholes, and permafrost classification.

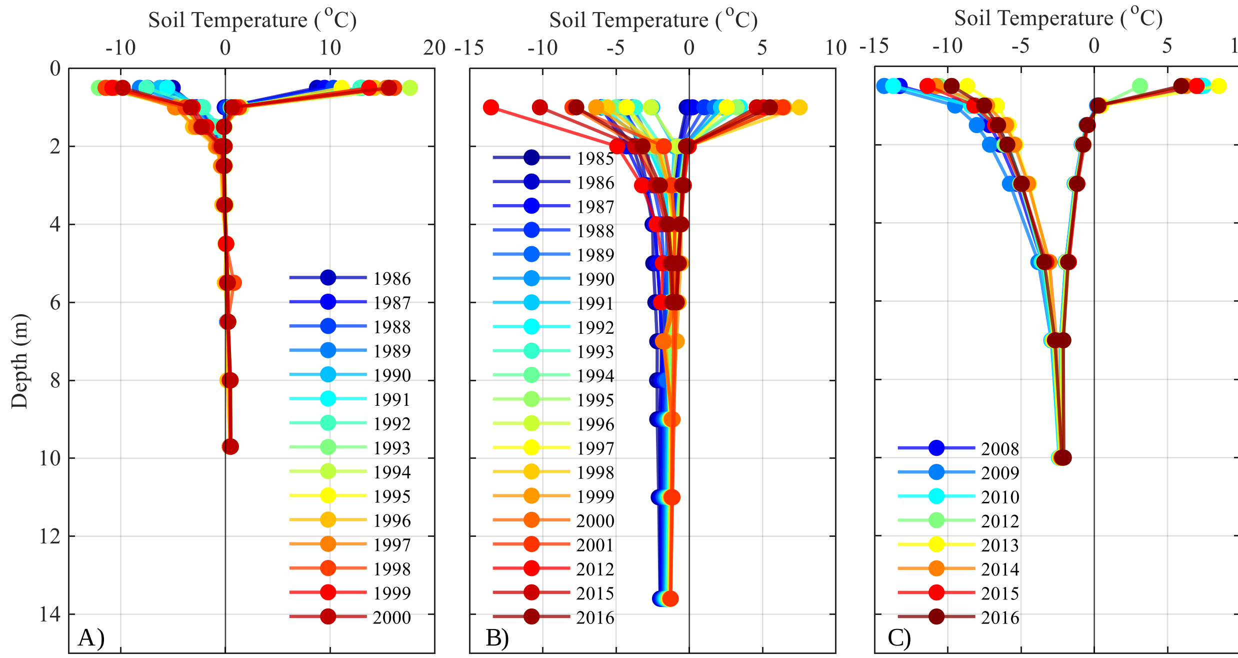



**Fig. 4.** Observed temperature envelopes at A) JMC, B) BWC, and C) HPC sites.

**Table 1.** Comparison of air temperature and precipitation for the three sites using the available meteorological stations for the period (1979-2016) – refer to **Table 3** for further information on the utilized meteorological stations.

| Site | Station ID | Mean annual air temperature (°C) | | Total annual precipitation (mm) | |
|---|---|---|---|---|---|
| | | Mean | Standard deviation | Mean | Standard deviation |
| JMC | 2202570 & 2202578 | -2.63 | 1.03 | 376.69 | 88.22 |
| BWC | 2202800 & 2202801 | -5.04 | 1.02 | 295.07 | 64.73 |
| HPC | 2202101 & 2202102 | -7.80 | 1.44 | 235.68 | 42.88 |

## 2.3 Climate forcing

As mentioned earlier, seven meteorological variables are required at a sub-daily time step to drive MESH. The ground-based stations in the neighbourhood of the study area provide hourly observations for air temperature, relative humidity, wind speed and barometric pressure, while total precipitation is provided daily, and longwave/shortwave radiations are not observed. Therefore, ground-based observations are not fully available and a model-based product is needed for the current study. Forcing dataset selection is constrained by the quality of the meteorological estimates and the overlap with the permafrost experimental datasets. As mentioned earlier, the available records are 1985-2016, 1986-2000, and 2008-2016 for BWC, JMC, and HPC, respectively; thus, the selected forcing dataset should begin before 1985 to enable model initialization and performance assessment. A few widely-used forcing datasets start prior to the 1980s, such as WFD (WATer and global CHange (WATCH) Forcing data), available from 1901 (Weedon et al., 2011), the Princeton dataset, available from 1901 (Sheffield et al., 2006), WFDEI (WFD with the ERA-Interim analysis), available from 1979 (Weedon et al., 2014), WFDEI-GEM-CaPA (WFD with the ERA-Interim analysis bias-corrected by GEM-CaPA), available from 1979 (Asong et al., 2020), and WFDE5 (bias-corrected WFD with the ERA5 reanalysis), available from 1979 (Cucchi et al., 2020). However, the WFD and Princeton datasets were discontinued in 2001 and 2012, respectively. The combined product of the Global Environmental Model (GEM; Côté et al., 1998), atmospheric forecasts and the Canadian Precipitation Analysis (CaPA; Mahfouf et al., 2007) showed considerable agreement with ground observations for the precipitation (Wong et al., 2017), but GEM-CaPA is not available prior to 2002 and the most recent version, the Regional Deterministic Reanalysis System (RDRS v2), is currently only available from 2000 to 2017 (Gasset et al., 2021).

Three forcing datasets are used in this study: WFDEI, WFDEI-GEM-CaPA (denoted WFDEI-GC hereafter), and WFDE5. Given that reasonable estimates of precipitation fields were obtained from WFDEI, as shown by Wong et al. (2017) for Canada, and the fact that it is available from 1979 with adequate temporal resolution (3 hours), WFDEI has an advantage. However, it has been found to be slightly biased relative to observation over Northern Canada (above 60°N) (Asong et al., 2020; Wong et al., 2017). It was therefore bias-corrected with the relatively short but more accurate GEM-CaPA product. However, the 40m estimates of GEM-CaPA (for temperature, humidity and wind speed) were used to bias-correct the surface-based estimates of WFDEI, yielding the WFDEI-GC dataset at a non-surface reference height (40m) that limits its applicability for some hydrologic models (Asong et al., 2020) but not MESH. Additionally, the WATCH Forcing Data methodology was applied to the ERA5 reanalysis data derived using the sequential elevation and monthly basis correction method in Weedon et al. (2011) for ERA5 reanalysis. WFDE5 was multi-variate bias-corrected for a limited number of meteorological variables (precipitation, temperature, and shortwave radiation) using the simple approach of rescaling the monthly average CRU ‘Climate Research Unit’ (and/or GPCC ‘Global Precipitation Climatology Center’) estimates (Weedon et al., 2014), unlike Asong et al. (2020) who used multi-variate quantile-mapping to correct the bias for the seven meteorological variables (conserving cross-correlations amongst them). Monte-Carlo (MC) simulations incorporating the three forcing datasets (*i.e.* WFDEI, WFDE5, and WFDEI-GC) and under the full range of parameter uncertainty was used to aid in selecting the best performing meteorological dataset for MESH/CLASS simulations for the period 1979-2016 (refer to **Fig. 1** and **Section** 2.4.3).

## 2.4 Experimental design and implementation

Experiments were designed to assess input uncertainty, model sensitivity and parameter identifiability of MESH in reproducing the observed permafrost conditions at the three sites with distinct climatic and geological conditions (**Fig. 1**). We conducted these comprehensive analyses using the MESH model (Pietroniro et al., 2007; Wheater et al., 2021) for the full set of LSM (*i.e.* CLASS) parameters, various model configurations, multiple permafrost variables/performance criteria, for three sites in Canada. We use two lenses for parameter analysis: GSA using the variogram Analysis of Response Surfaces (VARS: Razavi and Gupta, 2016a) and traditional threshold-based identifiability analysis. The point-scale experiments are adopted from Elshamy et al. (2020) and Abdelhamed et al. (2021), where MESH was utilized for investigating permafrost initialization in LSMs. We employed the soil layering scheme proposed by Elshamy et al. (2020) (**Table A1**), which extends to a depth of 51.24 m and has a fine discretization for the upper 2 meters of the soil (9 layers), in line with the observed ALT for the selected sites (**Fig. 4**). The lower boundary condition of the soil column (*i.e.* Neumann-type flux boundary condition), known as the geothermal flux, was set to zero since several studies underscored its limited/negligible impact on the simulated temperatures on a centennial timescale (Hermoso de Mendoza et al., 2020; Lawrence et al., 2008; Nicolsky et al., 2007; Sapriza-Azuri et al., 2018). Initializing MESH state variables (*i.e.* soil temperature and liquid/frozen contents) was achieved by spinning the first year of the climate record (*i.e.* Oct 1$^{st}$, 1979 - Sep 30$^{th}$, 1980) for 1000 cycles for each sampled set of parameters, as recommended by Abdelhamed et al. (2021).

### 2.4.1 Selection of parameters, variables, and metrics

Regarding MESH parameters, six groups of parameters representing canopy, soil texture, soil permeable depth, drainage, ponding, and snow cover processes were perturbed within their physical ranges to assess their influence on the permafrost dynamics (**Table 2**). The range of the canopy parameters was based on the lookup tables from the CLASS manual (Verseghy, 2012);

note that the three sites were parameterized as needleleaf forest, and hence, their canopy parameters have the same ranges of variation. Ponding, drainage, and snow cover parameters are identical across the three setups, with ranges taken from previous studies with the same model (*e.g.* Davison et al., 2016; Haghnegahdar et al., 2017) and textbook values (Dingman, 2015). Regarding SDEP (depth to the bedrock), we used the gridded bedrock depth dataset by Shangguan et al. (2017) to identify the upper limit, while the maximum root depth is used to define the lower limit. Although runoff generation processes (*i.e.* interflow, surface runoff and drainage from the soil column) are treated as vertical processes, they allow water to exit the system via the lateral horizons, which influences water stores and hence the hydraulic and thermal regime of the system. The last parameter group defines soil texture as sand, clay, and organic matter percentages. Since each soil layer has three descriptive parameters, and each model configuration has 25 soil layers, we grouped layers as appropriate and assigned the same values to each group's parameters. A new parameter, ODEP defining the depth of organic soil layers is introduced to reduce the number of parameters considered in the analysis, reduce computational cost, and facilitate a more straightforward analysis. ODEP is sampled over the given range and nudged to layer boundaries. It is used to divide the soil column into two horizons: horizon *i* with high organic content 'ORGM*i*' for all layers above ODEP, and horizon *j* with mineral soil texture and no organic content below it. This also prevents unrealistic combinations that could lead to model crashes and strange behaviour. The range of soil texture parameters (Sand % and Clay %) is determined from the Soil Landscapes of Canada (SLC) v2.2 (Keshav et al., 2019) after the U.S. Department of Agriculture (U.S. Department of Agriculture, 1951), while the range of soil organic content 'ORGM*i*' is identified from the available geological boreholes at each site.

**Table 2.** Parameters and their corresponding ranges for the model GSA for each site. Id (1:13) refers to canopy parameters, Id (14) refers to permeable soil depth, Id (15:20) refers to drainage/runoff parameters, Id (21:26) refers to soil texture/hydraulic parameters, Id (27) refers to snow-cover parameter, and Id (28:29) refers to ponding parameters.

| Id | Group | Parameter | Unit | Description | JMC/BWC/HPC Lower limit | JMC/BWC/HPC Upper limit |
|---|---|---|---|---|---|---|
| 1 | Canopy | LAMX | - | Annual maximum leaf area index | 1.5 | 2.5 |
| 2 | | LAMN | - | Annual minimum leaf area index | 0.5 | 1.5 |
| 3 | | LNZ0 | - | Natural logarithm of the roughness length | 2 | 3 |
| 4 | | ALVC | - | Average visible albedo when fully leafed | 0.02 | 0.04 |
| 5 | | ALIC | - | Average near-infrared albedo when fully leafed | 0.15 | 0.23 |
| 6 | | CMAS | $kg/m^2$ | Annual maximum canopy mass | 10 | 30 |
| 7 | | ROOT | m | Annual maximum rooting | 0.5 | 2 |
| 8 | | RSMN | s/m | Minimum stomatal resistance | 150 | 250 |
| 9 | | QA50 | $W/m^2$ | Reference value of incoming shortwave radiation for stomatal resistance formula | 20 | 40 |
| 10 | | VPDA | - | Vapor pressure deficit coefficient for stomatal resistance formula | 0.4 | 0.9 |
| 11 | | VPDB | - | Vapor pressure deficit coefficient for stomatal resistance formula | 0.8 | 1.3 |
| 12 | | PSGA | - | Soil moisture suction coefficient for stomatal resistance formula | 75 | 125 |
| 13 | | PSGB | - | Soil moisture suction coefficient for stomatal resistance formula | 2 | 8 |
| 14 | Permeable depth | SDEP | m | Soil permeable depth | 2 | 7.06/15.21/20.24 |
| 15 | Drainage/runoff | GRKF | - | The fraction of (horizontal) saturated soil conductivity moving in the horizontal direction | 0.001 | 1 |
| 16 | | KSAT | m/s | (Horizontal) saturated surface soil hydraulic conductivity | 0.0001 | 0.5 |
| 17 | | DRN | - | Drainage index controls water seepage from the bottom of the soil column | 0 | 1 |
| 18 | | DD | $km/km^2$ | Drainage density | 2 | 100 |
| 19 | | XSLP | - | Estimated average slope of the tile/GRU | 0.0001 | 0.4 |
| 20 | | MANN | - | Manning's roughness coefficient for overland flow generation "n" | 0.01 | 0.15 |
| 21 | Soil texture | SAND*i* | % | Percent sand in the soil of layers *i* | 0/13/0 | 12/30/32 |
| 22 | | CLAY*i* | % | Percent clay in the soil of layers *i* | 8/18/28 | 12/27/42 |
| 23 | | ORGM*i* | % | Percent organic matter in the soil of layers *i* | 0 | 60/30/15 |
| 24 | | SAND*j* | % | Percent sand in the soil of layers *j* | 0/13/0 | 12/30/32 |
| 25 | | CLAY*j* | % | Percent clay in the soil of layers *j* | 8/18/28 | 12/27/42 |
| 26 | | ODEP | m | Depth of the organic soil | 0.1 | 7/1/1 |
| 27 | Snow cover | ZSNL | m | Min depth to consider 100% cover of snow on the ground surface | 0.05 | 0.5 |
| 28 | Ponding | ZPLS | m | Max depth of water allowed to be stored on ground surface for snow-covered area | 0.05 | 0.5 |
| 29 | | ZPLG | m | Max depth of water allowed to be stored on ground surface for snow-free area | 0.05 | 0.5 |

ALT, Tmin and Tmax (refer to **Section 2.1** for definitions) were employed to assess the impact of parameter uncertainty on simulated permafrost dynamics under different forcing datasets, model sensitivities, and parameters identifiability. These three variables shed light upon the overall thermal regime using the annual envelopes (Tmax and Tmin), as well as a specific focus on ALT as the most critical and direct aspect used to describe permafrost dynamics. The BIAS and the Mean Absolute Error (MAE) of Tmin, Tmax and ALT were used to assess the quality of permafrost simulation at the annual time-scale. The error metrics were averaged over time (*i.e.* the record length) and vertical space (*i.e.* column depth for Tmax and Tmin). Although both metrics quantify the direct bias in the model residual, MAE avoids the bias compensation that can happen when one year has a positive bias and another a negative bias. Further, in a detailed case study on model performance assessment of behavioural parameter sets (**Section 3.5**), we considered additional permafrost characteristics (*i.e.* DZAA, PB, thermal offset, surface offset, date of maximum thaw) to ALT, Tmax, Tmin (refer to **Section2.1** for definitions).

#### 2.4.2 Global sensitivity analysis

Sensitivity analysis of model parameters (SA) can be beneficial in identifying the main factors (*e.g.* boundary/initial conditions, driving forcing and model parameters) controlling permafrost dynamics and overall model performance. According to Saltelli and Annoni (2010) and Razavi and Gupta (2015), SA traditionally serves three primary purposes: 1) ranking parameters' contribution to output variance, 2) filtering parameters with a negligible influence on output variance, and 3) mapping parameters' space to locate the regions with a satisfactory performance – exploring causalities between different processes/hypotheses and supporting decision-making are among the other benefits of SA (Razavi et al., 2021). Local sensitivity analysis (LSA) and global sensitivity analysis (GSA) are the main categories of SA. LSA explores output variability around a single reference point, which, regardless of its high simplicity and intuitiveness, is not appropriate for complex environmental models due to their non-linearity and significant parameter interactions (Saltelli & Annoni, 2010). In contrast, GSA evaluates model output variability over the entire feasible factor space, where a large sample of input factors is generated and output variation is analyzed (Saltelli & Annoni, 2010).

Controlled model experiments (often LSA with discrete factor space) have been integral for diagnosing model structure and validating potential modifications to enhance permafrost simulation in several LSM studies. For instance, Alexeev et al. (2007), Nicolsky et al. (2007), and Lawrence et al. (2008) improved simulated permafrost dynamics in CLM3 by assessing the sensitivity to soil layer geometries and textures. Sapriza-Azuri et al. (2018) also examined the sensitivity of soil column depth to parameters using CLASS LSM. Chadburn et al. (2015), Paquin and Sushama (2015) and Melton et al. (2019) examined the sensitivity of permafrost to various snow parameterizations, which improved the simulated permafrost extent for JULES and CLASS LSMs. Sapriza-Azuri et al. (2018) and Hermoso de Mendoza et al. (2020) also investigated the sensitivity of the evolution of soil temperatures to the geothermal heat flux. Chadburn et al. (2015), Melton et al. (2019) and Elshamy et al. (2020) explored the impact of the depth to bedrock on permafrost thermal/hydraulic regimes. Further, sensitivity to major input variables and parameters was necessary for developing and diagnosing the NEST model, which integrates the strength of permafrost models with LSMs (Yu Zhang et al., 2003). Lastly, the analysis of permafrost sensitivity to external forcing was central to quantifying the impact of input uncertainty and proposing methodological improvements for LSMs (Burke et al., 2020; McGuire et al., 2018; Paquin & Sushama, 2015; Slater & Lawrence, 2013). It is noteworthy that all the abovementioned studies employed SA informally, which has contributed to a variety of LSM diagnosis/development, but the used ‘what-if-scenario-based’ LSA has often been criticized for not being thorough enough to yield sound decisions/sensitivities (Saltelli & Annoni, 2010).

This study utilized the variogram analysis of response surfaces framework (VARS: Razavi and Gupta, 2016a). This provides a comprehensive spectrum of sensitivity information as it bridges the variance- and derivative-based approaches. For example, it produces sensitivity indices of the two most common GSA approaches, the derivative-based (Morris, 1991) and the variance-based methods (Sobol, 2001), while being more computationally efficient and statistically robust (Becker, 2020; Puy et al., 2021). To summarize global sensitivities, VARS integrates the directional variograms over a given perturbation scale (*e.g.* 10 %, 30%, and 50%) and produces a set of sensitivity indices called IVARS (Integrated Variograms Across a Range of Scales) (Razavi & Gupta, 2016a).

The STAR-VARS implementation (Razavi & Gupta, 2016b) is used here. This sampling strategy first generates star centers randomly using, *e.g.* Latin hypercube sampling, and then using a structured sampling approach generates the points on the star wings. Sampling is implemented with a resolution of $\Delta h = 0.1$ and with 100 star centers, as recommended by Razavi and Gupta (2016b), where star centers were selected using Progressive Latin Hypercube Sampling (Sheikholeslami and Razavi, 2017). This resulted in a total of 26,200 model evaluations (for 29 parameters) for each permafrost site (see **Section 2.4.1**). The normalized values of $IVARS_{50}$ (adds up to 100% to a “Ratio of Sensitivity”) are used to outline GSA results. This allows straightforward interpretation of sensitivity indices in terms of parameter importance and facilitates a consistent comparison across different metrics and cases. Model crashes due to infeasible combination of sampled parameters were handled by applying a data-filling strategy for the response surface following Sheikholeslami *et al.* (2019).

#### 2.4.3 Uncertainty and identifiability analyses

The study employed the Monte-Carlo (MC) procedure (**Fig. 1**) to assess the impact of parameter uncertainty on simulated permafrost dynamics, represented by ALT, Tmin and Tmax. The MC approach has been integral to various model analysis methodologies, such as the Generalized Likelihood Uncertainty Estimation Framework (GLUE: Beven and Binley, 1992), Regional Sensitivity Analysis (RSA: Spear and Hornberger, 1980), and Dynamic identifiability analysis (DYNIA: Wagener et al., 2003). MC analysis was also used for multi-objective calibration of an LSM (*e.g.,* (Houser et al., 2001)). Commonly, the MC procedure is based on sampling from uniformly distributed input spaces (*i.e.* model parameters). However, the high dimensionality of LSMs and the non-linearity of their response require a large number of samples, with high computational cost, especially as we also explore different meteorological forcing sets. To reduce the computational burden, a sampling strategy was used that conveys the maximum information from the model-output space with a minimal sample size (Sheikholeslami et al., 2021), based on the semi-structured parameter sampling scheme of STAR. Thus the STAR-based samples used for sensitivity analysis are also used to propagate the uncertainty of parameters to the simulated permafrost characteristics and to study their identifiability.

Parameter identifiability analysis investigates whether it is theoretically possible to have a unique parameter set for a given model structure, forcing data, observations, and response surface (objective function). Such analysis is vital to pinpoint sources of uncertainty and reduce them, leading to more credible model simulations (Guillaume et al., 2019). However, traditional applications of parameter identifiability analysis (via dotty plots or boxplots) involve a degree of subjectivity inherent in defining a behavioural threshold (*e.g.* acceptable performance accuracy as represented by a goodness of fit metric) and commonly do not account for parameter interactions. However, the concurrent application of both sensitivity analysis (forward problem) and identifiability analysis (inverse problem) in this case provides more insights into the parameter estimation over the whole response surface and at the global optimum (Gupta & Razavi, 2018). In other words, parameter estimation is examined by two different lenses to allocate and rule out various sources of uncertainty

The sampled parameter sets for each permafrost site were used for uncertainty and identifiability analyses to ensure consistency of explored input/output spaces. The uncertainty analysis was implemented to assess the combined impact of input forcing (*i.e.* meteorological data) and model parameters. The primary purpose of such analysis is to select the most appropriate forcing dataset that can encapsulate the observations, which can be achieved by examining the cumulative distribution function of the performance metrics. Parameter identifiability analysis was implemented for parameter sets that collectively fulfilled the three behavioural constraints (refer to **Section** 3.4 and **Table 6** for a discussion on the selected behavioural thresholds used for identifiability analysis). In other words, the proposed method aims to assess model simulations that satisfy all criteria (ALT, Tmin and Tmax) at once, *i.e.* using a multi-objective identifiability analysis.

# 3 Results and discussion

## 3.1 Gridded data sets assessment

A basic comparison of the three forcing data sets versus ground truth observation for the mean annual air temperature and total annual precipitation is provided in **Fig. 5**. The meteorological stations used for the comparison are listed in **Table 3**. Records at 2 stations at each site had to be merged to obtain a record for the analysis period and the overlap was assessed to ensure consistency. To facilitate the comparison of WFDEI-GC to other datasets, including ground observations, we applied an adiabatic lapse rate correction since the air temperature is provided at 40m and the ground observation is measured at ~2m height (the blue shading in **Fig. 5**). WFDE5 and WFDEI yielded similar mean annual air temperatures at HPC and JMC sites, noting that the two data sets outperformed WFDEI-GC at JMC and partially at HPC (1979-2005). The interannual variability was perfectly replicated at the JMC site (with a consistent overestimation of 1.25°C), but not at BWC and HPC, while the HPC site has a significant bias prior to 2004 (between 1-2 °C). Further, WFDEI-GC provides the warmest air temperature with an average bias of 2°C, 0.2°C, and 1.25°C for the HPC, BWC, and JMC sites, respectively. However, WFDE5 and WFDEI-GC deliver better estimates for the air temperature at BWC, while WFDEI persistently underestimates temperature values by 1-2°C. It is noteworthy that no single data product outperforms the others for the three sites collectively for the annual air temperature and similarly for monthly/daily temporal levels (figures not shown).

In general, the observed interannual variability for air temperature is better captured than precipitation among the three datasets for the whole period of comparison (1979-2016). The comparison of total annual precipitation sheds more light upon the issues/problems associated with these grid-based products. WFDEI-GC systematically overestimates the precipitation at the three sites, as does WFDE5 with a lesser magnitude but higher inter-annual variability. On the other hand, WFDEI displays a constant total annual precipitation value at HPC and BWC sites for several consecutive years, between 2008-2015 and 2006-2016, respectively. This could be attributed to the fact that CRU (which was used to constrain WFDEI) reverts to the monthly climatology when there is no data (Weedon et al., 2014). Even with the good match between WFDEI and ground observations, on average, the lack of interannual variability (repeated years) is critical in assessing permafrost initialization and dynamics, and thus it might not be advisable to use WFDEI in the current analysis. Lastly, both WFDE5 and WFDEI-GC offer similar results for the precipitation with no clear outperformance at the three sites and throughout the window of comparison.

**Table 3.** List of meteorological stations used for evaluating gridded data sets.

| Site | Station ID | Latitude | Longitude | Data Availability |
|---|---|---|---|---|
| HPC | 2202570 | 68.30 | -133.48 | 1957-2013 |
| | 2202578 | 68.32 | -133.52 | 2003-2021 |
| BWC | 2202800 | 65.28 | -126.80 | 1953-2012 |
| | 2202801 | 65.28 | -126.80 | 2003-2021 |
| JMC | 2202101 | 61.76 | -121.24 | 1963-2014 |
| | 2202102 | 61.76 | -121.24 | 2003-2021 |

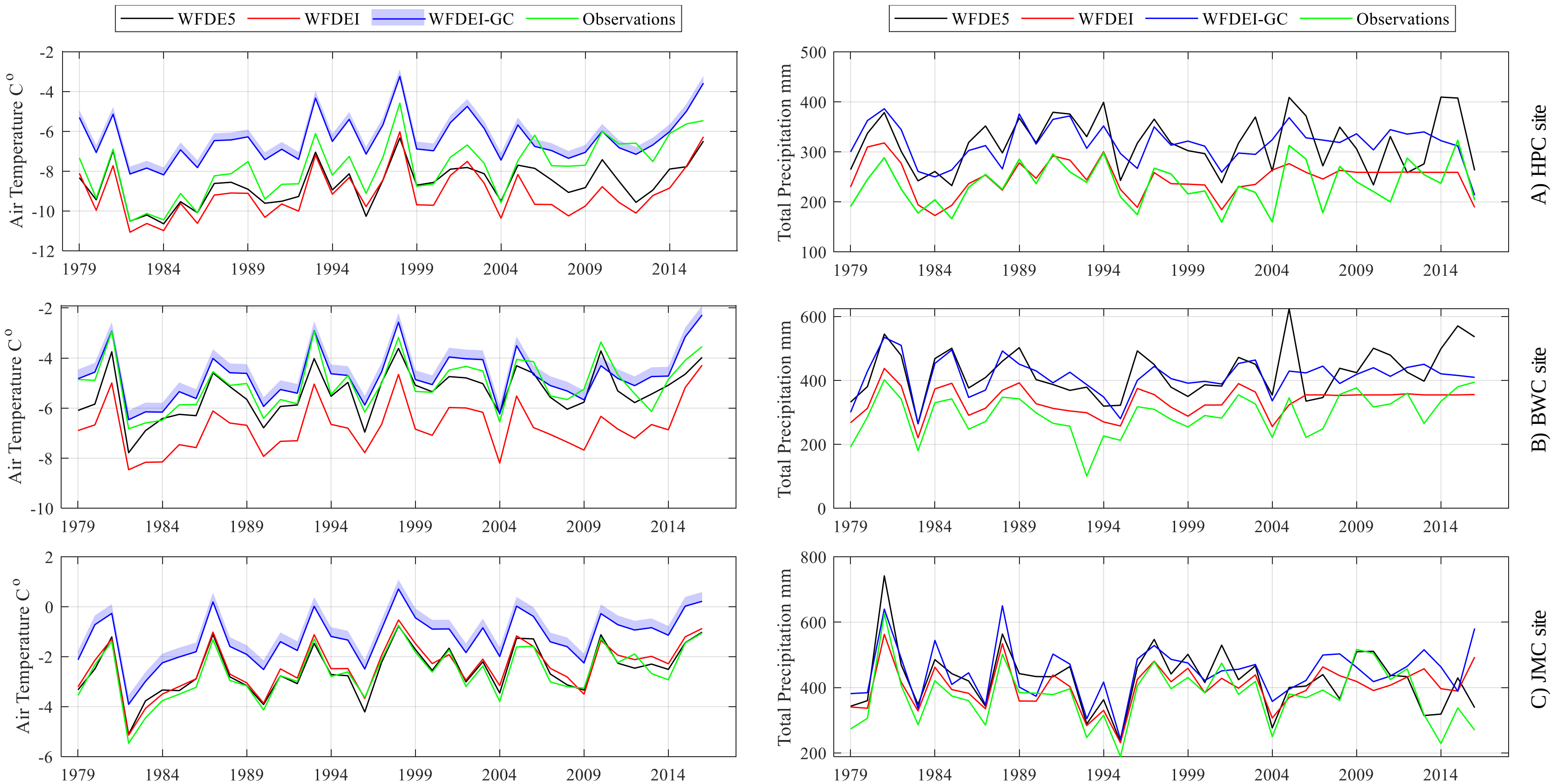



**Fig. 5.** Comparison of WFDE5, WFDEI, and WFDEI-GC to ground truth observations for mean annual air temperature (left column) and total annual precipitation (right panel) at A) HPC site, B) BWC site, and C) JMC site. The range of adiabatic lapse rate correction for WFDEI-GC air temperatures is displayed via the light blue shading.

### 3.2 Uncertainty analysis

The overarching goal of this section is to select the dataset that behaves well for most metrics and sites, and to investigate the associated uncertainty range. The impact of parameter uncertainty under different external forcing datasets is assessed by aggregating each modelled permafrost variable into a single error metric. This provides a general perspective on the effect of using (imperfect) forcing datasets on the quality/accuracy of model predictions. A summary of statistical measures of the cumulative frequency distributions (CDFs) for all experiments across the 26,200 model evaluations (**Table 4**) shows the best-performing dataset for each site and error-criterion in terms of distribution mean and range. The candidate forcing dataset should fulfill minimal mean (CDF at a frequency of 0.5) and minimal envelope (range) of variability, noting that having the mean of the CDF around zero is an additional criterion for BIAS-based assessments. Entries in bold font in **Table 4** correspond to the best climate datasets to replicate permafrost ground observations. It is clear that WFDEI-GC can reproduce observations using the BIAS and MAE error criterion at the three sites, except for Tmax BIAS at the three sites and ALT at the BWC site. Since the Tmax envelopes similarly encapsulate the observations (among the three datasets), with a slight advantage for WFDEI over WFDEI-GC, and ALT is simply a point on the Tmax envelope, the forcing dataset that has superior estimates for Tmin is selected for the rest of the study. Besides, forcing the three sites with the same climate facilitates a meaningful/comprehensive analysis and interpretation of results; using different input data products could yield misleading sensitivity and identifiability results. Thereby, we opted to use WFDEI-GC for the detailed analysis of parameter sensitivity and identification.

**Table 4.** Summary of PDFs statistical measures (mean [range]) for two performance metrics, three permafrost sites, three permafrost variables, and three climate datasets. Entries in bold indicate datasets that yield the best model performance.

| Site | Criterion | WFDE5 | | WFDEI | | WFDEI-GC | |
|---|---|---|---|---|---|---|---|
| | | BIAS | MAE | BIAS | MAE | BIAS | MAE |
| HPC | Tmin | -7.2 [-8.8:-1.9] | 7.5 [1.5:9.3] | -8.2 [-11:-5] | 8.8 [4.5:11] | **0.8 [-6:-5.5]** | **2.2 [1:5.8]** |
| | Tmax | 2.6 [0.9 :3.5] | 3.9 [1.3:4.5] | **1.4 [-4:2.2]** | 3.6 [1.6:4.4] | 1.8 [-2:4] | **1.7 [0.8:4.2]** |
| | ALT | -1.2 [-1.6:-0.5] | 1.3 [0.5:1.8] | -0.9 [-1.2:-0.2] | 0.7 [0.2:1.3] | **-0.5 [-3:0.5]** | **0.5 [0.1:3]** |
| BWC | Tmin | -2.9 [-4.5:0.75] | 3.4 [1:5.2] | -4.8 [-6.2:2] | 5.5 [2.5:7] | **0.8 [-3.5:2.5]** | **2.3 [0.8:3.6]** |
| | Tmax | 1.2 [0:2.25] | 1.7 [0.7:2.3] | **0 [-1.1:0.9]** | 1.8 [0.8:3.2] | 1.6 [-0.7:3.9] | **1.5 [0.3:4]** |
| | ALT | -1.5 [-4.2:-0.5] | 1.5 [0.4:4] | **-1 [-1.4:-0.3]** | **0.9 [0.3:1.5]** | -1.9 [-13:2] | 2.1 [0:11] |
| JMC | Tmin | -2.3 [-3.5:0] | 1.4 [0.8:3.3] | -2.2 [-4.2:1.7] | 2.3 [0.8:4.2] | **0.1 [-3.2:2.4]** | **0.9 [0.5:2.7]** |
| | Tmax | 2 [0.5:4] | 2 [1.2:3.9] | **1.4 [0.4:3.7]** | **1.8 [1:4]** | 2.6 [-0.5:5.7] | 2.7 [0.8:5.2] |
| | ALT | -1.1 [-8:0] | 1 [0.2:8] | -0.9 [-6.1:0.5] | 0.8 [0:6] | **0.2 [-2:1]** | **0.2 [0:2.4]** |

To gain insight into the associated uncertainty range under different datasets, the CDFs for the averaged BIAS of Tmin are presented in **Fig. 6**, and the CDFs for the averaged MAE of ALT are shown in **Fig. 7** (refer to **Fig. A1** for the CDFs of the averaged BIAS of Tmax). Several points can be observed:

- WFDEI-GC outperforms the other datasets in simulating Tmin at the three sites. A considerable number of parameter sets can replicate the observed Tmin without any bias. Still, the range of uncertainty is relatively large at the HPC site (range ~ ±5°C) compared to the other sites (range ~ ±2°C) forced by the same climate dataset. This can be attributed to an unsuccessful bias removal at HPC site and/or incorporating a wider/unfeasible range for model parameters,
- Both WFDE5 and WFDEI produce cooler Tmin (*i.e.* CDF mean < 0°C) at the three sites, noting that the site furthest north (HPC) experiences the coldest bias (~ -7°C on average) compared to the -4°C and -2°C biases at the other southern sites,
- WFDE5 provides better estimates of Tmin at the three sites compared to the original WFDEI; still, WFDE5 is inferior to WFDEI-GC,
- The range of uncertainty of Tmin for BWC and JMC forced by different climate datasets is almost identical, of the order of ~4°C. On the other hand, HPC does not depict the same behaviour, as both WFDEI and WFDE5 have a range of uncertainty of ~6°C compared to 10°C for WFDEI-GC,

- For ALT, WFDEI-GC offers relatively better estimates (MAE) at the HPC and JMC sites. Yet, WFDEI improves the identification of ALT for BWC site (MAE of 0.9m on average that varies around 0.3m and 1.5m) with respect to WFDEI-GC that does not yield better estimates of ALT (MAE 2.1m on average with a large range of variability between 0 and 11m),
- Unlike Tmin, WFDE5 could not improve the ALT simulation compared to the WFDEI; both give a biased estimate with a slight advantage for WFDEI (~ 0.3m improvement in the MAE of ALT),
- The distribution shape for ALT is not consistent among the three sites, even under the same forcing dataset, highlighting the model's non-linearity. On the other hand, unimodal and bimodal normal distribution characterize the Tmin case,
- No single dataset can collectively provide satisfactory model simulations (*i.e.* CDF encapsulate the observations with a mean of zero) for ALT and Tmin at the three sites.

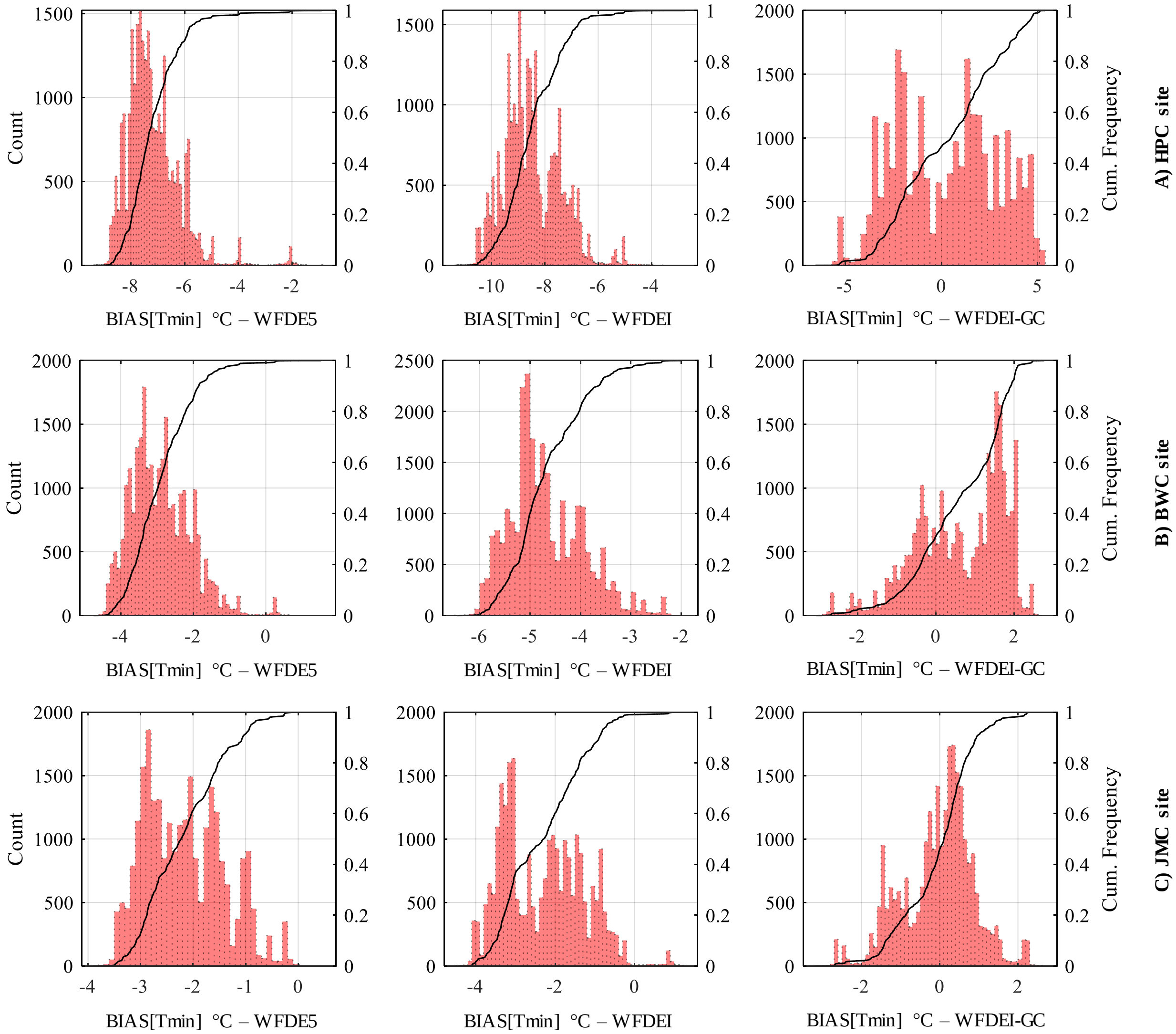



**Fig. 6.** Histograms (red bars) and cumulative frequency distributions (black lines) for the BIAS in simulated minimum annual temperature envelope (Tmin) at A) HPC site, B) BWC site, and C) JMC site under WFDE5 (left column), WFDEI (middle column), and WFDEI-GC (right column; the Count (left axis of each subplot) refers to the number of model evolutions.

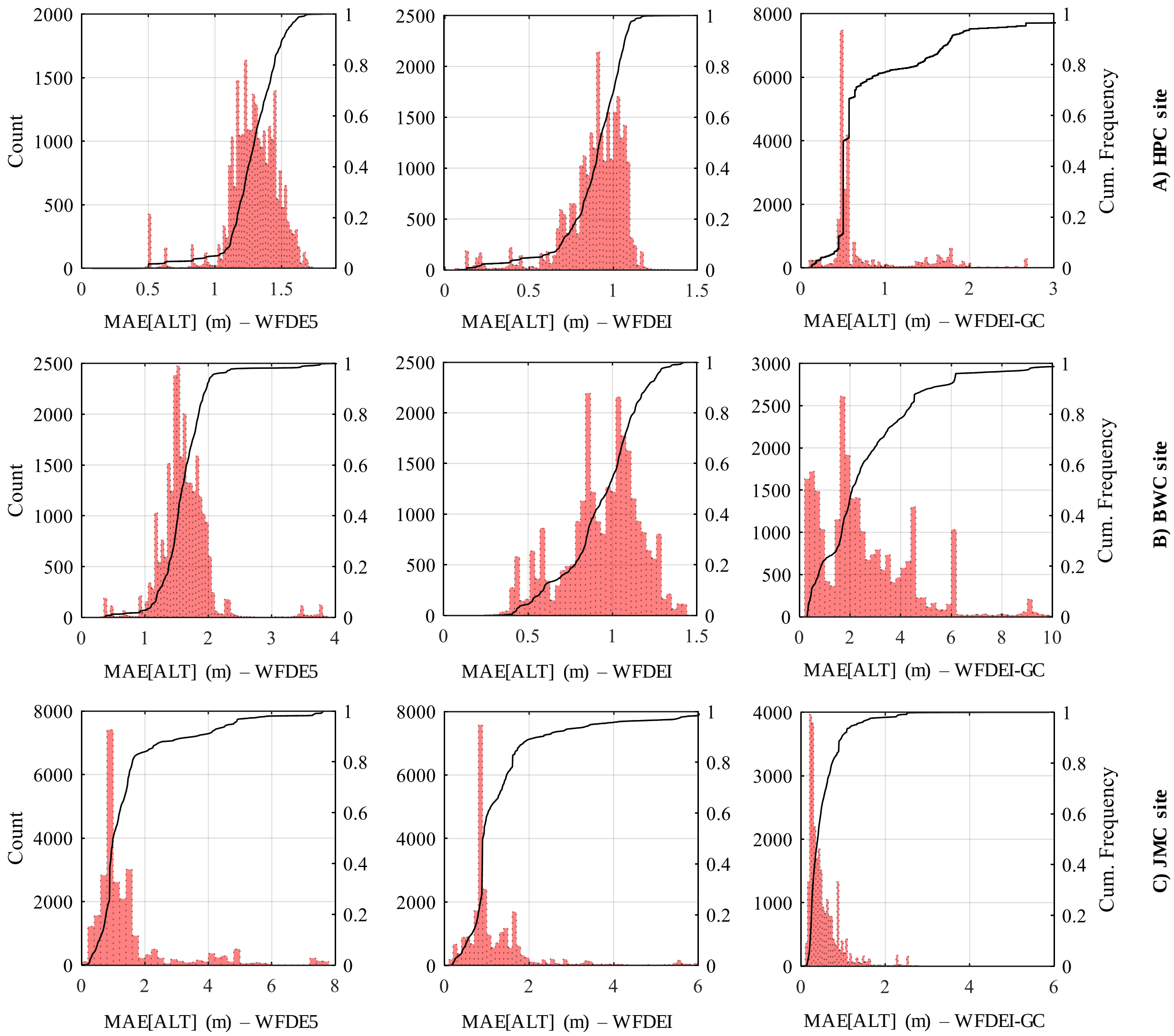



**Fig. 7.** Same as **Fig. 6**, but showing the MAE in simulated active layer thickness (ALT).

## 3.3 Sensitivity analysis

This section discusses GSA results based on the ratio of sensitivity (normalized $IVARS_{50}$). GSA response surfaces were constructed for both the BIAS and MAE of Tmax, Tmin, and ALT. Unfeasible parameter combinations that violate the numerical stability conditions triggered crashes of 6%, 1%, and 10% of the model runs for HPC, BWC and JMC, respectively. As noted above, a model emulation-based substitution technique handled these crashes (Sheikholeslami *et al.*, 2019). Bootstrapping was enabled for all GSA experiments, facilitating the assessment of the confidence in GSA results, ensuring the stability of the GSA algorithm, and accounting for randomness in sampling variability (refer to **Fig. A2** for the reliability of sensitivity indices).

**Fig. 8** shows that both the BIAS and MAE experiments yield consistent GSA results. For each permafrost variable, the experiments have similar sets of most sensitive parameters in terms of their relative ranking, with a minor impact on each parameter's absolute contribution. JMC is an exception in this regard, as the order of the two most influential parameters to ALT switches; the BIAS is dominated by ZSNL (26%) followed by SDEP (16%), while the MAE is controlled by SDEP (24%) and ZSNL (21%). Further, some parameters with moderate influence exhibit different behaviour/contribution as per the used metric. For instance, the contributions of XSLP and LAMN to BIAS[ALT] at JMC are 10% and 9%, which become 7% and 8% while considering MAE[ALT]. Such slight variation in parameter sensitivity ratio is amplified when calculating the cumulative influence for each family of parameters (**Fig. 9** for BIAS experiments and for **Fig. A3** MAE experiments). For example, the ponding and snow-cover parameters contribute 57% of the variability of Tmin BIAS for HPC site, altering to 47% for MAE, noting that the difference is distributed among the other parameter groups. Besides, the two metrics identify the same insensitive parameters throughout all experiments. Thus, we will focus on reporting the detailed sensitivity results for one metric, *i.e.* BIAS.

ZSNL has generally the most influence (~20%-55%) on ALT, Tmin, and Tmax sensitivities at the three sites, except Tmax of HPC site, for which ZSNL becomes the third most important parameter (~12%). Evidently, the clear agreement among the three sites stresses the importance of surface insulation (represented by ZSNL) in controlling the thawing and freezing fronts' depth and its annual temperature value. However, XSLP is the most influential parameter on Tmax sensitivities at the HPC site, contributing ~20% to BIAS[Tmax] variability. Although XSLP is recognized as a hydraulic parameter with no implicit impact on the thermal system, XSLP is crucial in the current 1-D (vertical) simulation because it determines the amount of water exiting the soil column via inter/surface flow mechanisms, which (as noted above) updates water stores and hence the hydraulic and thermal regime of the system.

Overall, Tmin is primarily dominated by three parameters, ZSNL, XSLP, and DD, contributing ~70% of the BIAS. The varying contribution of the three parameters is attributed to the combined impact of external forcing and parameter interactions among the sites. For example, compared to the HPC and JMC sites, the influence of ZSNL reduces (from ≥40% to 30%) at the cost of increasing the effect of DD and XSLP (~15% collectively) at the BWC site. Besides, the cold and dry conditions at HPC make the aerial snow coverage specification more pivotal to the insulation against heat loss in winter, which controls around half of the variability in Tmin. Further, the limited precipitation at HPC (247 mm/year) combined with shallower organic depth (ODEP<1m), little organic content (<15%), and less porosity (higher sand and clay contents) resulted in damping the impact of drainage and runoff parameters (XSLP and DD).

Tmax is highly affected by the amount and extent of accumulated snow (represented by ZSNL) in the previous winter. The magnitude of influence is similar for BWC and JMC sites (~35%), while the impact decreases at the colder HPC site to ~10%. This discrepancy is mainly driven by the external climate variables and the range of soil texture parameters, as highlighted earlier. Further, LNZ0 (implicitly the vegetation height) controls the simulation of Tmax, most intensely at HPC (~15%) compared to the other sites (~5% BWC and ~2% JMC), noting that LNZ0 affects canopy storage, interception, and latent/sensible heat flux partitioning. On the other hand, ORGM*i* has the utmost impact on Tmax at the JMC site, with a contribution of ~ 45% of the BIAS, noting that the other two sites showed less sensitivity to ORGM*i* (HPC ~ 1% and BWC ~7%). JMC site has a boarder range of perturbation for ORGM*i* (0-60%) compared to (0-30% BWC and 0-15% HPC) and for ODEP (0.1-7m) compared to (0.1-1m for both BWC and HPC) that match the site characteristics (Ednie et al., 2013; Smith et al., 2004, 2009, 2010; Smith, Chartrand, Duchesne, & Ednie, 2016; Smith, Chartrand, Duchesne, Ednie, et al., 2016). In contrast, the opposite case is associated with SDEP, as the upper limit for the JMC site is ~7m, which is significantly shallower than BWC (15.2m) and HPC (20.2m). The three sites share the same lower perturbation limit for SDEP of 2 m, defined from the maximum possible root depth.

Although ALT is extracted from Tmax, it does not depict the same sensitivities for all model parameters. For instance, SDEP is surprisingly crucial only for ALT, with a relatively negligible impact on Tmin and Tmax. On average, SDEP contributes ~14% of BIAS[ALT] variability at the three sites. Still, ZSNL exerts the most influence on ALT, with a ratio of sensitivity varying between 20% and 28% at the three sites. The other high-to-moderate parameters among ALT and Tmax experiments are identical in ranking but differ in individual contributions, including DD, XSLP, and ORGM*i* parameters. Notably, both LAMN and LNZ0 play a moderate role in the variability of the ALT, which collectively accounts for 10%-15% at the three sites. These two parameters have an almost similar magnitude of contribution for Tmax, with a slightly lower effect on Tmin (<9%). In general, the outcome of these comprehensive analyses reinforces the importance of surface insulation (LNZ0, LAMN and ZSNL) and subsurface regulation of heat by soil texture (ODEP and ORGM*i*), SDEP, and the drainage efficiency of the system (DD and XSLP).

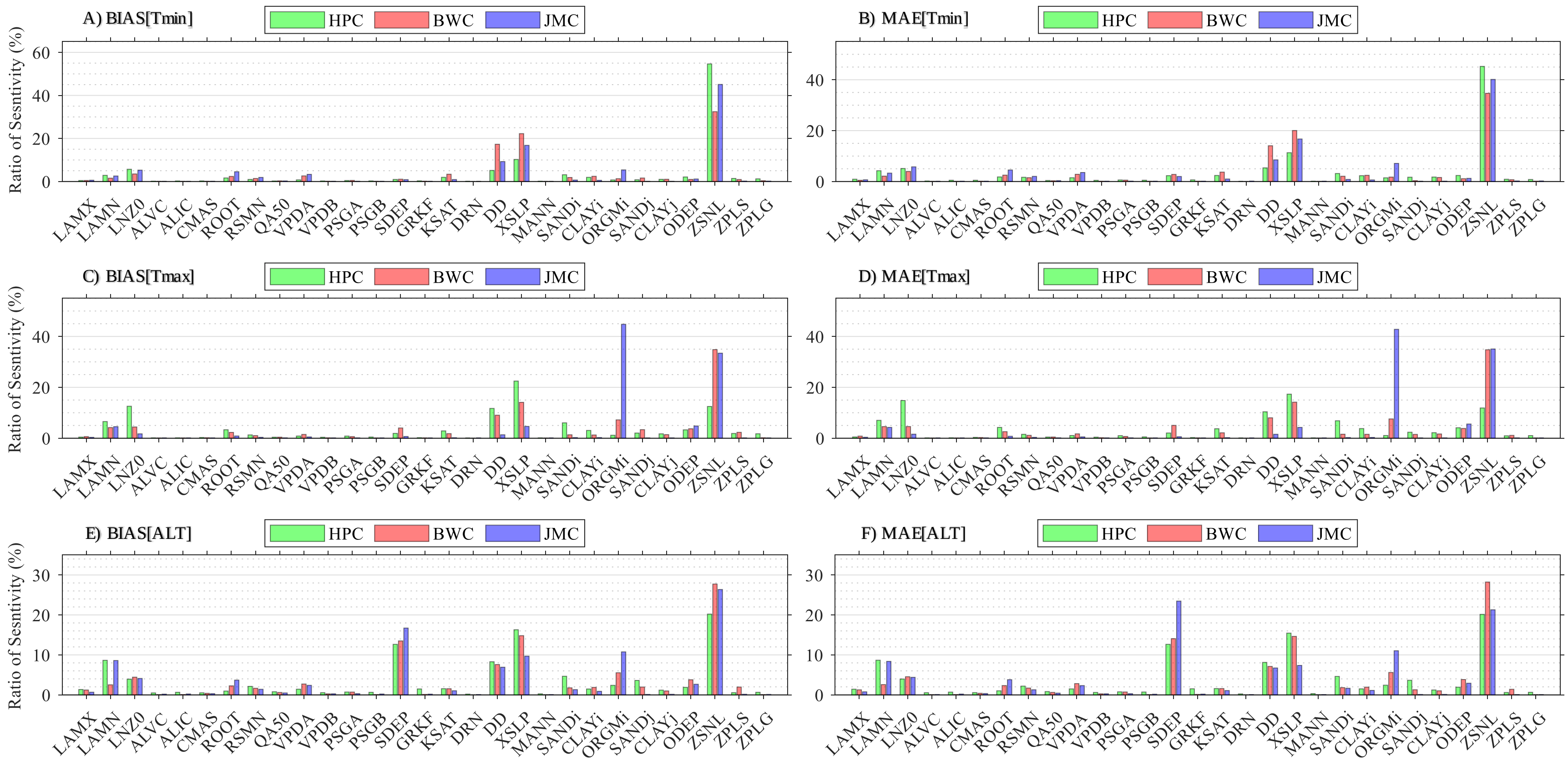



**Fig. 8.** Ratio of sensitivity of $IVARS_{50}$ for each parameter in all experiments for A) BIAS[Tmin], B) MAE[Tmin], C) BIAS[Tmax], D) MAE[Tmax], E) BIAS[ALT], and F) MAE[ALT]. Ratio of sensitivity of a parameter is calculated as the ratio of its respective sensitivity ($IVARS_{50}$) to the sum of the sensitivity indices of all model parameters (the ratios of sensitivity for the 29 parameter sum to one).

The study shows that a limited number of model parameters dominates the majority of the response surface variation across the selected error metrics and permafrost variables. Thus, in order to reduce the uncertainty in model predictions, it is crucial to identify/fine-tune these few but highly influential model parameters. Field observations and gridded remotely-sensed products provide a feasible approach in this regard (*e.g.* LAI, ORGM, SDEP), but this is not directly applicable/practical for all model parameters (*e.g.* ZSNL). Inference could be needed to relate some parameters to measured ones.

A substantial reduction in the number of free parameters (for model calibration) is achievable by fixing the values of insensitive parameters (*e.g.* to the median of a priori distribution), which reduces model dimensionality and computational cost of model calibration runs. Accordingly, a list of the most influential parameters for different permafrost aspects in MESH/CLASS is provided in **Table 5**. We report the 'very' important parameters, which have a ratio of sensitivity larger than 10% (similar to Haghnegahdar *et al.*, 2017). Since there are 18 GSA experiments (*i.e*. three sites x three permafrost variables x two performance metrics), the parameters listed in **Table 5** summarize the experiments' union at the HPC, BWC, and JMC sites. There are six 'very important' model parameters, namely ZSNL, ORGM*i*, XSLP, DD, SDEP, and LNZ0. Further, as a secondary goal of the sensitivity analysis, the following model parameters are entirely insensitive (<1% contribution): ALVC, ALIC, CMAS, GRKF, MANN, DRN, PSGA and PSGB.

**Table 5.** List of important parameters (ratio of sensitivity ≥1%) of the MESH model based on the global sensitivity analysis for different variables (and error metrics) and overall (union of all). Very important parameters (ratio of sensitivity ≥10%) are highlighted in bold.

| Rank | Tmin | Tmax | ALT | Overall |
|---|---|---|---|---|
| 1 | **ZSNL** | **ORGM*i*** | **ZSNL** | **ZSNL** |
| 2 | **XSLP** | **ZSNL** | **XSLP** | **ORGM*i*** |
| 3 | **DD** | **XSLP** | **SDEP** | **XSLP** |
| 4 | LNZ0 | **LNZ0** | **ORGM*i*** | **DD** |
| 5 | ORGM*i* | **DD** | LAMN | **SDEP** |
| 6 | ROOT | LAMN | DD | **LNZ0** |
| 7 | LAMN | SAND*i* | SAND*i* | LAMN |
| 8 | VPDA | ODEP | ODEP | SAND*i* |
| 9 | KSAT | SDEP | LNZ0 | ODEP |
| 10 | | ROOT | ROOT | ROOT |
| 11 | | KSAT | SAND*j* | SAND*j* |
| 12 | | CLAY*i* | VPDA | VPDA |
| 13 | | SAND*j* | | KSAT |
| 14 | | | | CLAY*i* |

In order to better understand the similarities and differences of GSA results across all BIAS experiments, the cumulative impact of each group/family of parameters (see **Table 2** and **Section 2.4.1**) is presented by pie-charts in **Fig. 9** (refer to **Fig. A3** for MAE results). This reveals that model performance at both JMC and HPC sites is mainly dominated by snow cover and drainage/runoff parameters, which collectively contribute ~72% to the variability in BIAS[Tmin], noting that the snow cover group has an absolute higher influence (55% for HPC and 45% for JMC). On the other hand, performance at the BWC site is primarily dominated by the same two parameter groups (75%), with a different proportional impact since the drainage/runoff group has more influence (43%) on the simulation of BIAS[ALT]. The ponding and permeable depth groups have minimal impact on Tmin throughout all the experiments, contributing to a maximum of 4% of Tmax variability. Unlike Tmin, the three sites depict a very different behaviour for Tmax in terms of the controlling group of parameters. For example, HPC is highly sensitive to drainage/runoff and canopy groups (37%+27%), which is not the case for

BWC that is dominated by snow cover and drainage/runoff group (35%+25%), nor for the JMC site that is controlled by soil texture and snow cover (50%+33%). The ponding and permeable depth groups have a minimal impact on Tmax for all experiments, with a maximum contribution of 5%. Lastly, the ALT GSA experiments yield relatively identical partitioning of parameters since no single group dominates more than 30% of the BIAS[ALT] variability. The ponding group of parameters has the most negligible impact on ALT (≤2%), while the permeable soil depth becomes more influential to ALT than Tmax and Tmin.

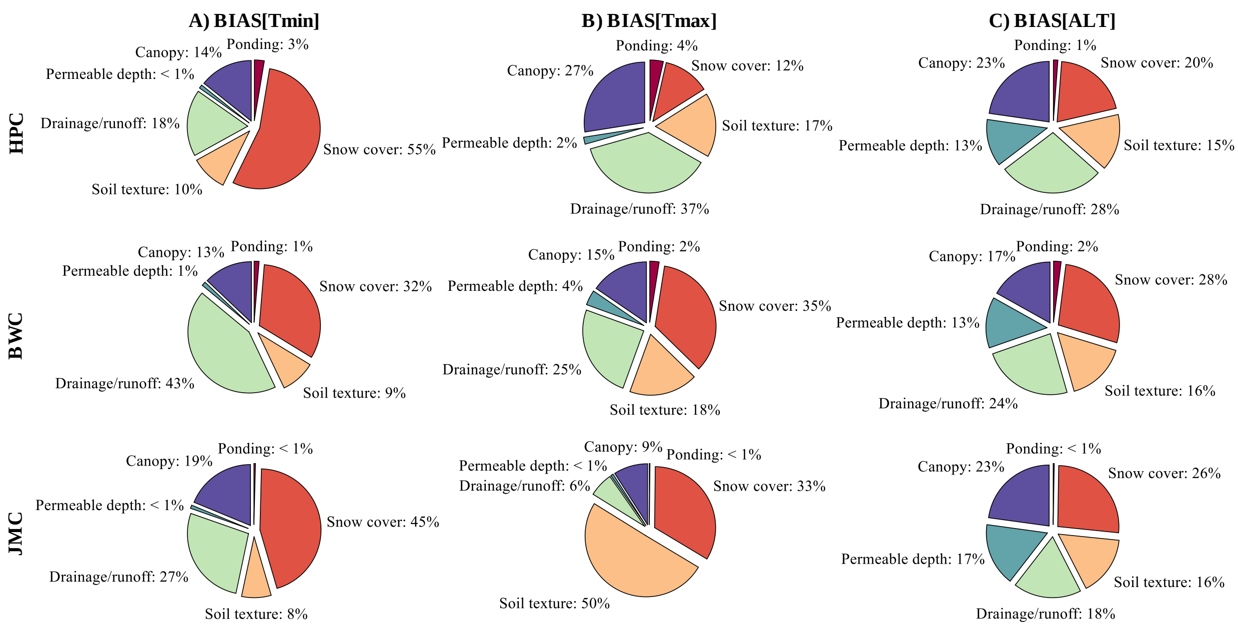



**Fig. 9.** Total ratio of sensitivity (based on $IVARS_{50}$) for each group of parameters (as indicated in **Table 2**) for the BIAS across all experiments.

## 3.4 Parameter identifiability

This section explores the degree of identifiability of MESH parameters in simulating ALT, Tmin, and Tmax. Instead of presenting identifiability results for each permafrost variable separately, we show parameter sets that simultaneously fulfil the behavioural threshold for the three permafrost variables. This reduces the explored response space and focuses on the global maxima regions. Further, instead of screening/selecting the best model runs based on an arbitrary percentage or number of the total simulations (*e.g.* 10% as in Wagener et al. (2003) or the best five parameter sets as in Houser et al. (2001)), we opted for the traditional approach in which the screening follows predefined thresholds of performance metrics. Employing this technique is not free of subjectivity in defining these thresholds; but we believe that filtering by a performance metric is relatively less subjective, which was applied for the MAE experiments as follows: ≤0.3m for MAE[ALT], and ≤1°C for MAE[Tmax/Tmin]. However, the experiments depicted a high discrepancy among sites for the BIAS case, making identifying and unifying these thresholds challenging; having a reasonable number of parameter sets was the other criterion in this assessment. Thereby, each site has a set of distinctive behavioural thresholds used for the identifiability analysis of its model parameters, as summarized in **Table 6**.

**Table 6**. Summary of behavioural thresholds per site and performance metric and the total number of parameter sets that fulfill such constraints.

| Site | Criterion | BIAS | | MAE | |
|---|---|---|---|---|---|
| | | Range | No. of sets | Range | No. of sets |
| HPC | Tmin & Tmax | ±0.5°C | 9 | ≤1°C | 7 |
| | ALT | ±0.25m | | ≤0.3m | |
| BWC | Tmin & Tmax | ±0.15°C | 34 | ≤1°C | 89 |
| | ALT | ±0.05m | | ≤0.3m | |
| JMC | Tmin & Tmax | ±0.25°C | 75 | ≤1°C | 13 |
| | ALT | ±0.1m | | ≤0.3m | |

The range of behavioural parameter sets for all sites, scaled between zero and one, is shown in **Fig. 10** (BIAS case) and **Fig. 11** (MAE case). ZSNL, the most influential parameter in most GSA experiments (see **Fig. 9**), is broadly identifiable (more constrained) among sites with a slightly different setting within its feasible range for each site. For instance, employing the MAE as the performance metric, the HPC model has an identifiable range for ZSNL of 0.6-0.8, while BWC and JMC have identifiable ranges of 0.45-0.7 and 0.55-0.95, respectively (**Fig. 11**). The three sites share the same identified normalized ZSNL value of ~0.7 (*i.e.* equivalent to ZSNL of 0.35m) with a high probability of occurrence (*i.e.* darker dots in **Fig. 11**). However, the analysis at JMC via the BIAS (**Fig. 10**) reveals a potential issue as the identifiable normalized value for the ZSNL was one (*i.e.* upper end of the range), probably highlighting an inadequately defined parameter range and/or other structural non-identifiability for the JMC model setup. Other experiments show similar behaviour of having an identifiable parameter value at the lower or upper limit of the feasible range (*e.g.* ORGM*i* of BWC with MAE); still, the proportion of such problematic parameters is insignificant among all experiments.

Irrespective of the employed performance metric, highly sensitive parameters (shaded in light red in **Fig. 10** and **Fig. 11**) other than ZSNL do not always depict clear identifiability. For example, the range of SDEP for the BWC setup covers almost the whole feasible space; 0.1-0.9 via BIAS and 0.1-0.75 via MAE. The HPC setup sustains the same issue of poorly identified SDEP, while it is relatively identifiable for JMC. The inability to determine such a critical parameter affects both the thermal and hydrologic simulation of MESH (Elshamy et al., 2020; Haghnegahdar et al., 2017). Further, the highly sensitive hydraulic parameters, XSLP and DD, are not well identified, especially the DD parameter for the BWC setup; XSLP is moderately constrained for HPC and JMC via MAE and BIAS, respectively. Further, the range of LNZ0 parameter for HPC differs depending on the utilized performance metric; MAE reduces the identifiability range (from 0.2-0.9 to 0.3-0.55). Lastly, a negligible impact of the employed metric on the well-identified ORGM*i* at the JMC site is observed.

Similarly, moderately influential parameters (shaded in light blue in **Fig. 10** and **Fig. 11**) do not depict a consistent response to the employed performance metric. For example, the identifiable range of LNZ0 for BWC is 0.5-1.0 and 0.05-0.35 for MAE and BIAS, respectively, highlighting the significant impact of the selected metric on the estimated parameter value. However, the behavioural value for the DD parameter for JMC setup does not change with the selected metric. Noting that a clear improvement on parameter identification for the JMC site is achieved when employing the MAE for the filtering/selecting behavioural solutions, including sensitive and insensitive parameters. This site shares the same behavioural solutions among BIAS and MAE experiments (*i.e.* same darker dots in **Fig. 10** and **Fig. 11**), which is not the case for the other two sites. Besides, the rest of the moderately influential parameters are less identified at HPC and BWC, especially LAMN, VPDB, and KSAT. Therefore, not every highly/moderately influential parameter is identifiable for all experiments/sites, underlining the potential parameter interaction and non-uniqueness of fluxes/states partitioning, hence the simulated permafrost.

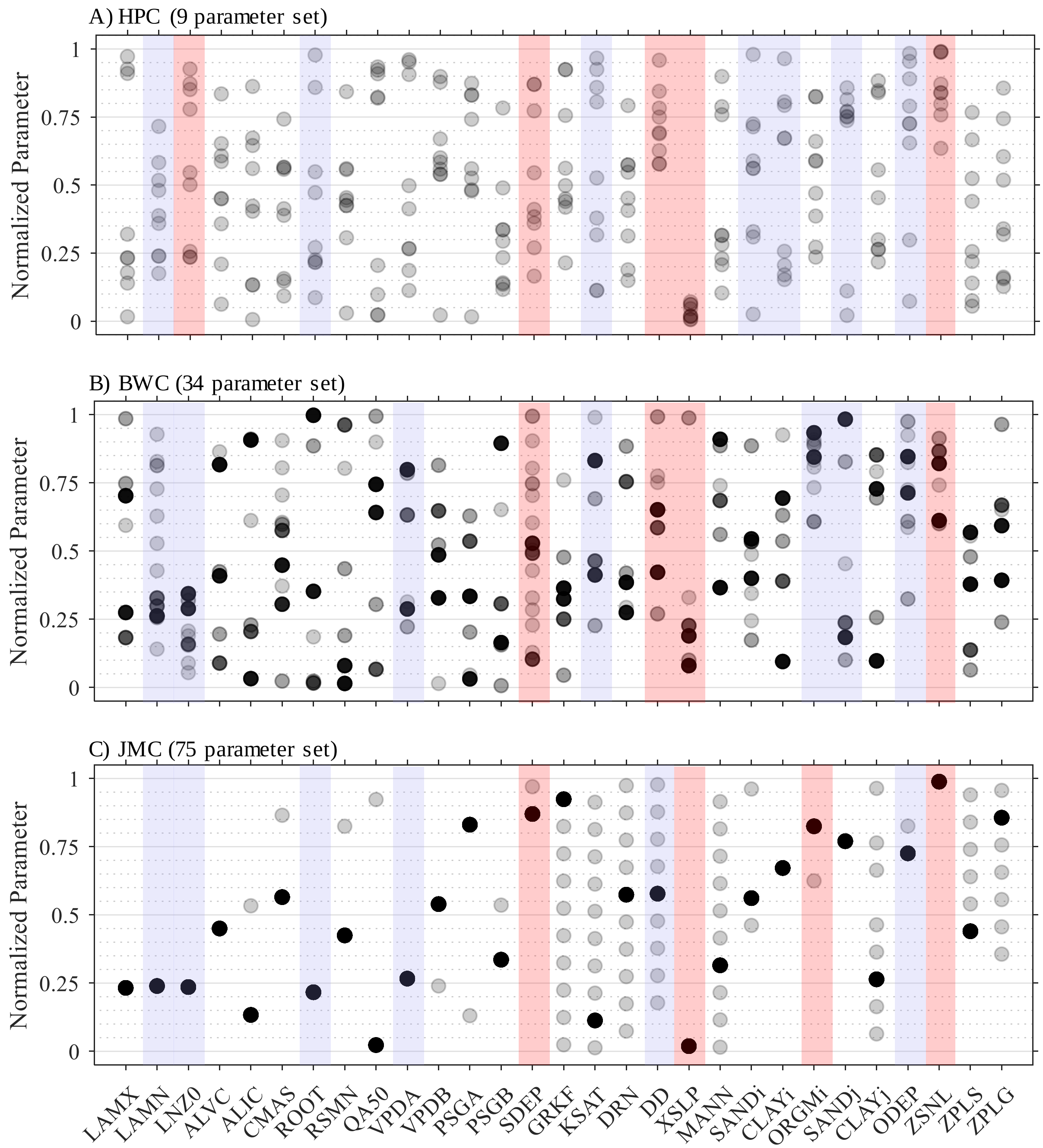



**Fig. 10.** Summary of parameter identifiability that simultaneously satisfies the ALT, Tmax, and Tmin performance criteria at A) HPC, B) BWC, and C) JMC sites. The behavioural parameter sets are filtered by BIAS performance error, where the number of parameter sets is 9, 34, and 75 for HPC, BWC and JMC sites, respectively; Light red shade refers to a high influential parameter (ratio of sensitivity ≥10%); Light blue shade refers to moderate influential parameter (10%> ratio of sensitivity ≥1%); Transparent markers/points are used that get darker when coinciding or overlapping.

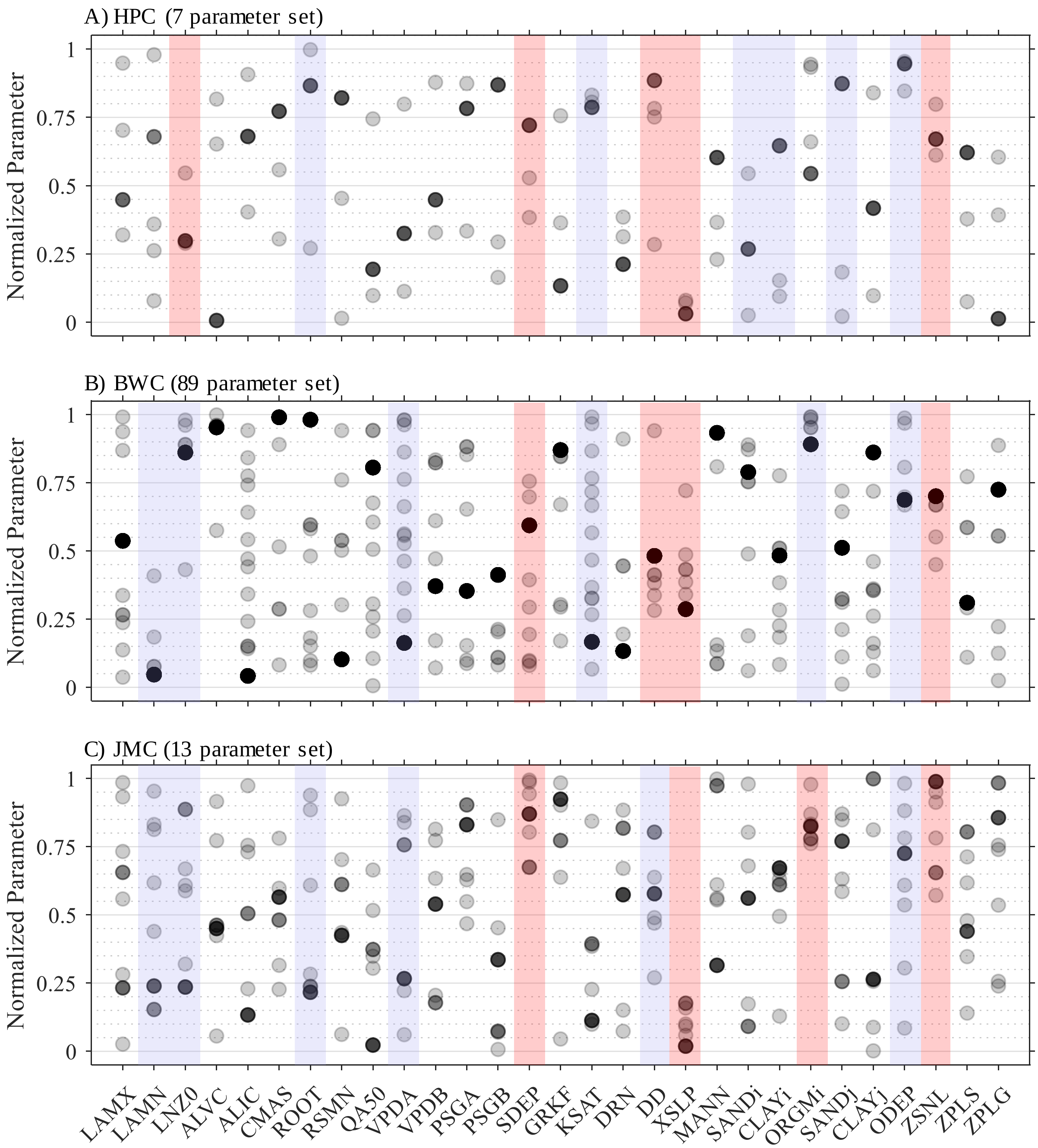



**Fig. 11.** Same as **Fig. 10**, but showing the MAE-based identifiability results, where the number of behavioural parameter sets is 7, 89, and 13 for HPC, BWC and JMC sites, respectively.

### 3.5 Model performance assessment

This section examines different facets of permafrost dynamics for the selected behavioural solutions based on the MAE metric at the JMC site (*i.e.* 13 parameter sets shown in **Fig. 11C**). This model experiment has long permafrost observations (1985-2000) and reasonably identifiable model parameters. In detail, **Fig. 12A-D** compare the uncertainty envelopes of ALT, MAGTp, PB and DZAA (refer to **Section 2.1** for definitions) with respect to their observed counterparts, extracted mainly from available temperature profiles. Although the maximum allowable average deviation from observed ALT was specified as 0.3m, as highlighted in **Table 6**, the envelope of variability corresponding to parameter sets that fulfill such conditions fluctuates between 0.3m-1m, broadly encapsulating the observations. The considerable range of variability is attributed to employing an aggregated performance metric (MAE in this case) that compresses the dynamical response of the system in time into a single objective function. One way to address such loss of information is by utilizing non-equal weights (or a form of penalty function) while performing the aggregation to constrain the filtering and ensure a behavioural solution replicates the observed behaviour. Another solution can be found in studying the identification of model parameters with finer temporal resolution using, *e.g.*, the DYNamic Identifiability Analysis (DYNIA: (Wagener et al., 2003)). However, even such a thorough approach is challenged by the unique nature of permafrost, as most of its descriptive variables are annual-based which limits the number of points used to calculate the metrics and are regulated by complicated interactions between surficial, sub-surficial and meteorological drivers.

The selected behavioural solutions produced warmer and thinner permafrost, as depicted in **Fig. 12B&C**. The simulated MAGTp is consistently higher than observations by around 1°C (uncertainty range: 0.25°C-1.75°C), except for 1996 and 1997, where few parameter sets could capture observed temperatures, noticing that warmer Tmin and/or Tmax reflect such overestimation in permafrost temperature. Similarly, but to a lesser extent, simulated PB is continually underestimated (*e.g.* two-folds during the 1990s onset) over the simulation period, with an exception at the beginning (1986-1988) and near the end (1995-2000). The envelope of PB depicts high interannual variability that is not shown by observations. Lastly, the DZAA is closely captured with an average variance of 1.5m, as shown in **Fig. 12D**. Noting that DZAA plays a pivotal role in simulating permafrost as it assesses the suitability of the selected soil column depth and identifies the presence of permafrost via calculating its corresponding soil temperature (TZAA) (Burke et al., 2020; Sapriza-Azuri et al., 2018).

**Fig. 12E&F** provide the temporal evolution of the BIAS and MAE of Tmax and Tmin, which were aggregated above to assess model uncertainty, sensitivity, and identifiability. Three distinctive remarks could be drawn for these figures. First, Tmin tends to be mostly colder than observed as indicated by the BIAS sign, while Tmax envelope is generally warmer. Secondly, Tmin depicts a higher uncertainty range and interannual variability compared to Tmax. Lastly, no clear trade-off between Tmax and Tmin was observed while using MAE or BIAS performance criteria due to error compensation. Although ALT is extracted from Tmax, it does not depict identical variance for behavioural model parameters compared to Tmax, which is compatible with sensitivity analysis results in this regard (**Section 3.3**).

**Fig. 12G** presents the temporal evolution of the date of maximum thaw. Remarkably, there is a wide range of uncertainty (100-150 days) for the date, driven by two distinctive clusters of simulations as shown by the solid lines (each one corresponds to a single parameter set) in **Fig. 12G**. Noting that four parameter sets yielded an earlier (questionable to happen) date between May (DOY-130) and June (DOY-180) throughout the simulation period, which cannot be verified or falsified due to the unavailability of thaw timing data. In contrast, the other cluster simulated the date of maximum thaw more reasonably, with values varying between DOY-250 and DOY-300; the years 1992 and 1993 were an exception as most of the cluster's members had earlier max thawing date (~DOY-200). Such a major discrepancy in maximum thaw timing highlights a potential challenge in LSM applications, even under a constrained parameterization for ALT and temperature envelopes, accentuating the need for additional constraint(s) for a better simulation of the freeze/thaw cycles.

**Fig. 12H** displays the envelopes of surface and thermal offsets (refer to **Section 2.1** for definitions). The figure reveals high variability for the surface offset above the ground, dominated by snow accumulation in winter and the shading effect of the canopy in summer. The envelope of surface offset has positive values with a varying mean of 5°C-10°C and interannual variability between of 2.2°C-5.5°C. Besides, the surface offset is the summation of winter and summer offsets, that when apportioned, highlights the significant influence of snow on keeping permafrost from losing heat in winter over the impact of canopy shading in summer. Winter offset could vary by up to ~10°C with an enormous interannual variability similar to that of MAGTp (**Fig. 12B)**, while summer offset has a weaker variance of ~5°C and a smaller interannual variability, as shown in **Fig. A4**. On the other hand, the thermal offset, occurring in the soil above the permafrost's horizon, displays a confined uncertainty in order of 0.1°C-1°C, noting that the thermal properties of soil texture above the permafrost table do exert the primary influence on the thermal offset and JMC is characterized by the presence of organic soil – ORGM*i* was well-identified according to the identifiability analysis for the JMC site based on MAE (**Fig. 11C**).

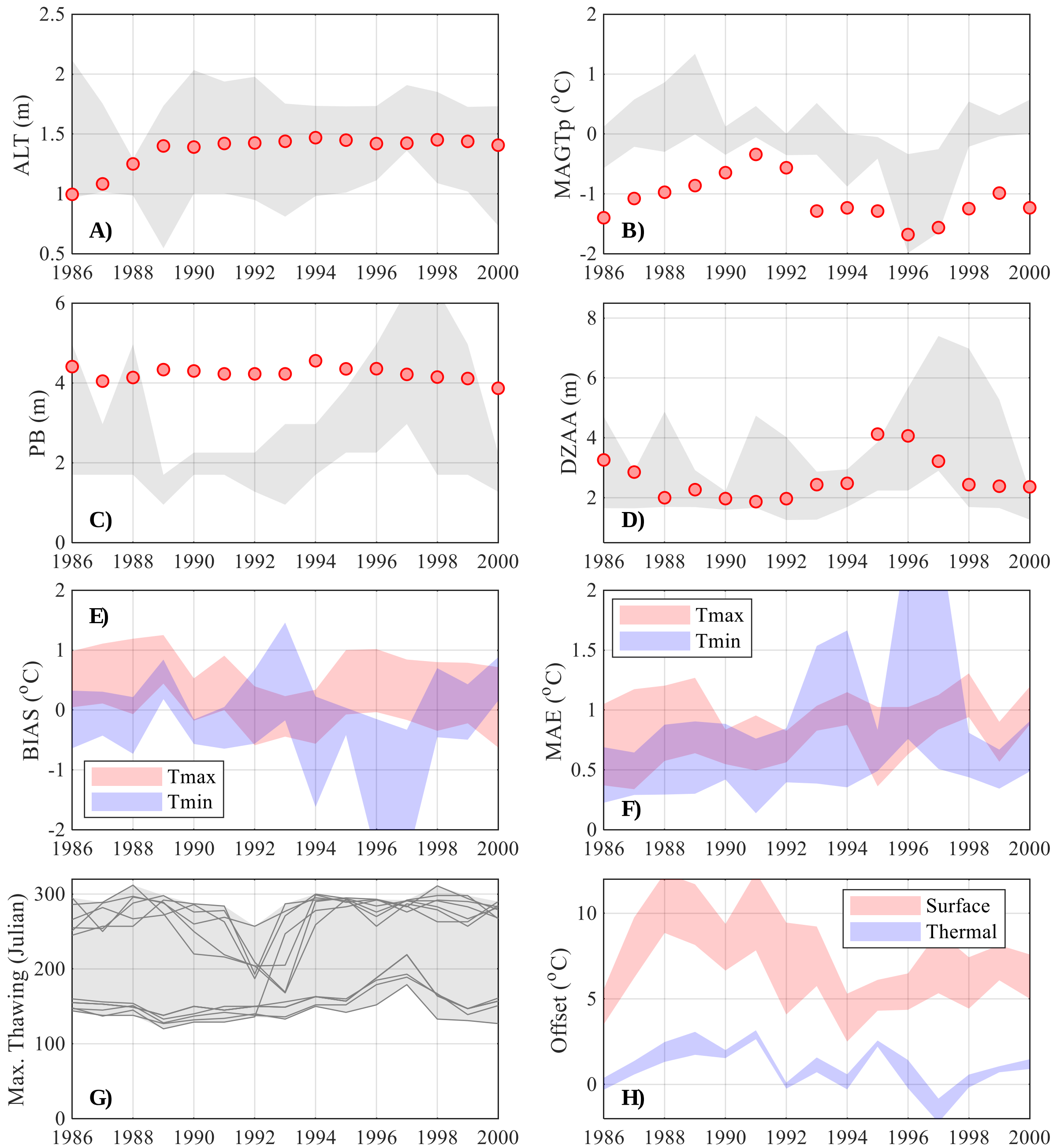



**Fig. 12.** Temporal evolution and the associated range of uncertainty for A) ALT, B) Mean Annual Ground Temperature at the top of permafrost (MAGTp), C) depth to the Base of Permafrost (BP), D)

Depth to the Zero Annual Amplitude (DZAA), E) BIAS of simulated ALT, F) MAE of simulated ALT, G) Date of Maximum thawing, and H) Surface and Thermal Offset. All presented variables correspond to the identifiable JMC experiments via the MAE (13 parameter set); refer to **Fig. 2** for a brief description of permafrost variables; shading denotes the range of uncertainty, while red marks denote the available observation; each solid line in subplot G corresponds to a single model evaluation (parameter set).

**Fig. 13** compares three different simulated temperature profiles for 1993, 1997, and 2000 at the JMC site for the behavioural parameter sets filtered by the MAE. The year 1993 (**Fig. 12** and **Fig. 13A**) has sound results for the ALT, DZAA, Tmax, and thermal offset at the cost of having warmer permafrost (overestimated MAGTp and Tmin by ~1°C), shallower permafrost thickness (underestimated PB by 2m), and highly divergent surface offset. On the other hand, the year 1997 (**Fig. 12** and **Fig. 13B**) gives good agreement for ALT, MAGTp, PB, DZAA, and a smaller uncertainty range for the maximum thaw date in exchange for the highest observed underestimation of Tmin by 2°C in addition to a considerable variation of surface offset by 5°C. Lastly, the year 2000 (**Fig. 12** and **Fig. 13C**) has balanced results for ALT, DZAA, Tmin, and Tmax, producing warmer (MAGTp overestimated by 1.2°C) and shallower permafrost thickness (PB underestimated by 2m) in addition to the immersive variability for the date of maximum thaw (DOY130-DOY290). Such comparison highlights the difficulty/complexity that modellers encounter while employing LSMs for permafrost-based applications, either investigative or predictive, regardless of the assessment scale.

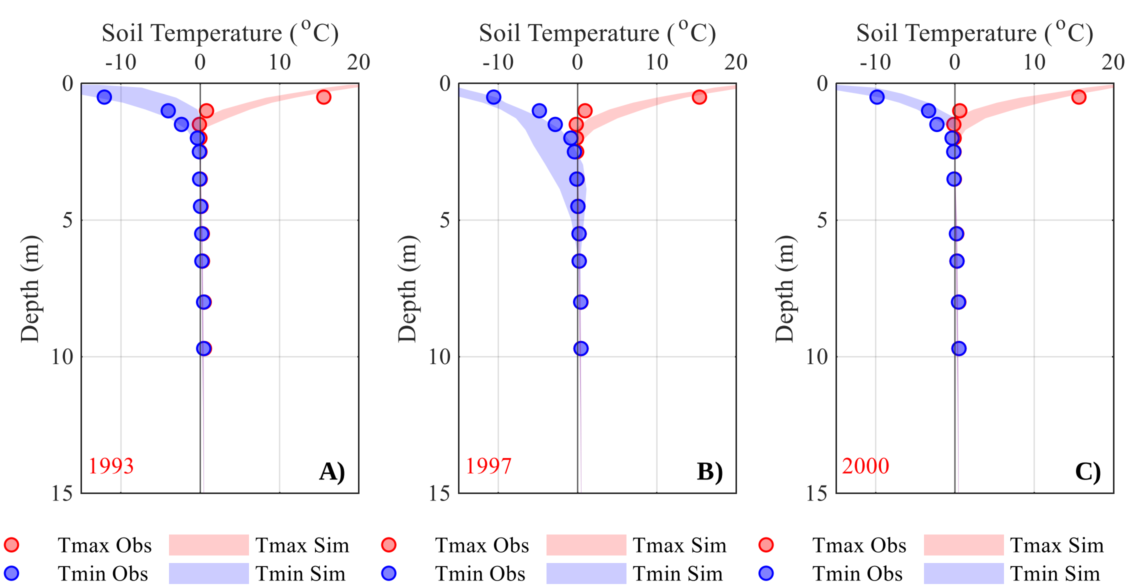



**Fig. 13.** The observed and simulated envelopes of temperature profiles (Tmin and Tmax) at the JMC site using identifiable parameters for MAE criteria at A) 1993, B) 1996, and C) 2000.

## 4 Summary and conclusions

Following extensive efforts to improve the realism of process representation in land surface models (LSMs), their complexity and dimensionality has increased remarkably, complicating model/parameter identification. Moreover, simulating the dynamics of perennially frozen soil, or permafrost, is further challenged by the significant thermal/hydraulic memories of a deep soil column and the limited availability of ground observations. We note that the simulation of the thermal regime dynamics of the soil column in cold regions directly influences the partitioning of water and energy fluxes, which, if not well constrained, could yield deceptive future projections for climate and hydrology. We inspected three interrelated issues regarding permafrost simulation in the MESH-CLASS LSM; the impact of input uncertainty (forcing data), the sensitivity of simulations to model parameters, and the identifiability of model parameters. Three experimental sites within the Canadian Mackenzie River Basin (MRB) were employed in the current study, characterized by various climatic conditions and permafrost zonation. We finally assessed model performance at one of the sites using a large set of permafrost characteristics versus observations as available.

The combined impact of climate forcing and parameter uncertainty on permafrost dynamics was assessed for various climate forcing data sets, characterized by different temporal/spatial resolutions and forecasting/reanalysis methods. Such characteristics can play a pivotal role in initializing model states, parameter/model identification/sensitivities, and the subsequent simulation quality. Three meteorological datasets were considered for the study; WFDEI, WFDEI-GEM-CaPA, and WFDE5. Comparing the three forcing datasets to ground stations highlighted the significant uncertainties that could be introduced to permafrost simulation due to persistent bias in air temperature and precipitation. Besides, it underlined associated issues with forcing datasets, such as a repeated climatology-based (WFDEI-CRU variant) precipitation when there are no data available, in addition to the noticeable warm bias of air temperature in the WFDEI-GEM-CaPA dataset, which exceeded the adiabatic lapse late correction (noting that dataset provided data at 40m altitude).

The cumulative frequency distributions (CDF) for the averaged BIAS and MAE of Tmin, Tmax, and ALT were utilized for evaluating the experiments. No single dataset could collectively provide satisfactory model simulations for the three characteristics at the three sites. For instance, WFDEI-GEM-CaPA can reproduce the observations for Tmin and ALT using the BIAS and MAE criteria at the three sites, except for ALT at BWC, where WFDEI slightly outperforms. Thereby, the WFDEI-GEM-CaPA dataset was nominated for forcing the models of the three sites, to avoid inconsistent results for parameter sensitivity and identifiability if different forcing sets were used for the different sites.

Different global sensitivity analysis (GSA) experiments were implemented using the variogram-based framework (VARS) to study the degree of sensitivity of permafrost variables to the perturbation of the MESH-CLASS parameters. Understanding, diagnosing, and developing models for permafrost simulation were the motives for such vital analysis, given the unavailability of a comprehensive (formal) GSA in the context of land surface modelling of permafrost. The traditional metric-based time-aggregate GSA was employed for different aspects of permafrost dynamics, noting that model crashes were handled by a model emulation-based substitution technique. The experiments accentuated the dominant role of parameters (ratio of sensitivity ≥10%) that describe heat insulation at the vegetation-soil interface, such as ZSNL, LNZ0, and ORGM*i*, and those that control the runoff generation processes, such as SDEP, DD, and XSLP. The ranking and contribution of these parameters vary among experiments based on the incorporated response surface variable of the GSA. Further, the study provides a list of the highly sensitive parameters for different permafrost characteristics. Remarkably, the water ponding-related parameters possess a limited-to-negligible influence for all GSA experiments. The study highlights model parameters that should be carefully fine-tuned and those with negligible impact on output variability that, if fixed (*e.g.* to the median of parameter range) and excluded from any subsequent model calibration, could reduce model dimensionality, the associated computational cost, and enhance parameter identification.

Parameter identifiability was also investigated in a multi-objective fashion to examine model parameterization and fidelity with a different lens, complimentary to sensitivity and uncertainty analyses. An approach incorporating additional constraints on permafrost simulations and tight behavioural thresholds (based on a predefined measure of model performance) yielded a small number of parameter sets that satisfy the multi-objective criteria (*i.e.* 7-89 out of 26,200 sets). The analysis underscored that not all highly and moderately (10%> ratio of sensitivity ≥1%) sensitive parameters were clearly identifiable among all experiments. Besides, the identifiable value/range for sensitive parameters differs among sites, highlighting the massive impact of external forcing and predefined parameter ranges.

Permafrost dynamics, represented by various facets, were extensively examined for one of the experiments (*i.e.* the MAE-based JMC experiment) to explore the uncertainty corresponding to these designated behavioural solutions. Even though ALT, Tmin, and Tmax are mainly replicated by the behavioural parameter sets with significant interannual variability, other descriptive factors of permafrost dynamics were not appropriately reproduced, such as MAGTp, PB, the date of maximum thaw and the surface offset. Further, a qualitative comparison of the simulated temperature profiles (at different years) and other permafrost variables highlighted the challenges that modellers encounter while configuring LSMs for permafrost-related applications.

Despite the fact that the outcomes of this study were specific to the MESH-CLASS model and limited to the selected evaluation sites and methods, they are practically beneficial for advancing modelling practices, especially permafrost-related applications. The study highlighted the complexities and challenges of LSM application in cold regions and shed light upon the possible approaches to address such obstacles. That being said, a joint multi-objective GSA and multi-objective identifiability analyses promote an improved understanding of LSM structure, reduces predictive uncertainties, and facilitates efficient model calibration. Further, there is a pressing need to develop improved forcing datasets that rectify the problems of the current versions of datasets in terms of systematic biases and lack of interannual variability in air temperature and precipitation for some datasets; other meteorological variables were not assessed due to data availability constraints. Besides, additional improvement is required in the MESH-CLASS model to enhance the realism of permafrost simulation, as reflected by the simulation's quality/uncertainty, *e.g.* producing cooler Tmin and misrepresenting surface/thermal offsets, possibly due to insufficient insulation. Proposed modifications include the snow component which is still simulated via a single layer and the surface canopies where no explicit treatment of canopy litter and moss is available. Lastly, future studies could be directed towards generalizing the outcomes of these analyses to other observational sites with more data and to other regions, as well as exploring the extendibility of the work to various regional and global models with varying complexity in large-scale applications.

**Acknowledgement**

This research was supported financially by the Canada Excellence Research Chair (CERC) Programme in Water Security, Integrated Modelling Program for Canada (IMPC) and Razavi's Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant. The authors would like to thank the Information & Communications Technology (ICT) at the University of Saskatchewan, for providing continuous support in using the High-Performance Computing (HPC) research cluster, Plato and Copernicus.

## References


Abdelhamed, M. S., Elshamy, M. E., Wheater, H. S., and Razavi, S. (2021). Hydrologic-land surface modelling of the Canadian sporadic-discontinuous permafrost: initialization and uncertainty propagation. Authorea Preprints, April 26, 2021, under review, https://doi.org/10.22541/au.161941151.16015163/v1

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

Alexeev, V. A., Nicolsky, D. J., Romanovsky, V. E., & Lawrence, D. M. (2007). An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost. *Geophysical Research Letters*, *34*(9). https://doi.org/10.1029/2007GL029536

Andresen, C. G., Lawrence, D. M., Wilson, C. J., McGuire, A. D., Koven, C., Schaefer, K., et al. (2020). Soil Moisture and Hydrology Projections of the Permafrost Region: A Model Intercomparison. *The Cryosphere Discussions*, (14), 445–459. https://doi.org/10.5194/tc-2019-144

Asong, E., Elshamy, M., Princz, D., Wheater, H., Pomeroy, J., Pietroniro, A., & Cannon, A. (2020). High-resolution meteorological forcing data for hydrological modelling and climate change impact analysis in the Mackenzie River Basin. *Earth System Science Data*, *12*(1), 629–645. https://doi.org/10.5194/essd-12-629-2020

Barros, V. R., Field, C. B., Dokken, D. J., Mastrandrea, M. D., Mach, K. J., Bilir, T. E., et al. (2014). IPCC: Climate Change 2014 Impacts, Adaptation, and Vulnerability Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://doi.org/10.1007/s13398-014-0173-7.2

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*, 1–12. https://doi.org/10.1038/sdata.2018.214

Becker, W. (2020). Metafunctions for benchmarking in sensitivity analysis. *Reliability Engineering and System Safety*, *204*(August). https://doi.org/10.1016/j.ress.2020.107189

Beven, K. (2006). A manifesto for the equifinality thesis. In *Journal of Hydrology* (Vol. 320, pp. 18–36). https://doi.org/10.1016/j.jhydrol.2005.07.007

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

Burke, E. J., Zhang, Y., & Krinner, G. (2020). Evaluating permafrost physics in the Coupled Model Intercomparison Project 6 (CMIP6) models and their sensitivity to climate change. *The Cryosphere*, *14*(9), 3155–3174. https://doi.org/10.5194/tc-14-3155-2020

Chadburn, S. E., Burke, E. J., Essery, R. L. H., Boike, J., Langer, M., Heikenfeld, M., et al. (2015). Impact of model developments on present and future simulations of permafrost in a global land-surface model. *Cryosphere*, *9*(4), 1505–1521. https://doi.org/10.5194/tc-9-1505-2015

Chen, F., & Dudhia, J. (2001). Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. *MONTHLY WEATHER REVIEW*, *129*, 569–585.

Côté, J., Desmarais, J. G., Gravel, S., Méthot, A., Patoine, A., Roch, M., & Staniforth, A. (1998). The operational CMC-MRB global environmental multiscale (GEM) model. Part II: Results. *Monthly Weather Review*, *126*(6), 1397–1418. https://doi.org/10.1175/1520-0493(1998)126<1397:TOCMGE>2.0.CO;2

Cucchi, M., P. Weedon, G., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., et al. (2020). WFDE5: Bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*, *12*(3), 2097–2120. https://doi.org/10.5194/essd-12-2097-2020

Davison, B., Pietroniro, A., Fortin, V., Leconte, R., Mamo, M., & Yau, M. K. (2016). What is Missing from the Prescription of Hydrology for Land Surface Schemes? *Journal of Hydrometeorology*, *17*(7), 2013–2039. https://doi.org/10.1175/JHM-D-15-0172.1

DeBeer, C. M., Wheater, H. S., Carey, S. K., & Chun, K. P. (2016). Recent climatic, cryospheric, and hydrological changes over the interior of western Canada: A review and synthesis. *Hydrology and Earth System Sciences*, *20*(4), 1573–1598. https://doi.org/10.5194/hess-20-1573-2016

Devoie, É. G., Craig, J. R., Connon, R. F., & Quinton, W. L. (2019). Taliks: A Tipping Point in Discontinuous Permafrost Degradation in Peatlands. *Water Resources Research*, *55*(11), 9838–9857. https://doi.org/10.1029/2018WR024488

Dingman, S. L. (2015). *Physical hydrology*. *Waveland press.* (Vol. 2). https://doi.org/10.1177/030913337800200111

Dobinski, W. (2011). Permafrost. *Earth-Science Reviews*, *108*, 158–169. https://doi.org/10.1016/j.earscirev.2011.06.007

Ednie, M., Chartrand, J., Smith, S. L., Duchesne, C., & Riseborough, D. W. (2013). Report on 2011 field activities and collection of ground thermal and active layer data in the Mackenzie Corridor completed under Northwest Territories science licence #14918. *Geological Survey of Canada, Open File*, *7416*, 1–69. https://doi.org/10.4095/291982

Elshamy, M. E., Princz, D., Sapriza-Azuri, G., Abdelhamed, M. S., Pietroniro, A., Wheater, H. S., & Razavi, S. (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

Everdingen, R. O. (1998). *MULTI-LANGUAGE GLOSSARY of PERMAFROST and RELATED GROUND-ICE TERMS*. Calgary, Alberta, CANADA T2N 1N4. https://doi.org/10.2307/1551636

Gasset, N., Fortin, V., Dimitrijevic, M., Carrera, M., Bilodeau, B., Muncaster, R., et al. (2021). A 10km North American precipitation and land-surface reanalysis based on the GEM atmospheric model. *Hydrology and Earth System Sciences*, *25*(9), 4917–4945.

https://doi.org/10.5194/hess-25-4917-2021

Gibson, C. M., Chasmer, L. E., Thompson, D. K., Quinton, W. L., Flannigan, M. D., & Olefeldt, D. (2018). Wildfire as a major driver of recent permafrost thaw in boreal peatlands. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05457-1

Guillaume, J. H. A., Jakeman, J. D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., et al. (2019). Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software*, *119*(July), 418–432. https://doi.org/10.1016/j.envsoft.2019.07.007

Gupta, H. V., & Razavi, S. (2018). Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models. *Water Resources Research*, *54*(11), 8692–8717. https://doi.org/10.1029/2018WR022668

Haghnegahdar, A., Razavi, S., Yassin, F., & Wheater, H. (2017). Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty. *Hydrological Processes*, *31*(25), 4462–4476. https://doi.org/10.1002/hyp.11358

Harris, C., Arenson, L. U., Christiansen, H. H., Etzelmüller, B., Frauenfelder, R., Gruber, S., et al. (2009). Permafrost and climate in Europe: Monitoring and modelling thermal, geomorphological and geotechnical responses. *Earth-Science Reviews*, *92*(3–4), 117–171. https://doi.org/10.1016/j.earscirev.2008.12.002

Hegginbottom, J. A., Dubreuil, M. A., Harker, P. T. (1995). Canada, permafrost, The National Atlas of Canada, Natural Resources Canada, Geomatics Canada, MCR Series no. 4177, https://doi.org/10.4095/294765.

Hermoso de Mendoza, I., Beltrami, H., MacDougall, A. H., & Mareschal, J.-C. (2020). Lower boundary conditions in land surface models - effects on the permafrost and the carbon pools: a case study with CLM4.5. *Geoscientific Model Development*, *13*(3), 1663–1683. https://doi.org/10.5194/gmd-13-1663-2020

Hjort, J., Karjalainen, O., Aalto, J., Westermann, S., Romanovsky, V. E., Nelson, F. E., et al. (2018). Degrading permafrost puts Arctic infrastructure at risk by mid-century. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07557-4

Houser, P. R., Gupta, H. V., Shuttleworth, W. J., & Famiglietti, J. S. (2001). Multiobjective calibration and sensitivity of a distributed land surface water and energy balance model. *Journal of Geophysical Research Atmospheres*, *106*(D24), 33421–33433. https://doi.org/10.1029/2000JD900803

Husain, S. Z., Alavi, N., Bélair, S., Carrera, M., Zhang, S., Fortin, V., et al. (2016). The Multibudget Soil, Vegetation, and Snow (SVS) Scheme for Land Surface Parameterization: Offline Warm Season Evaluation. *Journal of Hydrometeorology*, *17*(8), 2293–2313. https://doi.org/10.1175/JHM-D-15-0228.1

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

Keshav, K., Haghnegahdar, A., Elshamy, M., Gharari, S., & Razavi, S. (2019). Aggregated gridded soil texture dataset for Mackenzie and Nelson-Churchill River Basins, Federated Research Data Repository (FRDR), Canada, https://doi.org/10.20383/101.0154

Kokelj, S. V., & Jorgenson, M. T. (2013). Advances in thermokarst research. *Permafrost and Periglacial Processes*, *24*(2), 108–119. https://doi.org/10.1002/ppp.1779

Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J., & Harrington, R. A. (1993). Grouped response units for distributed hydrologic modelling. *Journal of Water Resources Planning and Management*, *119*(3), 289–305.

Kouwen, N., Soulis, R., Seglenieks, F., Bingeman, A., & Davison, B. (1993). An Introduction to WATFLOOD and WATCLASS, 1–13.

Lamontagne-Hallé, P., McKenzie, J. M., Kurylyk, B. L., Molson, J., & Lyon, L. N. (2020). Guidelines for cold-regions groundwater numerical modeling. *Wiley Interdisciplinary Reviews: Water*, (June), 1–26. https://doi.org/10.1002/wat2.1467

Lawrence, D. M., & Slater, A. G. (2008). Incorporating organic soil into a global climate model. *Climate Dynamics*, *30*(2–3), 145–160. https://doi.org/10.1007/s00382-007-0278-1

Lawrence, D. M., Slater, A. G., Romanovsky, V. E., & Nicolsky, D. J. (2008). Sensitivity of a model projection of near-surface permafrost degradation to soil column depth and representation of soil organic matter. *Journal of Geophysical Research: Earth Surface*, *113*(2), 1–14. https://doi.org/10.1029/2007JF000883

Lawrence, D. M., Slater, A. G., & Swenson, S. C. (2012). Simulation of present-day and future permafrost and seasonally frozen ground conditions in CCSM4. *Journal of Climate*, *25*(7), 2207–2225. https://doi.org/10.1175/JCLI-D-11-00334.1

Mahfouf, J. F., Brasnett, B., & Gagnon, S. (2007). A Canadian precipitation analysis (CaPA) project: Description and preliminary results. *Atmosphere - Ocean*, *45*(1), 1–17. https://doi.org/10.3137/ao.v450101

McGuire, A. D., Lawrence, D. M., Koven, C., Clein, J. S., Burke, E., Chen, G., et al. (2018). Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), 3882–3887. https://doi.org/10.1073/pnas.1719903115

McMillan, H. K., Westerberg, I. K., & Krueger, T. (2018). Hydrological data uncertainty and its implications. *WIREs Water*, *5*(6), 1–14. https://doi.org/10.1002/wat2.1319

Mekonnen, M. A., Wheater, H. S., Ireson, A. M., Spence, C., Davison, B., & Pietroniro, A. (2014). Towards an improved land surface scheme for prairie landscapes. *Journal of Hydrology*, *511*, 105–116. https://doi.org/10.1016/j.jhydrol.2014.01.020

Melton, J. R., Verseghy, D. L., Sospedra-Alfonso, R., & Gruber, S. (2019). Improving permafrost physics in the coupled Canadian Land Surface Scheme (v.3.6.2) and Canadian Terrestrial Ecosystem Model (v.2.1) (CLASS-CTEM). *Geoscientific Model Development*, *12*(10), 4443–4467. https://doi.org/10.5194/gmd-12-4443-2019

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., et al. (2020). Polar regions. In H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M.

Tignor, E. Poloczanska, et al. (Eds.), *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (pp. 203–320). https://doi.org/10.1016/S1366-7017(01)00066-6

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161–174. https://doi.org/10.1080/00401706.1991.10484804

Nelson, F. E., Anisimov, O. A., & Shiklomanov, N. I. (2002). Climate change and hazard zonation in the circum-arctic permafrost regions. *Natural Hazards*, *26*(3), 203–225. https://doi.org/10.1023/A:1015612918401

Nicolsky, D. J., Romanovsky, V. E., Alexeev, V. A., & Lawrence, D. M. (2007). Improved modeling of permafrost dynamics in a GCM land-surface scheme. *Geophysical Research Letters*, *34*(8), 2–6. https://doi.org/10.1029/2007GL029525

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., et al. (2019). Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale. *Earth-Science Reviews*, *193*(March), 299–316. https://doi.org/10.1016/j.earscirev.2019.04.023

Pan, X., Li, Y., Yu, Q., Shi, X., Yang, D., & Roth, K. (2016). Effects of stratified active layers on high-altitude permafrost warming: A case study on the Qinghai-Tibet Plateau. *Cryosphere*, *10*(4), 1591–1603. https://doi.org/10.5194/tc-10-1591-2016

Paquin, J. P., & Sushama, L. (2015). On the Arctic near-surface permafrost and climate sensitivities to soil and snow model formulations in climate models. *Climate Dynamics*, *44*(1–2), 203–228. https://doi.org/10.1007/s00382-014-2185-6

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*, 1633–1644. https://doi.org/https://doi.org/10.5194/hess-11-1633-2007

Pietroniro, A., & Soulis, E. D. (2003). A hydrology modelling framework for the Mackenzie GEWEX programme. *Hydrological Processes*, *17*(3), 673–676. https://doi.org/10.1002/hyp.5104

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., et al. (2007). Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences*, *11*(4), 1279–1294. https://doi.org/10.5194/hess-11-1279-2007

Pomeroy, J. W., MacDonald, M. K., Dornes, P. F., & Armstrong, R. (2016). Water Budgets in Ecosystems. In E. A. Johnson & Y. E. Martin (Eds.), *Hydrological Processes of Aquatic and Terrestrial Ecosystems* (pp. 88–132). Cambridge, United Kingdom: Cambridge University Press.

Prentice, I. C., Liang, X., Medlyn, B. E., & Wang, Y. P. (2015). Reliable, robust and realistic: The three R's of next-generation land surface modelling. *Atmospheric Chemistry and Physics Discussions*, *15*(17), 5987–6005. https://doi.org/10.5194/acpd-14-24811-2014

Puy, A., Becker, W., Lo Piano, S., & Saltelli, A. (2021). A comprehensive comparison of total-order estimators for global sensitivity analysis. *International Journal for Uncertainty Quantification*. https://doi.org/10.1615/int.j.uncertaintyquantification.2021038133

Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of ‘“global”’ sensitivity in Earth and Environmental systems models. *Water Resources Research*, *51*, 3070–3092. https://doi.org/10.1002/2014WR016527

Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, *52*, 423–439. https://doi.org/10.1002/2015WR017559.A

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

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., et al. (2021). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling and Software*, *137*(December 2020). https://doi.org/10.1016/j.envsoft.2020.104954

Riseborough, D., Shiklomanov, N., Etzelmuller, B., Gruber, S., & Marchenko, S. (2008). Recent Advances in Permafrost Modelling. *Permafrost and Periglacial Processes*, *19*, 137–156. https://doi.org/10.1002/ppp.615

Rodell, M., Houser, P. R., Berg, A. A., & Famiglietti, J. S. (2005). Evaluation of 10 methods for initializing a land surface model. *Journal of Hydrometeorology*, *6*(2), 146–155. https://doi.org/10.1175/JHM414.1

Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, *25*(12), 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

Sapriza-Azuri, G., Gamazo, P., Razavi, S., & Wheater, H. S. (2018). On the appropriate definition of soil profile configuration and initial conditions for land surface-hydrology models in cold regions. *Hydrology and Earth System Sciences*, *22*(6), 3295–3309. https://doi.org/10.5194/hess-22-3295-2018

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., et al. (2015). Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), 171–179. https://doi.org/10.1038/nature14338

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., et al. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*. https://doi.org/10.1126/science.275.5299.502

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., & Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *9*(1), 65–88. https://doi.org/10.1002/2016MS000686

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, *19*(13), 3088–3111. https://doi.org/10.1175/JCLI3790.1

Sheikholeslami, R., Razavi, S., & Haghnegahdar, A. (2019). What should we do when a model crashes ? Recommendations for global sensitivity analysis of Earth and environmental systems models. *Geoscientific Model Development*, *12*, 4275–4296. https://doi.org/https://doi.org/10.5194/gmd-12-4275-2019

Sheikholeslami, R., Gharari, S., Papalexiou, S. M., & Clark, M. P. (2021). VISCOUS: A Variance-Based Sensitivity Analysis Using Copulas for Efficient Identification of Dominant Hydrological Processes. *Water Resources Research*, *57*(7), 1–24. https://doi.org/10.1029/2020wr028435

Slater, A. G., & Lawrence, D. M. (2013). Diagnosing present and future permafrost from climate models. *Journal of Climate*, *26*(15), 5608–5623. https://doi.org/10.1175/JCLI-D-12-00341.1

Smith, S., Burgess, M. M., Riseborough, D., Coultish, T., & Chartrand, J. (2004). *Digital Summary Database of Permafrost and Thermal Conditions – Norman Wells Pipeline Study Sites*. Geological Survey of Canada, Open File 4635, Ottawa, Canada.

Smith, S., Chartrand, J., Nguyen, T.-N., Riseborough, D. W., Ednie, M., & Ye, S. (2009). *Geotechnical database and descriptions of permafrost monitoring sites established 2006-07 in the central and southern Mackenzie Corridor*. Geological Survey of Canada, Open File 6041, Ottawa, Canada.

Smith, S., Throop, J., Ednie, M., Chartrand, J., Riseborough, D. W., & Nixon, F. M. (2010). *Report on 2009 Field Activities and Ground Thermal Data Collection in the Mackenzie Corridor Completed Under N.W.T. Science Licence # 14582*. Geological Survey of Canada, Open File 6695, Ottawa, Canada.

Smith, S., Chartrand, J., Duchesne, C., Ednie, M., Smith, S. L., Duchesne, C., & Riseborough, D. W. (2016). Report on 2013 field activities and collection of ground thermal and active layer data in the Mackenzie Corridor. *Geological Survey of Canada Open File 8125*, *7659*, 1–105. https://doi.org/10.4095/295596

Smith, S., Chartrand, J., Duchesne, C., & Ednie, M. (2016). *Report on 2015 field activities and collection of ground thermal and active layer data in the Mackenzie Corridor, Northwest Territories*. Geological Survey of Canada, Open File 8125, Ottawa, Canada. https://doi.org/10.4095/299296

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1–3), 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Soulis, E. D., Snelgrove, K. R., Kouwen, N., Seglenieks, F., & Verseghy, D. L. (2000). Towards closing the vertical water balance in Canadian atmospheric models: Coupling of the land surface scheme class with the distributed hydrological model watflood. *Atmosphere - Ocean*, *38*(1), 251–269. https://doi.org/10.1080/07055900.2000.9649648

Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in peel inlet-II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, *14*(1), 43–49. https://doi.org/10.1016/0043-1354(80)90040-8

U.S. Department of Agriculture (1951). Soil survey manual, USDA Handbook 18. Washington, D.C.: Government Printing Office

Verseghy, D. (1991). Canadian Land Surface Scheme for GCMS I. Soil model. *International Journal of Climatology*, *11*, 111–133. https://doi.org/10.1002/joc.3370110202

Verseghy, D. (2000). The Canadian land surface scheme (CLASS): Its history and future. *Atmosphere - Ocean*, *38*(1), 1–13. https://doi.org/10.1080/07055900.2000.9649637

Verseghy, D. (2012). *CLASS – The Canadian land surface scheme (version 3.6) - technical documentation. Internal report, Climate Research Division, Science and Technology Branch, Environmental Canada*.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, *17*(2), 455–476. https://doi.org/10.1002/hyp.1135

Walvoord, M. A., & Kurylyk, B. L. (2016). Hydrologic Impacts of Thawing Permafrost—A Review. *Vadose Zone Journal*, *15*(6), 1–20. https://doi.org/10.2136/vzj2016.01.0010

Weedon, G., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., ÖSterle, H., et al. (2011). Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology*, *12*(5), 823–848. https://doi.org/10.1175/2011JHM1369.1

Weedon, G., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, *50*, 7505–7514. https://doi.org/10.1002/2014WR015638

Wheater, H., Pomeroy, J., Pietroniro, A., Davison, B., Elshamy, M., Yassin, F., et al. (2021). Advances in modelling large river basins in cold regions with Modélisation Environmentale Communautaire-Surface and Hydrology (MESH), the Canadian hydrological land surface scheme. Authorea Preprints, July 27, 2021, under review, https://doi.org/10.22541/au.162738548.88997933/v1

Wong, J. S., Razavi, S., Bonsal, B. R., Wheater, H. S., & Asong, Z. E. (2017). Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. *Hydrology and Earth System Sciences*, *21*(4), 2163–2185. https://doi.org/10.5194/hess-21-2163-2017

Wu, Y., Verseghy, D. L., & Melton, J. R. (2016). Integrating peatlands into the coupled Canadian Land Surface Scheme (CLASS) v3.6 and the Canadian Terrestrial Ecosystem Model (CTEM) v2.0. *Geoscientific Model Development*, *9*(8), 2639–2663. https://doi.org/10.5194/gmd-9-2639-2016

Yassin, F., Razavi, S., Wheater, H., Sapriza-Azuri, G., Davison, B., & Pietroniro, A. (2017). Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach. *Hydrological Processes*, *31*(19), 3320–3333. https://doi.org/10.1002/hyp.11267

Yokohata, T., Saito, K., Takata, K., Nitta, T., Satoh, Y., Hajima, T., et al. (2020). Model improvement and future projection of permafrost processes in a global land surface model. *Progress in Earth and Planetary Science*, *7*(1). https://doi.org/10.1186/s40645-020-00380-

w

Zhang, Yinsuo, Carey, S., & Quinton, W. (2008). Evaluation of the algorithms and parameterizations for ground thawing and freezing simulation in permafrost regions. *Journal of Geophysical Research Atmospheres*, *113*(17), 1–17. https://doi.org/10.1029/2007JD009343

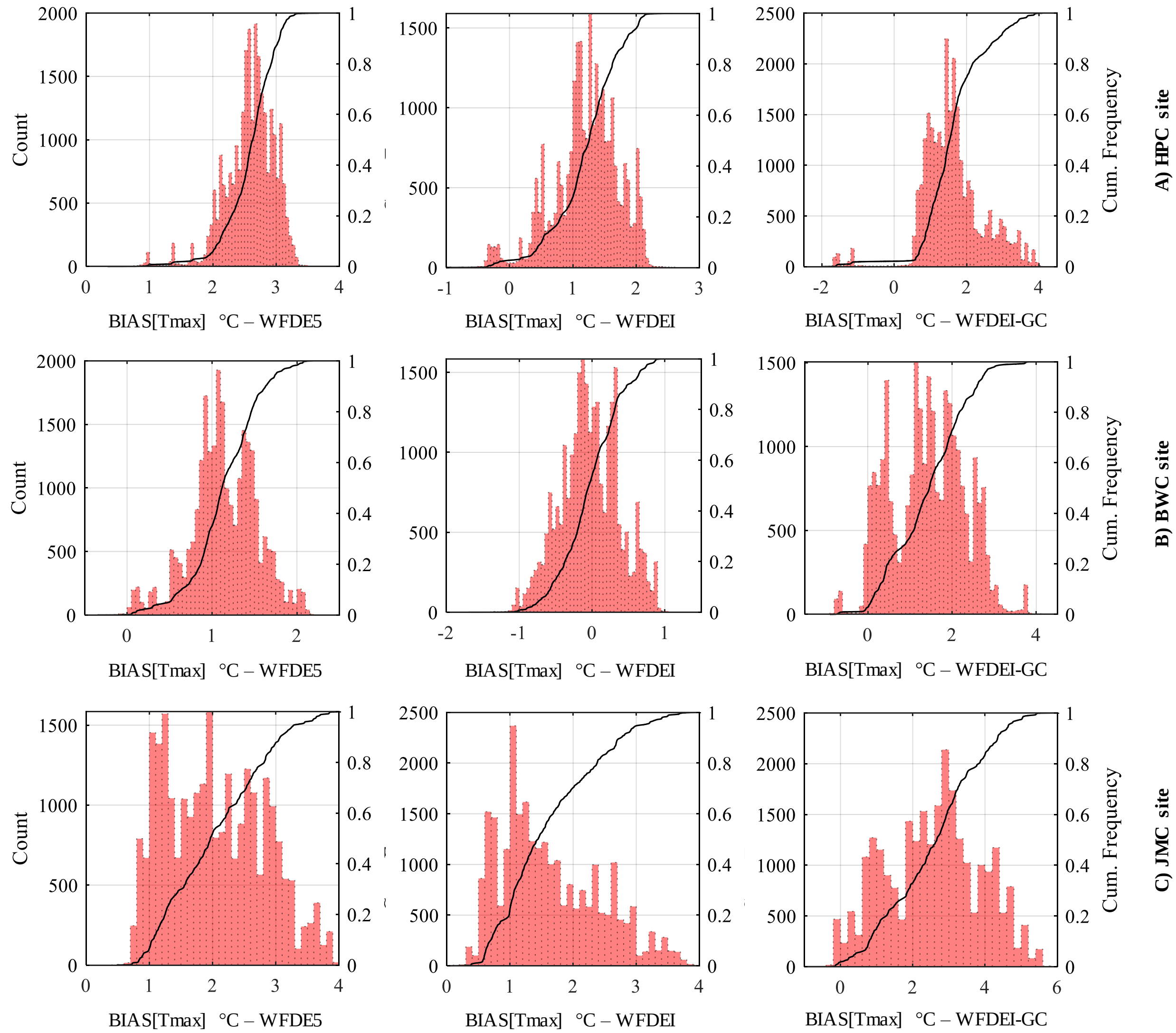
Zhang, Yu, Chen, W., & Cihlar, J. (2003). A process-based model for quantifying the impact of climate change on permafrost thermal regimes. *Journal of Geophysical Research D: Atmospheres*, *108*(22). https://doi.org/10.1029/2002jd003354

Zhang, Yu, Chen, W., & Riseborough, D. W. (2008). Transient projections of permafrost distribution in Canada during the 21st century under scenarios of climate change. *Global and Planetary Change*, *60*(3–4), 443–456. https://doi.org/10.1016/j.gloplacha.2007.05.003

## Appendix

**Table A1.** Soil profile layering scheme for the two sites, adopted form (Elshamy et al., 2020).

| Layer | Thickness | Layer | Thickness |
|---|---|---|---|
| 1 | 0.1 | 14 | 1.48 |
| 2 | 0.1 | 15 | 1.78 |
| 3 | 0.11 | 16 | 2.11 |
| 4 | 0.13 | 17 | 2.48 |
| 5 | 0.16 | 18 | 2.88 |
| 6 | 0.21 | 19 | 3.33 |
| 7 | 0.28 | 20 | 3.81 |
| 8 | 0.37 | 21 | 4.34 |
| 9 | 0.48 | 22 | 4.9 |
| 10 | 0.63 | 23 | 5.51 |
| 11 | 0.8 | 24 | 6.17 |
| 12 | 0.99 | 25 | 6.87 |
| 13 | 1.22 | | |




1290

1291 **Fig. A1**. Same as **Fig. 6**, but showing the BIAS in maximum annual temperature envelope (Tmax).

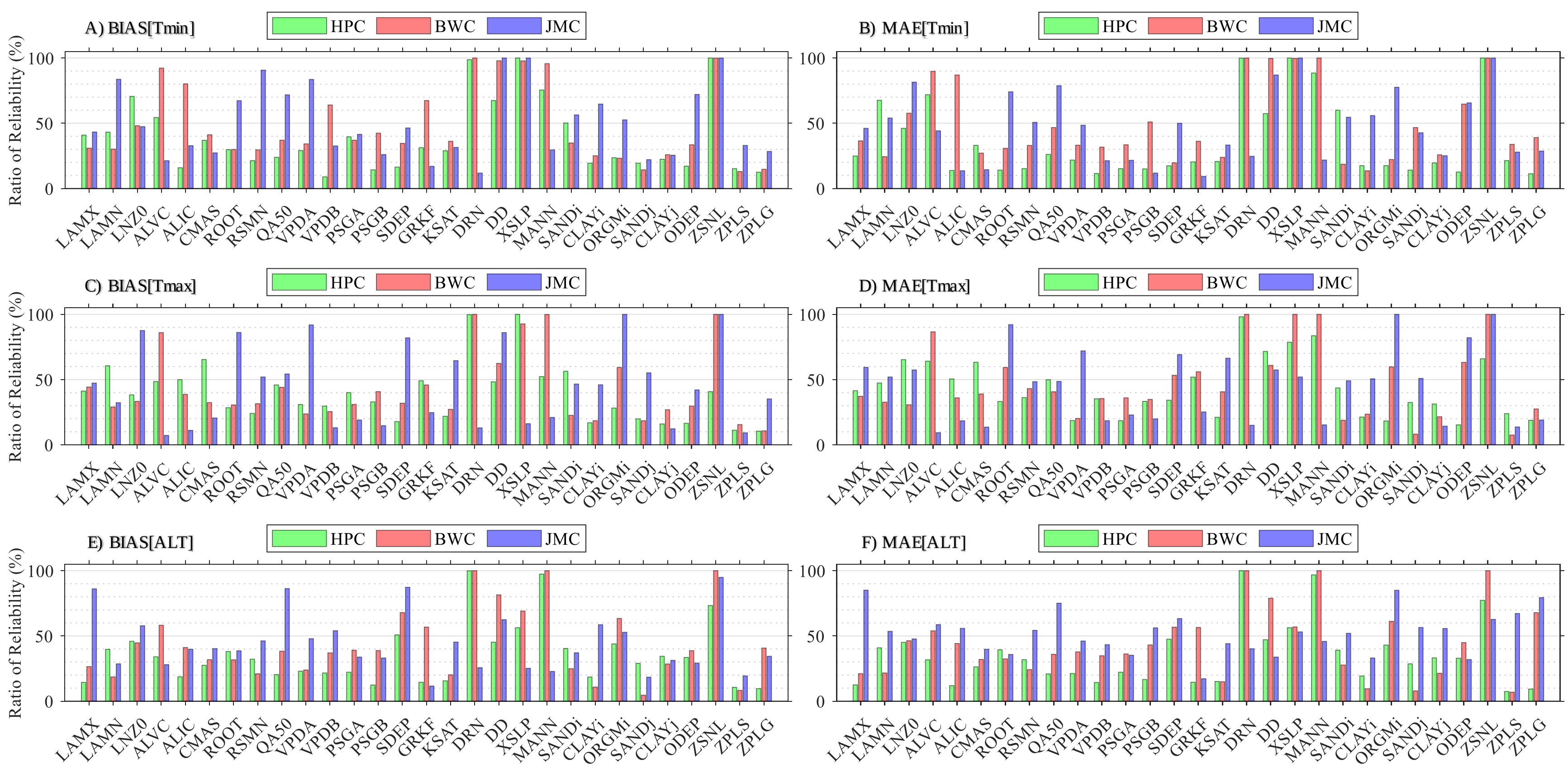



**Fig. A2.** Ratio of reliability of IVARS$_{50}$ for each parameter in all experiments for A) BIAS[Tmin], B) MAE[Tmin], C) BIAS[Tmax], D) MAE[Tmax], E) BIAS[ALT], and F) MAE[ALT].

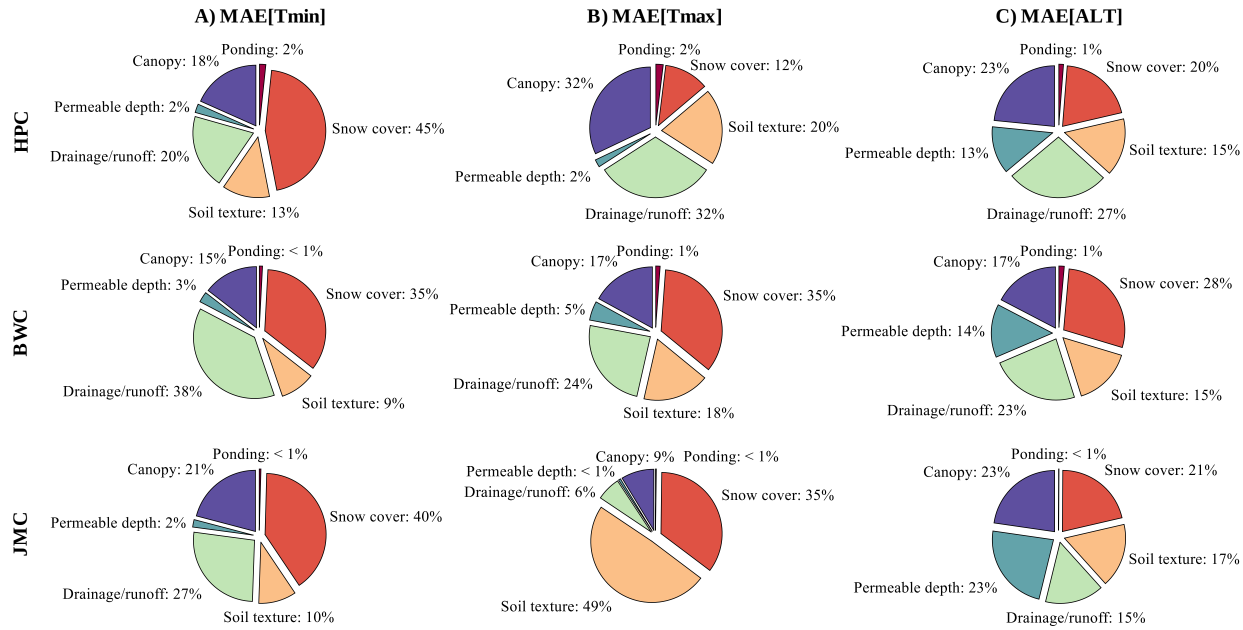



**Fig. A3.** Total ratio of sensitivity (based on $IVARS_{50}$) for each group of parameters (as indicated in **Table 2**) for the MAE across all experiments.

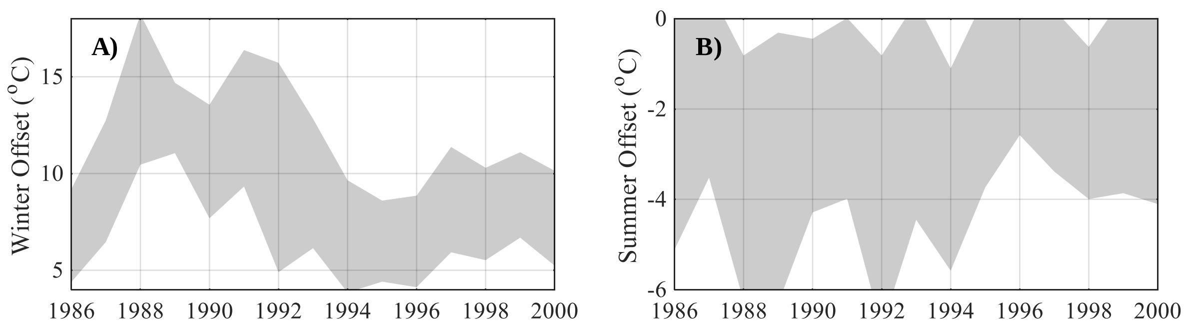



**Fig. A4.** Temporal evolution and the associated range of uncertainty for A) Winter offset by accumulated snow, and B) Summer offset by the canopy's shading; gray shading denotes the range of uncertainty.