The Ability of Hydrologic-Land Surface Models to Concurrently Simulate Permafrost and Hydrology

Mohamed S. Abdelhamed[1,2,3], Mohamed E. Elshamy[1,4], Saman Razavi[1,2,5,6], Howard S. Wheater[1,4,5,7]

[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]Institute for Water Futures, Mathematical Science Institute, Australian National University, Canberra, Australia.

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

*Corresponding author: Mohamed.abdelhamed@usask.ca

Running Title: H-LSM challenges to model permafrost and hydrology

Keywords: Permafrost, Hydrologic-land surface models, Active layer thickness, Temperature envelope, Streamflow, Liard River Basin.

Abstract

Hydrologic-land surface models (H-LSMs) provide physically-based understanding and predictions of the current and future states of the world's vast high-latitude permafrost regions. Two major challenges, however, hamper their parametrization and validation when concurrently representing hydrology and permafrost. One is the high computational complexity, exacerbated by the need to include a deep soil profile to adequately capture the freeze/thaw cycles and heat storage. The other is that soil-temperature data are severely limited, and traditional model validation, based on streamflow, can show the right fit to these data for the wrong reasons. There are few observational sites for such vast, heterogeneous regions, and remote sensing provides only limited support. In light of these challenges, we develop 16 parametrizations of a Canadian H-LSM, MESH, for the sub-arctic Liard River Basin and validate them using three data sources: streamflows at multiple gauges, soil temperature profiles from few available boreholes, and multiple permafrost maps. The different parametrizations favor different sources of data and it is challenging to configure a model faithful to all three data sources, which are at times inconsistent with each other. Overall, the results show that: (1) surface insulation through snow-cover primarily regulates permafrost dynamics after model initialization effects decay, relatively long time and (2) different parametrizations yield different partitioning patterns of solid-vs-liquid soil-water and produce different low-flow but similar high-flow regimes. We conclude that, given data scarcity, an ensemble of model parametrizations is essential to provide a reliable picture of the current states and future spatio-temporal co-evolution of permafrost and hydrology.

1   Introduction

Expanding knowledge of Earth system governing processes, the revolution in computing, and the pressing need to investigate/predict various interrelated facets of Earth systems have collectively led to the development of more complex numerical Earth system models. Hydrologic-Land Surface Models (H-LSMs) are one such advancements that have significantly improved atmospheric predictions and hydrological flow simulations. H-LSMs have evolved from first generation, simple process representations, such as the bucket model for moisture storage with a lumped treatment of vegetation and snow (Manabe, 1969), to third generation, state-of-the-art representations that account for feedback loops, vegetation dynamics, carbon pools, and surface/subsurface flow routing (Pitman, 2003; Prentice et al., 2015; Sellers et al., 1997). Yet, model representations of natural Earth system(s) remain limited for several reasons, including the limited scalability of small-scale observations of governing processes, the unavailability of validation data, process complexity, and the models' large degrees of freedom. In cold regions, configuring an H-LSM to simulate hydrological responses also presents difficulties associated with the limited understanding of water and energy process interactions. Upscaling process understanding from local-scale to larger domains and evaluating model performance adds further complexity to model development and diagnosis (Pomeroy et al., 2007). Perennially frozen ground, or *permafrost*, is an example of a crucial feature of cold regions for which understanding of thermal-hydrological behaviour is limited; yet, at the same time, this feature dominates the hydrological functionality over large areas.

Permafrost is ground material that sustains cryotic conditions (*i.e.,* temperature ≤ 0 °C) for no less than two successive years. Around a quarter of the Northern Hemisphere's land is underlain by permafrost, highlighting its substantial

influence on hydrology, climatology, and ecology (Walvoord and Kurylyk, 2016; Woo et al., 1992). Permafrost is overlain by a dynamic soil layer prone to seasonal thawing/freezing cycles called the *active layer* (van Everdingen,

55   1998). Permafrost thaw/degradation can happen in a heterogeneous manner, both vertically (upward, downward, or both) and spatially (Farquharson et al., 2019; Harris et al., 2009; Morse et al., 2016; Romanovsky et al., 2007; Smith et al., 2012; Zhao et al., 2010), which in turn affects the partitioning of water/heat stores above and below ground surface, hydrologic connectivity of landscape features, and streamflow seasonality. Further, there is strong dependency between energy and water fluxes/stores within permafrost (Riseborough et al., 2008; Walvoord and

60   Kurylyk, 2016), which necessitates appropriate numerical model formulations to simulate their coevolution (in addition to carbon). In this regard, Earth System Models (ESMs), particularly the land-surface component (H-LSMs), have been employed (and advanced over time) for accurate, robust, and reliable simulation of thermo-hydrological processes in permafrost regions (Abdelhamed et al., 2022a; Alexeev et al., 2007; Elshamy et al., 2020; Ji et al., 2022; Lawrence and Slater, 2008) and future projections under climate change scenarios (Burke et al., 2020; Lawrence and

65   Slater, 2005; Zhao et al., 2022).

Permafrost dynamics are simulated in H-LSMs through the coupled evolution of water and heat (and carbon for some H-LSMs) across atmosphere-vegetation-soil interfaces, which can be linked to atmospheric models through online (two-way) coupling to represent critical feedback loops related to permafrost state changes (*e.g.,* the carbon pool in permafrost). Recently, several improvements have been introduced to enhance the representation of permafrost

70   dynamics in H-LSMs. These include deeper soil columns to capture the freeze/thaw cycle and resolve the heat storage of underlying ground (Alexeev et al., 2007; Hermoso de Mendoza et al., 2020), incorporating the insulation effect of snow (using a multi-layer scheme) (Chadburn et al., 2015) and soil organic matter explicitly (Lawrence and Slater, 2008; Park et al., 2013), carbon pool processes and associated vegetation dynamics (Melton et al., 2019), and representation of lateral taliks and micro-topographic processes (Aas et al., 2019; Devoie et al., 2019). However, these

75   improvements are not incorporated into all current models due to computational and/or numerical limitations and/or the absence of upscaling methodologies.

The paucity of *in situ* permafrost observations restrains model development, diagnosis, and assessment, amplified by the spatial heterogeneity of permafrost and the non-uniformity of the thaw process (pattern and rate), as highlighted by Lamontagne-Hallé et al. (2020), Pastick et al. (2014), and Walvoord and Kurylyk (2016). Moreover, current remote

80   sensing technologies remain incapable of (directly) detecting permafrost characteristics (*e.g.,* soil temperature, ice content, and thaw/freeze depth) (Westermann et al., 2015). However some indirect/non-comprehensive permafrost indicators are currently detectable and could be used to delineate the presence of permafrost (Hachem et al., 2009; Muskett and Romanovsky, 2011), such as land surface temperature (*e.g.,* MODIS LST: Hulley and Hook, 2017), total water storage: (TWS: Tapley et al., 2004) and surface topography (*e.g.,* SRTM: Farr and Kobrick, 2000). Building on

85   the correlation between permafrost occurrence/status and those indicators, several mapping techniques for permafrost have been introduced. For instance, Gruber (2012), Chadburn et al. (2017), and Aalto et al. (2018) correlated the spatial occurrence of permafrost with the mean annual air temperature (MAAT), elevation, and surficial condition. Another practical approach can be found in thermal models (*e.g.,* TTOP: Smith and Riseborough, 1996) that

incorporate a transfer function between the MAAT and mean annual ground temperature (MAGT) to locate the
90   freeze/thaw front; the work of Obu et al. (2019) is one example that also combines land cover and precipitation as
inputs. Machine learning approaches can be used, but these require a large suite of input data (*e.g.,* soil texture, climate,
vegetation cover) to compile permafrost variables, such as the gridded product of Ran et al. (2022).

The significant spatial heterogeneity of permafrost, as well as the scarcity of permafrost data over space/time, has
hindered the development of a unified upscaling approach for mapping permafrost over large domains. As a result, it
95   is challenging for modellers to validate any H-LSM against gridded permafrost products (over large domains) and
localized ground observations (either thermal cables or thaw tubes). Furthermore, modelling the evolution of
permafrost has historically disregarded information about hydrologic fluxes/states, *e.g.,* streamflow,
evapotranspiration, and basin storage. By incorporating these aspects, modellers are expected to gain more insights
into the spatio-temporal evolution of water-energy states and fluxes and identify synergies or trade-offs that could
100 inform future efforts for model parameterization and data collection, and eventually minimize the uncertainty in model
projections.

Here, we aim to assess the capability of an H-LSM to simultaneously replicate observed permafrost dynamics at the
local/point scale (*vs* observational sites) and basin scale (*vs* gridded permafrost products), in addition to the streamflow
dynamics that integrate various dominant underlying hydrological processes. The MESH H-LSM (Pietroniro et al.,
105 2007; Wheater et al., 2022) is used; model experiments are configured based on previous studies with the same model.
The assessment is performed on the Canadian Liard River Basin (LRB), which is dominated by sporadic-discontinuous
permafrost and has a significant streamflow contribution to the sub-continental Mackenzie River Basin (MRB).
Several gridded permafrost products with different levels of physics, complexity, and spatial/temporal coverage are
used as additional indicators of simulation quality. The remainder of the article is organized as follows: **Section 2**
110 presents the methods, case study, data sources, and model implementation, **Section 3** reports the results of all
experiments, and the article closes with a summary and conclusions in **Section 4**.

2   Models, datasets, and methods

2.1   Study area

The Liard River Basin (LRB) is a large basin (area 275,000 km^2, elevation range 140–2,700 m; **Fig. 1**) in subarctic
115 Canada underlain by discontinuous and sporadic permafrost that is sensitive to the effects of climatic warming
(DeBeer et al., 2016; Hayashi et al., 2004; Woo, 2012). The LRB is a mostly natural river system with insignificant
surface water storage (*i.e.,* no large lakes) or water management (**Fig. 1**). The basin has a high runoff coefficient
(around 0.55) and contributes more than one quarter of the annual flow to the 1.8 million km^2 Mackenzie River Basin
(MRB) (Woo, 2012). The LRB has received attention from researchers in various hydrological and Earth system-
120 related studies, including peatland hydrology in discontinuous permafrost (Connon et al., 2014; Hayashi et al., 2004;
Quinton et al., 2009), the role of the mountain snow accumulation/melt on simulated discharge (Woo and Thorne,
2006), the improvement in streamflow and snow water equivalent simulation from assimilating terrestrial water
storage data (Bahrami et al., 2020), the hydrologic implications of climate changes under general circulation model

(GCM) and scenario uncertainties (Shrestha et al., 2019; Thorne, 2011), and the surface water balance closure for the
125 MRB and its sub-basins (Louie et al., 2002; MacKay et al., 2003).

The Liard River headwaters lie in the Yukon, British Colombia, Alberta, and the Northwest Territories. Short, cool summers and long, cold winters describe the climate of the basin (Jacques and Sauchyn, 2009), which has a mean annual temperature of approximately –2.9 °C and mean annual total precipitation of around 490 mm (40% as snowfall). The LRB can be divided into five distinct sub-basins (*i.e.,* Upper Liard, Central Liard, Lower Liard, Fort
130 Nelson, and Petitot), as shown in **Fig. 1F**. Boreal forest (with alpine tundra) dominates the Upper and Central Liard sub-basins (**Fig. 1C**), while sub-arctic forest and alpine tundra cover the Lower Liard sub-basin. Boreal forest and wetlands dominate the Petitot and Fort Nelson sub-basins, and small agricultural areas exist near Fort Nelson (Burn et al., 2004b).

Fig. 1D shows the permafrost extent and categories for the LRB based on the Canadian Permafrost Map
135 (Hegginbottom et al., 1995; Brown et al., 1998) in addition to the available permafrost monitoring sites (see **Section2.4**). About 57% of the basin is underlain by sporadic permafrost (the southern part of the basin), while discontinuous permafrost underlies the northern 33% of the basin. Hydrologically, the basin is characterized by low flows in winter months and a rising hydrograph starting in late April and May with a peak occurring in June, followed by a gradual decline to winter low-flow conditions in December (Burn et al., 2004a). The annual minimum flow occurs
140 in late March or early April due to snowmelt and the mean annual flow at the outlet from 2000 to 2016 was 2600 m³/s. River flow is supplied mostly by snowmelt and rainfall, with a slight contribution from small glaciers, and groundwater is the primary source of low flows in winter (Woo, 2012). The Upper and Central Liard catchments contribute around one-half of the flow at the outlet from 38% of the basin's contributing area.

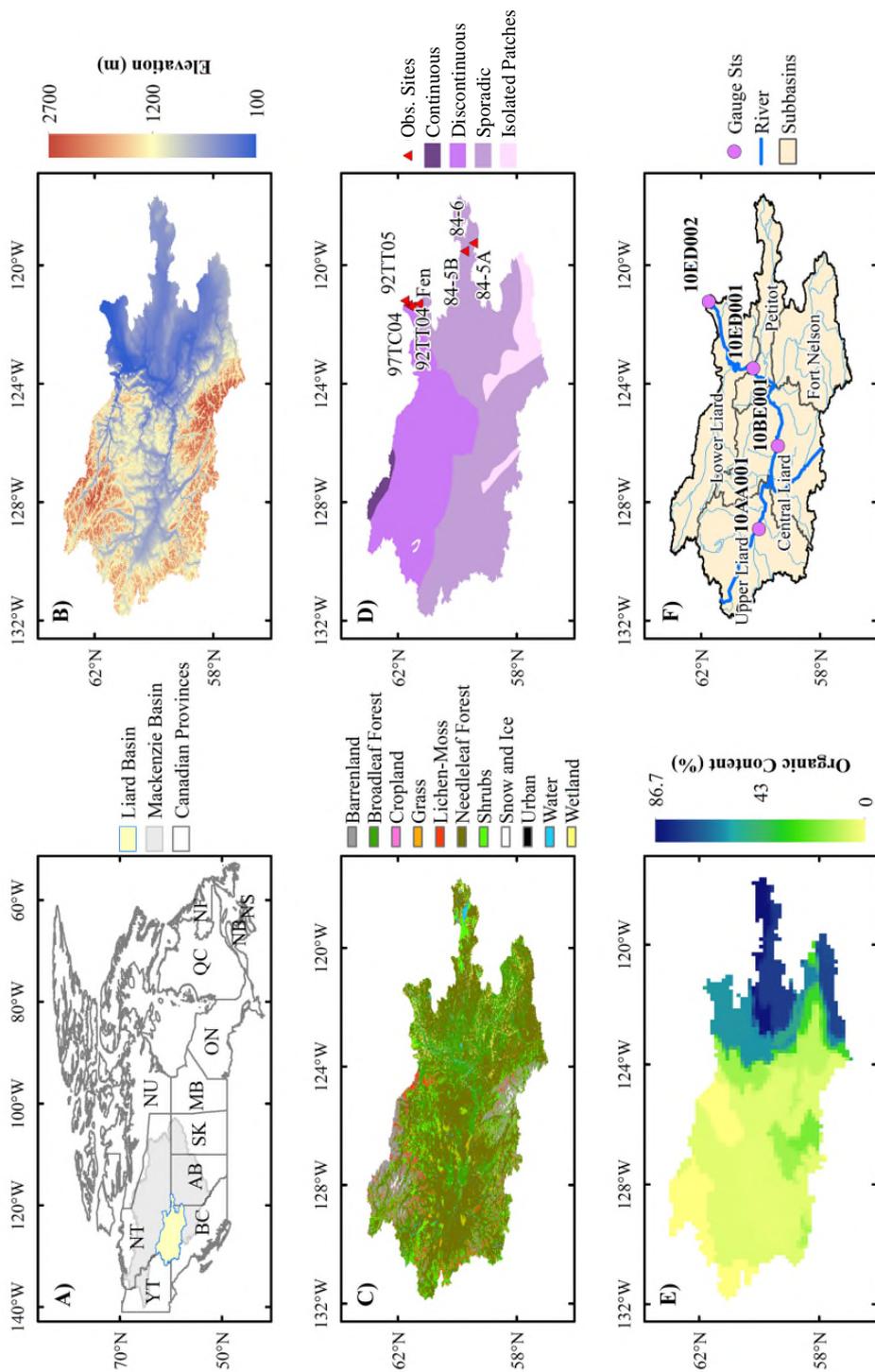



Fig. 1. (A) Location map of the Liard River Basin (LRB) and Mackenzie River Basin (MRB) in Canada, (B) LRB topography, (C) LRB land cover, (D) LRB permafrost classification according to Brown et al. (1998) and the utilized observational sites, (E) LRB organic content, and (F) LRB subbasins and gauge stations.

145

2.2 Model description

The **M**odélisation **E**nvironmenntale communautaire – **S**urface et **H**ydrology (MESH) (Pietroniro et al., 2007; Wheater et al., 2022) is the model utilized in this study. MESH has a semi-/fully distributed grid-based modelling architecture encompassing a land surface component that quantifies vertical energy and water fluxes in cold regions (CLASS: Verseghy, 1991, 2000) or (SVS: Husain et al., 2016), an algorithm for lateral movement of surface and subsurface flow (WATROF: Soulis et al., 2000) and/or (PDMROF: Mekonnen et al., 2014), and a grid-to-grid hydrologic river routing scheme (WATFLOOD: Kouwen et al., 1993a). The spatial heterogeneity within each grid is conserved by using the Grouped Response Unit (GRU) concept (Kouwen et al., 1993b). Fluxes/states are calculated by default at a half-hour time step at the tile level (GRUs mapped onto grids) and then aggregated to grids. Seven meteorological forcing variables drive the land surface component: precipitation, air temperature, shortwave radiation, longwave radiation, specific humidity, air pressure, and wind speed.

Organic matter can be configured in MESH following two approaches: (1) fully organic soil (FOS hereafter) using three predefined organic peat types (*i.e.,* fibric, hemic, and sapric) based on the work of Letts et al. (2000), and (2) mineral soil with organic content (MSO hereafter), allowing the coexistence of mineral and organic matter. Thermal properties are the same for all peat sub-types, noting that peat has higher heat capacity (2.5×10^6 Jm^{-3} K^{-1}), and lower thermal conductivity (0.25 Wm^{-1} K^{-1}) than mineral soils. Further, peat has higher porosity (0.93, 0.88, 0.83 for the three sub-types, respectively, compared to 0.49 for clay and 0.36 for sand), higher retention capacity (0.275, 0.62, and 0.705, respectively), and higher residual water content (0.04, 0.15, and 0.22, respectively) compared to mineral soils.

MESH has been under continuous development by Environment and Climate Change Canada (ECCC) and several Canadian universities as an investigative and predictive model for various application scales and regions. Developments include simulation of blowing snow transport and sublimation within grids via the Prairie Blowing Snow Model (PBSM, MacDonald et al., 2009), and improved representation of base-flow routing (Luo et al., 2012) and reservoir operations (Yassin et al., 2019). MESH has been widely used in Canada at various scales of applications: small-scale watersheds (< 100 km^2) such as Pelly's lake (Berry et al., 2017), Bosworth Creek (Sapriza-Azuri et al., 2018), and Jean Marie Creek (Abdelhamed et al., 2022a); medium-scale watersheds (>100 to < 1000 km^2) such as Bright Water Creek (Budhathoki et al., 2020) and White Gull Creek (Davison et al., 2016; Razavi and Gupta, 2016); and large-scale river basins (> 1000 km^2), such as the Liard River Basin (0.275 million km^2; Bahrami et al., 2020), the Canadian portion of the Yukon River Basin (0.288 million km^2; Elshamy et al., 2022), the Saskatchewan River Basin (0.40 million km^2; Yassin et al., 2017, 2019), the Great lakes (1.0 million km^2; Deacu et al., 2012; Haghnegahdar et al., 2014), and the Mackenzie River Basin (1.78 million km^2; Elshamy et al., in preparation). Interested readers are referred to Wheater et al. (2022) for a comprehensive list of developments and applications.

For the current study, CLASS version 3.6 is used as the underlying LSM and WATROF as the underlying runoff generation algorithm. CLASS solves the coupled water and energy balances for a user-specified soil column, noting that the default system has three layers with thicknesses of 0.1, 0.25, and 3.75 m, respectively, generalized across the modelled basin. A deeper soil column with a power-function layer discretization is used in the current study (Supplement **Table S1**). Descriptive soil parameters (defining the thermal and hydraulic regimes) are tied to soil

texture percentages of sand, clay, and organic matter. For each soil layer, temperature and moisture content (liquid and frozen) evolve during each time step based on the solution of coupled water-energy balance equations. Boundary

185   conditions are the surface exchanges of heat and moisture with the atmosphere, no lateral movement of heat or moisture between adjacent grids, and no heat flux at the bottom of the soil column. Spatial heterogeneity is incorporated by subdividing each grid into tiles based on land cover. Further details are provided in the CLASS manual (Verseghy, 2012).

2.3   Climate forcing

190   The meteorological variables needed to force MESH were obtained from version 2 of the W5E5 dataset (Lange et al., 2021), a merged dataset that combines the water and global change forcing data (WFD) with the ERA5 reanalysis (WFDE5: Cucchi et al., 2020) data over land and the ERA5 reanalysis (Hersbach et al., 2020) over the ocean, in addition to precipitation data from version 2.3 of the Global Precipitation Climatology Project (GPCP). W5E5 V2 covers the entire globe at 0.5° (~55 km) spatial and daily temporal resolution and is available from 1979 to 2019.

195   W5E5 V2 was selected for two reasons: 1) it overlaps with data for the selected permafrost observational sites (*i.e.,* Petitot River sites have been functional since 1985 (see **Fig. 3** and **Table 1**) and 2) it was used for bias correcting climate model data for climate change impact assessments carried out in phase 3b of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3b: Lange, 2019), from which we utilized the bias-corrected and downscaled future projection for 10 CMIP6 models to assess the concurrent temporal/spatial evolution of permafrost and hydrology for

200   the same study area in a forthcoming paper (Abdelhamed et al., in preparation). W5E5 V2 was remapped to the same grid as the LRB setup (0.125°×0.125°) using bilinear interpolation.

As noted above, MESH is driven by seven meteorological variables at a sub-daily time step, necessitating temporal disaggregation of W5E5 V2 climate data from the original daily time-step into a 3-hour time step, following Melton and Arora (2016). Linear interpolation is utilized for surface pressure, specific humidity, and wind speed. For

205   radiation, longwave radiation is uniformly distributed throughout the day, while shortwave radiation is distributed diurnally over the day based on the grid's latitude and the day of the year. Similar to shortwave radiation, disaggregated air temperature preserves the diurnal cycle. Lastly, total daily precipitation is stochastically disaggregated into a 3-hour time step. However, because this approach does not preserve the diurnal cycle of specific humidity, it can lead to unrealistic sub-daily values, resulting in more frequent oversaturation (*i.e.,* relative humidity exceeding 100%) and

210   persistent negative evapotranspiration (condensation) events. To rectify this problem and knowing that specific humidity, relative humidity, and air temperature are related, we assumed a constant relative humidity throughout the day (*i.e.,* at the daily value provided by W5E5 V2) and used its value along with the disaggregated sub-daily air pressure and temperature to modify the specific humidity and synthetically impose the diurnal cycle (Supplement **Section S2**).

2.4 Permafrost variables, ground observations, and gridded datasets

215 We focused on three permafrost variables to evaluate the simulated permafrost using ground observations (**Fig. 3**), namely the two annual temperature envelopes (T_{min} and T_{max}) and the active layer thickness (ALT), as shown in **Fig. 2**. These variables can be extracted from the continuously simulated soil temperature for each tile in each grid in MESH. Temperature envelopes provide a comprehensive measure of simulation quality and are used to derive all

220 other permafrost variables. ALT is deemed the most important factor used by both hydrologists and meteorologists to describe permafrost evolution over time in terms of aggradation/degradation (Farquharson et al., 2019; Park et al., 2013; Zhang et al., 2005).

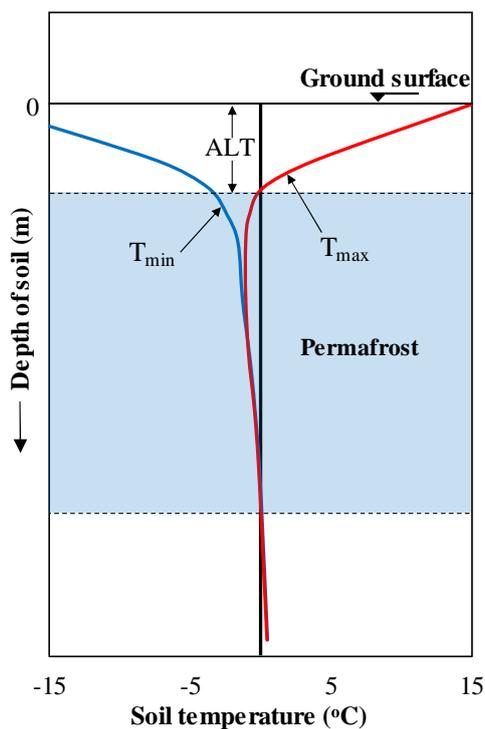



Fig. 2. Schematic of the soil column showing variables used to represent permafrost dynamics, modified after Abdelhamed et al.
225 (2022a).

However, these variables are limited with respect to describing the vertical distribution of permafrost and are of secondary importance when it comes to assessing the spatial distribution of permafrost in grid-based models. In this regard, two interrelated variables are widely used: 1) permafrost extent (PE), defined as the proportion of an area (grid pixel) underlain by permafrost (also called permafrost probability), and 2) permafrost area (PA), defined as PE

230 multiplied by grid pixel area. Another subjective indicator, known as the permafrost region (PR), can be used, defined as the area of a grid pixel with a PE larger than a predefined probability threshold (*e.g.*, ≥0.02 as in Gruber (2012) and ≥0.50 as in Burke et al. (2020)). In the current study, we opted for PE and PA, in addition to ALT, to assess the spatio-

temporal evolution of permafrost, noting these three variables are averaged based on the fractional coverage area of GRUs in each grid. Furthermore, we assumed permafrost would exist underneath any tile if at least two soil layers

235 (out of the 25-layer soil column) are cryotic through each simulation year; the one-year cycle was adopted to facilitate the evaluation of model simulations against annual temperature envelopes and annual active layer thickness observations. We opted for two soil layers, instead of one, to circumvent uncertainties of having 1) warm permafrost (*i.e.,* negative temperatures but very close to zero) in a single thin layer, which could be a numerical error, and 2) permafrost only existing in the very thin topsoil layer (between 0.1 and 0.2 m). Alternatively, we can limit our search

240 to a specific depth (*e.g.,* 2 m as in Burke et al. (2020)), but this approach is subjective and does not account for deep permafrost, as indicated by the observed temperature envelopes in the LRB (**Fig. 3**).

Permafrost ground observations in Canada are limited to experimental sites and/or boreholes made during (and after) the construction of large-scale infrastructure for maintenance and monitoring purposes. In this study, seven sites within the LRB are available (**Fig. 3** and **Table 1**); six provide a continuous time series of the annual soil temperature

245 envelopes (T_{min} and T_{max}), while the Manners Creek site (two boreholes) offers thaw-tube-based annual ALT. Notably, the selected sites vary in surficial vegetation cover, permafrost zonation, and temporal coverage. Regarding vegetation cover, sites with complex land cover (*e.g.,* Wrigley Highway) will be compared to GRUs with the corresponding land cover characteristics (as possible) in the corresponding grid cell. It is noteworthy that the available sites are concentrated near the outlet of the basin and thus are not representative of the continuous permafrost region in the

250 basin's far north nor the isolated patches in the southeastern part of the basin.

Table 1. Summary of permafrost monitoring sites[§].

Site name	Borehole	Latitude	Longitude	Data	Vegetation Cover	Permafrost Zone
Scotty Creek	Fen	61.306	-121.301	2012-2015	Graminoid (grasses), live moss over peatland	Sporadic
Petitot River North A	84-5A-T4	59.759	-119.516	1985-2000	Recovering burn (burned 2004), originally black spruce, shrubs, moss woodland (peat plateau)	Sporadic
Petitot River North B	84-5B-T4	59.759	-119.516	1985-1995		
Petitot River S	84-6-T6	59.460	-119.24	1985-2000	Peat plateau preceded by unfrozen fen	Sporadic
Wrigley Highway (Open black spruce)	99TC03	61.657	-121.344	2007-2015	Small black spruce thicket with willow shrub, 100% cover of moss with lichen and boreal heath	Discontinuous
Liard Spruce	97TC04	61.545	-121.394	2007-2015	Boreal, wetland shrub and sedge	Discontinuous
Manners Creek[*]	92TT-5	61.767	-121.185	1993-2005	Poplar, aspen	Discontinuous
	92TT-4	61.770	-121.193	1993-2007		

[§]*Scotty Creek is an experimental site operated by Wilfrid Laurier University* (Quinton and Marsh, 1999)*, and the other six sites are part of the Norman Wells-Zama pipeline monitoring program* (Smith et al., 2004)*,*
[*]*denotes sites that provide only active layer thickness measurement using a thaw tube.*

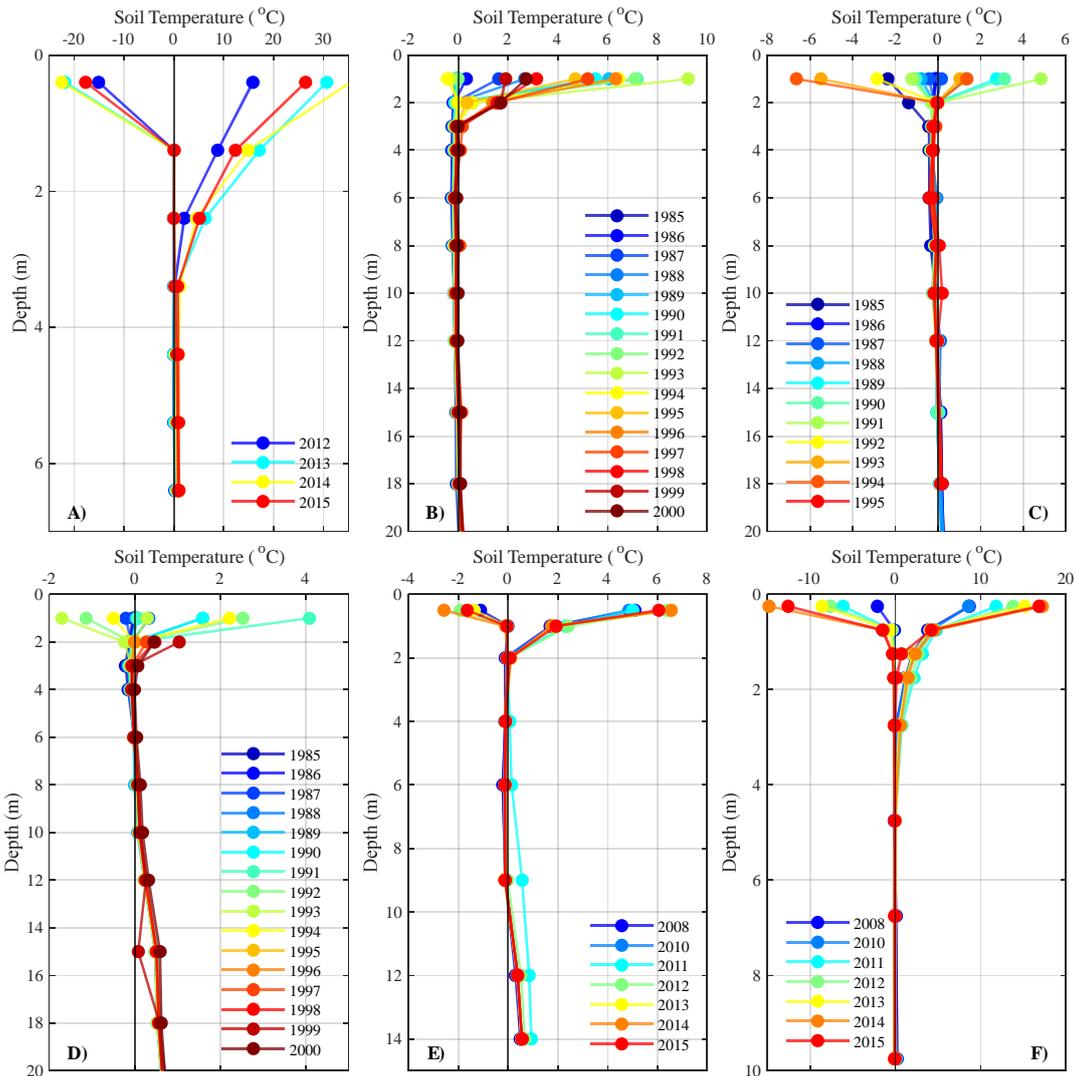



Fig. 3. Observed temperature envelopes at A) Scotty Creek (Fen), B) Petitot River North (84-5A-T4), C) Petitot River North (84-5B-T4), D) Petitot River South (84-6-T6), E) Wrigley Highway (99TC03), and F) Liard Spruce (97TC04) sites.

As highlighted earlier, the absence of a unified understanding of permafrost spatial heterogeneity has resulted in a wide range of mapping techniques for permafrost. Here, we explore several feasible permafrost products that can aid

260 in quantitatively assessing the quality of permafrost simulations. Five gridded products that provide different permafrost variables were utilized in our study (**Table 2**). These can be classified into three groups: 1) statistical approaches (*i.e.,* Aalto et al. (2018), Chadburn et al. (2017), and Gruber (2012)) based on the relationship between MAAT and the occurrence/extent of permafrost while accounting for variability in snow cover, vegetation, and subsurface properties (and terrain ruggedness for the last two datasets); 2) thermal/equilibrium models (*i.e.,* Obu et al.

265 (2019)) that employ a transfer function between air and ground temperatures to locate the freeze/thaw front; and 3) a machine learning approach (*i.e.,* Ran et al. (2022)) that requires a comprehensive set of inputs on soil density, organic content, precipitation, solar radiation, leaf area index, and snow cover duration. Further, we incorporated the International Permafrost Association (IPA) dataset, which provides permafrost zonation/distribution originally in a 1:10,000,000 paper map for the 1960-1990 period (Brown et al., 1998). This dataset was used to assess the simulated

270 PA, not PE, as it only provides the zonation boundaries, without further information on the variability within each zone. All datasets are remapped for the LRB extent using 0.125°×0.125° resolution and the first-order conservative scheme (Jones, 1999) for interpolation, except Chadburn et al. (2017), which is remapped using bilinear interpolation due to its coarser resolution (~0.5°×0.5°) than the LRB model setup (0.125°×0.125°).

Table 2. List of permafrost gridded datasets used in the study.

Dataset	Size (km)	Period	Variables	Approach	Forcing Dataset	Evaluation	Ensemble	Permafrost delineation
Gruber (2012)	1	1960-1990	PE	Statistical	CRU TS 2.0 & NCEP reanalysis	IPA PF area/region	5	Function of MAAT
Chadburn et al. (2017)	55	1960-1990	PE *	Statistical	WFDEI	IPA PF area/region	-	Function of MAAT
Aalto et al. (2018)	1	2000-2014	ALT, TZAA**	Statistical	WorldClim-GMFD	797(MAGT) & 303(ALT)	1000	-
Obu et al. (2019)	1	2000-2016	PE, MAGT	Thermal model	ERA-interim & MODIS LST	GTN-P 920(MAGT)	200	% of ensemble members with MAGT ≤ 0°
Ran et al. (2022)	1	2000-2016	PE, ALT, TZAA	Machine learning	WorldClim & MODIS LST	1002(MAGT) & 452(ALT)	1000	% of multi-model ensemble members with MAGT ≤ 0°

* Three permafrost extent maps are available based on three different relationships between MAAT and permafrost extent.

** TZAA denotes the temperature at the depth of zero annual amplitude.
Fig. 4 compares PA and the associated zonation for all gridded datasets; PE is shown in **Fig. 8**. These comparisons shed light on the discrepancies/commonalities among the datasets in terms of spatial patterns, zonation and total area. We employed the intermediate ("most-likely" as per the authors) permafrost extent map of Chadburn et al. (2017) to produce the PE map; the maximum and minimum values of each zone incorporated the calculations of PA for the Chadburn et al. (2017) and Brown et al. (1998) datasets. The total area underlain by permafrost (*i.e.,* permafrost fraction/probability per grid multiplied by grid area) was comparable for Gruber (2012), Chadburn et al. (2017), and Obu et al. (2019), varying between 87,000 and 90,000 km^2. However, the partitioning of each product's area into zones varies among these three products, as shown in **Fig. 4**, except for the continuous permafrost zone (*i.e.,* 5-10%). The digitized version of Brown et al. (1998) had the largest total area of permafrost at ~112,000 km^2, and the machine-learning product of Ran et al. (2022) had the smallest at ~64,000 km^2; however, the two products have dissimilar partitioning of the permafrost zones. Additional analysis on permafrost zonation is provided in Supplement **Section S5.1**.

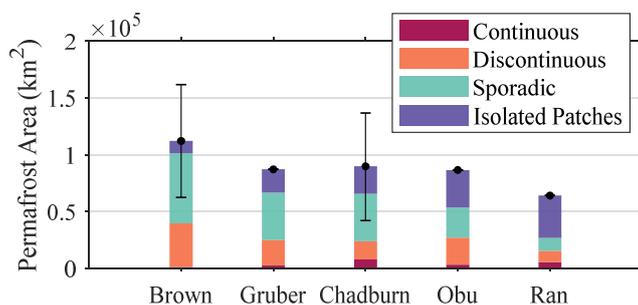



Fig. 4. Permafrost area for the LRB (total and partitioned per zonation) and the associated range of variability (error bars) based on the employed gridded products.

2.5   Modelling experiments - design and implementation

Experiments were designed to assess the capability of the MESH model to concurrently simulate the dynamics of permafrost, both spatially and vertically, in addition to the hydrology, considering the uncertainty of the most influential model parameters for the LRB. The model configuration was extracted from a larger MRB model (Elshamy et al., in preparation) that could replicate the observed streamflow (at different gauges, including the outlet of the LRB), different hydrological fluxes/states (including evapotranspiration and snowpack), and permafrost. The MRB model was configured with a deep soil column to 51.24 m using a power-function-based layer discretization (Supplement **Table S1**), with a regular grid size of 0.125°×0.125°. However, parameter estimation for the MRB model was based on a combined product of the Global Environmental Model (GEM) (Côté et al., 1998) atmospheric forecasts and the Canadian Precipitation Analysis (CaPA) (Mahfouf et al., 2007), or GEM-CaPA in short, which is a different forcing dataset from that used in the current study. We note that GEM-CaPA could not be used in the current study due to its short record length (not available before 2002) and the fact that it was not employed in the bias correction and downscaling of ISIMIP3b GCMs; refer to **Section 2.3** for further discussion.

305 Thus, the first step was to assess the compatibility of the estimated parameter values (based on GEM-CaPA) to the W5E5 V2 dataset. **Fig. 5** shows the simulated streamflow at the basin outlet (station 10ED002), forced with W5E5 V2 and GEM-CaPA meteorological products. The two simulated hydrographs are similar (KGE ≈ 0.8) and reasonably capture both the high and low flow events; still, the W5E5 V2 hydrograph depicts minor/modest differences for the volumetric bias (−16% 'underestimation' compared to −11% for GEM-CaPA) and the low flows (NSE-Log of 0.48

310 compared to 0.65 for GEM-CaPA). The assessment was executed for the three other stations (**Fig. 1F**) and showed similar patterns (Supplement **Section S3**). Likewise, the daily basin-average water-balance states/fluxes were comparable for the same model configuration under W5E5 V2 and GEM-CaPA forcings, in terms of quantities, trend, and variability (Supplement **Section S3**). These preliminary checks could not be extended to permafrost thermal profiles as the original model (MRB) was evaluated against three sub-watersheds outside the LRB; however, all of

315 the new experimental configurations (based on W5E5 V2) are evaluated against the available permafrost sites and datasets (**Table 2**).

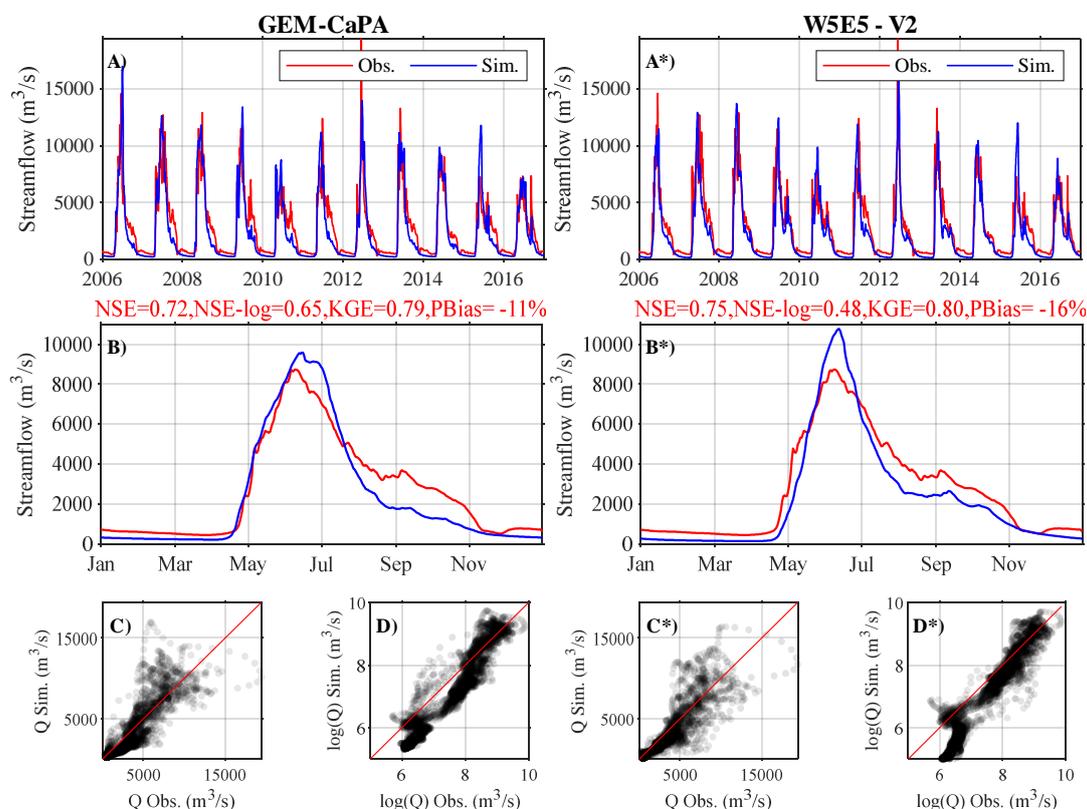



Fig. 5. Measured and simulated daily discharge of the LRB outlet (10ED002) under W5E5 V2 (left column) and GEM-CaPA (right column), for the original model configuration from 2006 to 2016, are compared using (A/A*) hydrographs, (B/B*) the long-term
320 mean daily discharge, and scatter plots for (C/C*) normal values and (D/D*) logarithmic values. Performance metrics are calculated for the whole period and are provided above subplots (B/B*) in red font.

The second step was to configure several experiments to incorporate the uncertainty in model parameters. In this regard, we followed a progressive approach (one-factor-at-a-time) to develop the experiments, with emphasis on the aspects of model configuration most sensitive to permafrost simulation, ending up with 16 model experiments,

325   summarized in **Table 3**. Based on a previous global sensitivity analysis (Abdelhamed et al., 2022b), permafrost simulation was found to be highly sensitive to two sets of parameters, representing surface insulation by snow cover (represented by the minimum snow depth to consider 100% ground snow cover, ZSNL) and soil texture, especially related to organic matter (type, vertical distribution, and depth of organic soil 'ODEP'), which collectively contribute ≥50% of the total sensitivity of permafrost simulation. The remaining parameters are either entirely insensitive (eight

330   parameters with <1% contribution) or have low-to-moderate sensitivity (15 parameters with 1-10% contribution). Further, these parameters (*i.e.*, snow cover and soil texture) were also among the most sensitive in the simulation of streamflow, as shown by Haghnegahdar et al. (2017).

The ZSNL parameter was perturbed for the dominant GRUs (*i.e.,* needleleaf forest (NLF), broadleaf forest (BLF), grass, and shrubs that collectively represent 84% of the LRB area; Supplement **Table S2**). Further, organic matter

335   parameters were perturbed over each grid cell of the LRB setup; all tiles within a grid cell are assigned the same soil texture value. As highlighted earlier in **Section 2.1**, the organic soil can be configured following two distinct methods that were assessed in the configured experiments. The depth of organic soil was also represented as either a fixed depth (top six layers, 0.85 m) or a 'spatially' varying depth based on the relationship between ODEP and the soil organic content (Supplement **Fig. S1** and **Table S3**). The vertical distribution of organic soil was configured using

340   either uniform values, values decaying with depth, or values decaying with soil layer number. All model setups were spun up to initialize state variables (*e.g.,* soil temperature, liquid/frozen soil content) by looping over a single hydrological year of climate record (hydrological year: October 1 to September 30) for 200 cycles for each setup, as recommended by Abdelhamed et al. (2022a). In this regard, two different single years were incorporated into our assessment (*i.e.,* 1979-1980 and 1980-1981), where the two years had almost the same total annual precipitation (~540

345   mm) and differed in terms of MAAT, at −1.5 and −3°C for 1979-1980 and 1980-1981, respectively. Computationally, running a single year of the LRB H-LSM took 20 mins on a full high-performance computing node (with 32 Xeon Gold Quad-Core CPU and a total memory of 326 GB), and hence each experiment required around three days of continuous running.

Table 3. Proposed model configurations, with the original model configuration denoted by *.

Exp ID	Snow Cover (ZSNL)				Organic Matter			Spin-up year
	Grass	**Shrubs**	**NLF**	**BLF**	**Type**	**Distribution**	**ODEP**	
1*	0.26	0.10	0.60	0.60	FOS	Uniform	Fixed	1979-1980
2	0.26	0.10	0.60	0.60	MSO	Uniform	Fixed	1979-1980
3	0.26	0.10	0.60	0.60	MSO	By Layer	Fixed	1979-1980
4	0.26	0.10	0.60	0.60	MSO	By Depth	Fixed	1979-1980
5	0.26	0.10	0.30	0.30	MSO	By Depth	Fixed	1979-1980
6	0.26	0.10	0.15	0.30	MSO	By Depth	Fixed	1979-1980
7	0.13	0.05	0.15	0.15	MSO	By Depth	Fixed	1979-1980
8	0.13	0.05	0.15	0.15	MSO	By Depth	Fixed	1980-1981
9	0.13	0.05	0.15	0.15	MSO	Uniform	Fixed	1980-1981
10	0.13	0.05	0.15	0.15	MSO	Uniform	Fixed	1979-1980
11	0.13	0.05	0.15	0.15	FOS	Uniform	Fixed	1979-1980
12	0.13	0.05	0.15	0.15	FOS	Uniform	Fixed	1980-1981
13	0.13	0.05	0.15	0.15	MSO	Uniform	Varying	1980-1981
14	0.13	0.05	0.15	0.15	MSO	Uniform	Varying	1979-1980
15	0.13	0.05	0.15	0.15	MSO	By Depth	Varying	1979-1980
16	0.13	0.05	0.15	0.15	MSO	By Depth	Varying	1980-1981

350

As highlighted earlier, the assessment also focuses on the quality of simulated hydrology in the LRB in terms of the main water-balance states/fluxes, and the generated streamflow at four gauge stations (**Fig. 1F**). The selected stations are located on the main stem of the Liard River and correspond to different zones in the basin, including a headwaters station (10AA001), two intermediate stations (10BE001 and 10ED001), and the outlet station (10ED002) (Supplement

355 **Table S4**). Four performance metrics are utilized to assess different aspects of the simulated hydrographs for 2000-2016: the Nash-Sutcliffe efficiency (NSE), Nash-Sutcliffe efficiency of log-transformed flow (NSE-Log), Kling-Gupta efficiency (KGE), and percentage bias (PB). We also investigated the range of variability introduced to the basin-averaged evapotranspiration, change of storage, and surface runoff in response to the various proposed model parameterizations.

360 **3   Results and discussion**

3.1   Permafrost evaluation

3.1.1   Ground observations

The primary indicator of permafrost simulation quality can be inferred from the reproducibility of the observed temperature envelopes (*i.e.,* T_{max} and T_{min}), comparing the ground observation time series to that of corresponding

365 model grid cell and land cover. For example, **Fig. 6** shows the observed and simulated temperature envelopes for the Wrigley Highway site (99TC03) in 2014; the site is covered by evergreen shrubs (amongst other vegetation) and, thus, the simulated values are extracted for the shrubs GRU of the corresponding model cell.

Overlaying all simulated envelopes facilitates visual comparison of the performance of model experiments to the observations, quantifies the range of variability associated with all experiments, and highlights any odd model
370 response. In this regard, 10 experiments (*i.e.*, 3, 4, 5, 6, 7, 8, 9, 13, 15, and 16) form no permafrost at all, which is seen from the positive T_{min} envelope of these experiments (below the active layer), which is shifted to the right by 1.0 to 3.0 °C. On the other hand, two experiments (*i.e.,* 1 and 11) yield cooler permafrost conditions than the observations; the envelope is shifted to the left by -1.8 and -1.4 °C, respectively. The remaining four experiments are ranked as follows: Exp. 12 provides the best performance for temperature envelopes and ALT, followed by Exp. 2, and the third-
375 best performance is equally achieved by Exps. 14 and 10 that have a relatively deeper ALT with a warm-biased T_{max} compared to Exps. 12 and 2.

However, the Wrigley Highway site, as many of the other sites (**Table 1**), has a complex canopy, which necessitate comparison of the observed temperature envelopes against multiple modelled GRUs because CLASS does not allow the overlaying of different canopy types. The relevant single-year envelopes and associated performance metrics for
380 the needleleaf forest tile of the Wrigley Highway site are provided in **Fig. S15** and **Table S12** in the Supplement. The best-performing experiments are as follows: Exp. 15 has the best match to ground observations, followed by Exps. 6 and 9. Obviously, the ranking based on the needleleaf forest GRU is completely different from that of the shrubs GRU, highlighting that one model configuration does not always work for all GRUs, particularly for sites with a complex canopy.

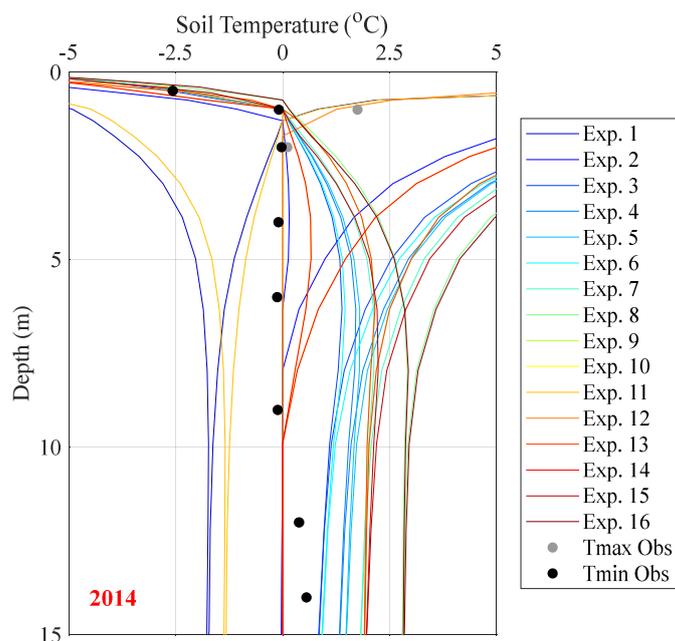
385 For a quantitative assessment, three statistical performance metrics are calculated for T_{max} and T_{min} for all experiments. These metrics are the root mean square error (RMSE), mean absolute error (MAE), and the BIAS, whenever observations were available. The error metrics were averaged over space (soil column) and time (whole record length). This approach was selected due to its generality, simplicity, and suitability to assess different aspects of the whole soil column. **Table 4** provides these statistics for the Wrigley Highway site for the shrubs GRU. Excluding experiments
390 that did not form permafrost and using the provided statistics in **Table 4**, the overall ranking of model experiments can be obtained: Exps. 10 and 14 have the best agreement to ground observations, followed by Exp. 12, and then Exp. 2; such ranking differs from the single-year approach. For the remaining sites with observed temperature profiles, a similar comparison of the 16 model configurations versus ground observations and the associated statistical measures are provided in **Section S4** in the Supplement.

395

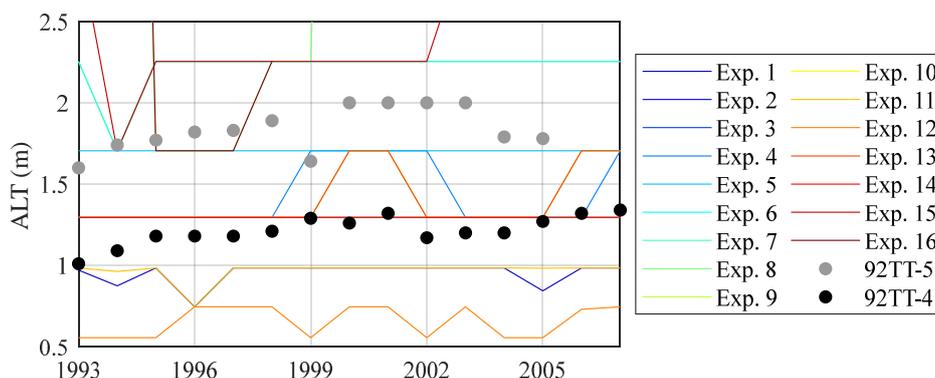
Fig. 6. Observed and simulated temperature profiles at the Wrigley Highway site for the shrubs GRU for the 16 model experiments. The profiles correspond to the same year (2014) for all of the experiments.

Table 4. Summary of performance metrics for temperature profiles at the Wrigley Highway site for the shrubs GRU for the 16 400 model experiments. Each metric is averaged over the whole observation period. Refer to **Table 3** for further information about the configured experiments.

Exp. ID	Tmin			Tmax		
	RMSE	BIAS	MAE	RMSE	BIAS	MAE
1	5.71	5.09	5.09	3.68	-1.50	3.49
2	2.08	1.60	1.80	5.09	-4.04	4.31
3	2.03	0.81	1.85	6.29	-5.28	5.35
4	2.18	0.47	2.05	6.61	-5.76	5.76
5	2.24	0.39	2.12	6.65	-5.85	5.85
6	2.01	0.75	1.83	6.44	-5.41	5.47
7	2.34	0.12	2.21	6.90	-6.14	6.14
8	2.85	-0.26	2.63	7.35	-6.79	6.79
9	2.57	0.39	2.45	6.36	-5.79	5.79
10	1.91	1.30	1.74	5.55	-4.45	4.71
11	5.10	4.44	4.44	3.46	-1.66	3.27
12	2.88	2.44	2.60	3.01	-1.83	2.72
13	2.57	0.39	2.45	6.36	-5.79	5.79
14	1.91	1.30	1.74	5.55	-4.45	4.71
15	2.39	0.05	2.25	6.89	-6.19	6.19
16	2.85	-0.28	2.62	7.32	-6.79	6.79

The thaw-tube-based ALT data from the Manners Creek site were also used, where two observational tubes in close proximity (*i.e.,* 92TT-5 and 921TT-4) were active between 1993 and 2007. The two sites are located within the same

405   model grid cell and have the same vegetation cover, hence their time series of observed ALT are compared against the same simulated ALT for the needleleaf forest GRU. **Fig. 7** shows the observed and simulated ALT while **Table S14** (in the Supplement) shows the statistical metrics for each experiment. However, because the difference in observed ALT is quite significant (0.35 to 0.83 m), the ranking of experiments depicts a trade-off between the two tubes, which highlights the spatial heterogeneity of permafrost over short distances. The best-performing experiments

410   for the 92TT-5 site are Exps. 5 and 9, while the best for the 92TT-4 site are Exps. 2, 10, and 14. However, we opted to give a higher weight to the 92TT-4 site in our subsequent analysis due to the consistency of its observations (no odd shift in the observed values as occurs in 1999) and completeness of the data (the record at 92TT-5 is shorter by 2 years than at 92TT-4).

415   **Fig. 7.** Simulated and observed ALT for Manners Creek sites (92TT-5 and 92TT-4) for the needleleaf forest GRU.

To summarize the results for all sites, associated GRUs per site, and observed variables, the overall ranking of experiments is provided in **Table 5**. A lower ranking value corresponds to higher simulation quality (*i.e.,* 1 is the best and 16 is the worst). We excluded the experiments that did not form permafrost (denoted by *x* in **Table 5**) for more straightforward interpretation. The table confirms the complexity of permafrost-based H-LSM evaluation and the

420   difficulty of finding a single configuration that can replicate the observed permafrost for sites with complex canopy types.

Table 5. Average ranking of experiments based on performance against different permafrost sites and the associated GRU. Lower ranking values denote better simulation quality (*e.g.,* 1 is best and 16 is worst). *x* denotes the failure to form permafrost for a given model experiment (unlike the observation) and exclusion from the ranking. NLF: Needleleaf Forest and BLF: Broadleaf Forest. Refer to **Table 3** for further information about the experimental configurations.

Exp. ID	Scotty Creek			Petitot River South		Petitot River North		Wrigley Highway		Liard Spruce	Manners Creek	
	Fen			84-6-T6		84-5A-T4	84-5B-T4	99TC03		97TC04	92TT-5	92TT-4
	Wetland	NLF	BLF	Wetland	Shrubs	Wetland	Wetland	Shrubs	NLF	Wetland	NLF	NLF
1	3.3	14.8	9.2	3.3	5.5	12.0	11.2	5.0	14.0	2.8	9.3	6.7
2	4.0	14.2	9.5	5.5	3.5	10.5	9.5	3.3	12.8	3.0	8.0	2.0
3	x	13.0	7.5	4.3	x	5.0	6.3	x	11.7	x	7.0	5.0
4	x	11.7	6.5	x	x	5.8	6.3	x	10.7	x	6.0	8.0
5	x	7.5	3.2	x	x	5.7	6.7	x	5.0	x	4.3	10.3
6	x	1.7	2.8	x	x	7.0	6.0	x	4.0	x	5.3	12.0
7	x	1.7	x	x	x	6.7	6.0	x	4.0	x	5.3	12.0
8	x	4.3	x	x	x	x	x	x	x	x	16.0	16.0
9	x	4.5	x	x	x	2.2	2.0	x	4.2	x	6.0	6.0
10	2.0	7.2	2.0	2.5	2.0	5.0	6.0	2.3	5.0	2.5	7.0	3.0
11	2.7	11.2	7.0	1.8	4.0	9.0	7.7	4.0	10.0	3.2	9.0	6.3
12	x	10.2	4.3	x	3.0	x	x	3.0	8.0	x	9.7	7.7
13	x	4.5	x	x	x	2.2	2.0	x	4.2	x	6.0	6.0
14	2.0	7.2	2.0	2.5	2.0	5.0	6.0	2.3	5.0	2.5	7.0	3.0
15	x	3.5	x	x	x	x	x	x	3.5	x	12.0	14.0
16	x	x	x	x	x	x	x	x	x	x	15.0	15.0

3.1.2 Gridded datasets

In this section, we compare the simulated PE (and PA) and ALT to different gridded products. The selected datasets reflect different mapping techniques, including the level of physics, forcing datasets, spatial resolution, and temporal coverage (refer to **Table 2** and **Section 2.4** for further discussion). **Fig. 8** shows the simulated PE versus four gridded products, where the simulated extent (and associated area) is averaged from 2000-2016, the baseline time period for the Obu et al. (2019) and Ran et al. (2022) datasets; the other two datasets correspond to 1960-1990. Several points can be observed:

- The spatial extents of the four gridded products differ, except for the far north, south, and southeast parts of the LRB. For instance, the Gruber (2012) and Chadburn et al. (2017) products characterize the finger-shaped basin outlet by low values (~0.3), while the Obu et al. (2019) and Ran et al. (2022) products yield very high (~1) and very low values (~0), respectively.
- The PA of the gridded products does not vary as much as the PE, because 31.5 to 32.6% of the LRB is underlain by permafrost for all datasets except Ran et al. (2022), which mapped a smaller permafrost percentage (23.3%). A comprehensive assessment of the temporal evolution of the PA is provided in **Section S5.2** in the Supplement.
- Exp. 1 (the original configuration), which had satisfactory performance for streamflow (**Fig. 5**), overestimates the PE over the basin compared to the gridded products. Similar overestimation, to varying degrees, is found for Exps. 2, 3, and 4, which were configured using different organic matter types and vertical distributions (**Table 3**).

- Switching the organic matter type has a minor impact on the PA in Exp. 1 vs. Exp. 2 (reduction of 2,700 km^2), but a more significant impact in Exp. 10 vs. Exp. 11 (reduction of 20,000 km^2). The latter pair of experiments has a reduced ZSNL value and, hence, the slight change in the organic matter configuration could have a greater impact on the simulated permafrost as organic matter is a major driver of insulation in this case.

- The impact of the vertical distribution of organic soil is relatively pronounced, as the non-uniform decay of organic matter (Exps. 3 and 4) reduces the PA by 15,000 km^2.

- The gradual reduction in ZSNL value, the most sensitive permafrost parameter in the MESH model (Abdelhamed et al., 2022b), resulted in a drastic reduction in PE (and PA of ~77,000 km^2) in Exp. 4 vs. Exp. 7. ZSNL describes heat insulation at the soil surface and represents the amount and extent of accumulated snow; a reduction in ZSNL value means snow cover reaches 100% earlier and thus insulates the soil from extreme cold in winter and results in less permafrost.

- Experiments spun up with a warmer year (1980-1981) had a lower PE (more or less consistent over the basin, except for the region near the outlet that had high organic matter content; see **Fig. 1E**) and smaller PA when compared to the experiments spun up with a cooler year (1979-1980).

- The impact of the spin-up year for the LRB can be estimated as a reduction in PA of ~30,000 km^2, representing 11% of the LRB area. This can be demonstrated by comparing experiment pairs 7 vs. 8, 10 vs. 9, 11 vs. 12, 14 vs. 13, and 15 vs. 16.

- Employing a variable (gridded) depth of organic soil, instead of a fixed value (0.85 m), resulted in a slight reduction in permafrost extent and total area (~7,000 km^2), as shown by comparing Exps. 9 and 13.

- Visually, four experiments (Exps. 8, 9, 13, and 16) have a reasonable agreement with the four gridded datasets.

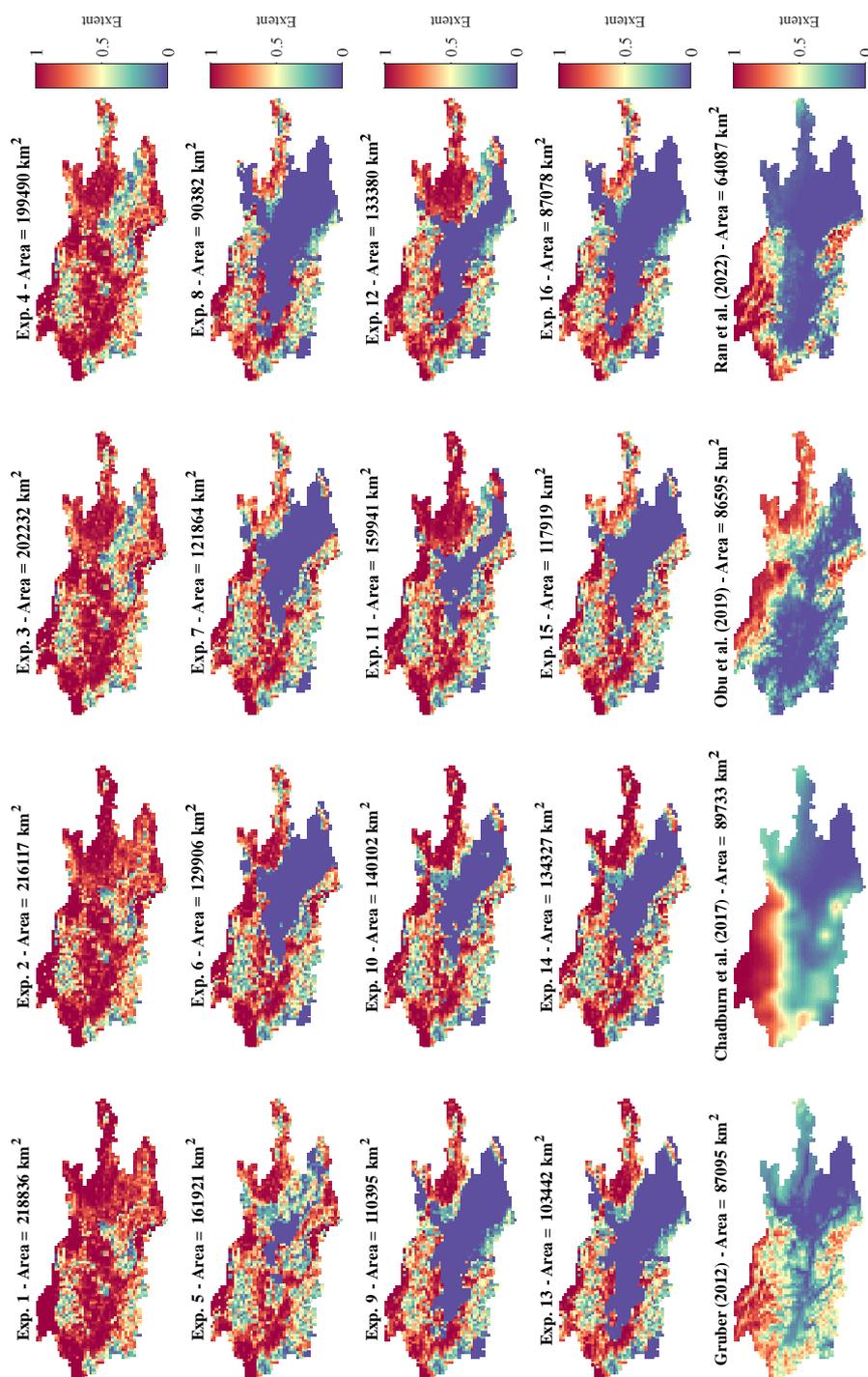



Fig. 8. Comparison of the simulated permafrost extent for the 16 model experiments to four gridded permafrost products (shown in the bottom row). Permafrost extent is averaged over 2000–2016, and the total area underlain by permafrost is given above each subplot.

ALT sheds light upon the thermal regime and describes permafrost dynamics. **Fig. 9** shows the simulated ALT versus the Aalto et al. (2018) and Ran et al. (2022) gridded permafrost products; the other datasets used to assess PE do not provide ALT (see **Table 2**) while Aalto et al. (2018) did not provide PE. The simulated ALT is averaged for 2000 to 2016, which is the representative period for the Ran et al. (2022) dataset; the other dataset corresponds to 2000-2014.

475 Several points can be observed:

- The two datasets agree, to a certain level, on the spatial extent of the ALT over the LRB, except for the middle of the basin where Ran et al. (2022) has fewer pixels underlain by permafrost. The two datasets also agree on the ALT values, for which the basin-average difference is around 10 cm (as shown Supplement **Fig. S20**).

480 - The original configuration (Exp. 1) overestimates ALT in the far north and south (red cells in **Fig. 9**). The same pattern is produced by Exps. 2, 3, and 4, where the changes in the organic soil configuration did not reduce these high ALT values.

- Similar to the PE analysis, the impact of the spin-up year on the simulated ALT is more pronounced than the organic matter configuration (Exps.1-4), where a clear reduction in deep ALT pixels is observed. This is

485 shown by comparing experiment pairs 7 vs. 8, 10 vs. 9, 11 vs. 12, 14 vs. 13, and 15 vs. 16.

- Similar to the PE analysis, the utilization of a variable organic depth instead of a constant value has a minor impact on the simulated ALT, as depicted by comparing Exps. 13 and 9.

- Lastly, comparing the two gridded products (**Fig. 9**) to ground observations (**Fig. 3**) suggests that gridded products tend to underestimate ALT, as four out of six sites used in the current study have ALT with a depth

490 between 2 and 4 m, whereas the max estimated ALT value for the two products is 2.2 m.

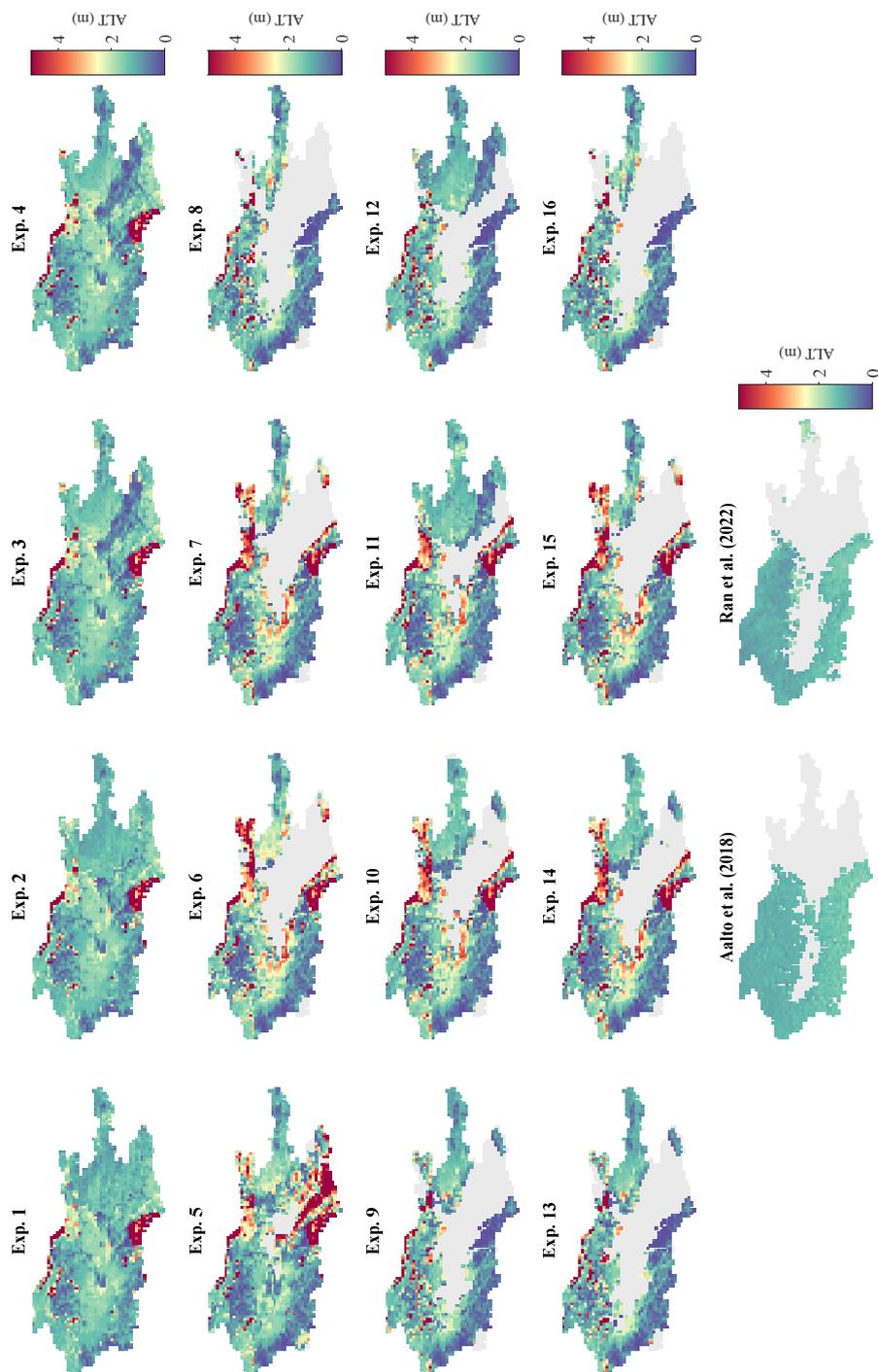



Fig. 9. Comparison of the simulated active layer thickness (ALT) for the 16 model experiments with the two gridded permafrost products (shown in the bottom row). The simulated ALT is averaged over 2000-2016. Gray shading represents area with no permafrost in simulations and gridded products.

The visual assessments for PE (**Fig. 8**) and ALT (**Fig. 9**) provide a qualitative indicator of simulations but lack an
495 explicit quantitative measure. This can be rectified by comparing the cumulative frequency distributions (CDFs) of
the simulated and gridded products and/or performing a grid-to-grid statistical comparison for the simulated and
observed variables. The CDFs facilitate an additional visual comparison of the distribution of each experiment to the
gridded-based CDFs in only one figure (**Fig. S18** for PE and **Fig. S19** for ALT), which also can be assessed statistically
using the Kolmogorov–Smirnov test. However, this approach offers no information on the spatial pattern/correlation
500 or systemic bias. In the latter approach, different statistical measures/tests can be directly applied in a grid-to-grid
manner, including correlation and bias. However, the re-mapping/re-gridding applied for each product introduces
further uncertainty (**Section 2.4**) and thus can affect assessment reliability. Nevertheless, we opted to utilize the grid-
to-grid statistical measures to quantitatively assess and systematically rank the experiments.

A summary of Spearman's rho, difference (or error/bias), and RMSE for the simulated PE and ALT of all experiments
505 is presented in **Table S15** in the Supplement, calculated for the representative period for each dataset (see **Table 2**).
Noting that the Gruber (2012) and Brown et al. (1998) datasets correspond to 1960-1990 and our simulation started
later (1979 or 1980), we compared these datasets to the average values from the starting year till the end of 1990. For
PE, Exps. 8 and 16 have the highest correlation to the Gruber (2012), Chadburn et al. (2017), and Ran et al. (2022)
datasets with correlation values of 0.55, 0.66, and 0.58, respectively, confirmed by visual assessment (**Fig. 8**).
510 However, another pair of experiments, namely Exps. 12 and 13, compare better to the Obu et al. (2019) dataset. **Table
S15** (in the Supplement) provides an estimate of the difference and RMSE, which can be used to select the best-
performing configuration for each dataset. For instance, Exp. 13 is better than Exp. 12 when compared to the Obu et
al. (2019) dataset, with a lower difference (0.06 rather than 0.17) and a lower RMSE (0.72 instead of 0.77). A similar
assessment can be done for ALT. For instance, Exp. 16 is the best configuration versus the Aalto et al. (2018) and Ran
515 et al. (2022) datasets, shown by minimal difference (−0.06 and −0.18) and RMSE (1.5 and 1.5) across all of the
experiments. Nevertheless, all experiments have negative correlations with the gridded products, varying between
−0.04 and −0.42 for Aalto et al. (2018) and −0.06 and −0.48 for Ran et al. (2022), which brings into question the
suitability of these datasets for the current study area.

Lastly, we rank all of the experiments from 1 to 16 based on the calculated statistics in **Table S15**, where, as above, 1
520 corresponds to the best and 16 to the worst experiment. For each experiment, we determine the rank for each individual
metric and then average the rank for the three statistical measures per experiment (**Table 6**). The averaging ensures
the final ranking integrates both the spatial correlation and the bias of each simulation. According to **Table 6**, Exp. 16
is the best-performing configuration for most gridded permafrost products, except the Obu et al. (2019) dataset for
which Exp. 13 is relatively better. On the other hand, the worst-performing configuration versus the gridded permafrost
525 extent datasets is Exp. 1 (*i.e.,* the original configuration), while Exps. 5 and 6 have the poorest performance against
the Aalto et al. (2018) and Ran et al. (2022) datasets, respectively.

Table 6. Average ranking of experiments based on their performance against different permafrost gridded products for the period 2000-2016. Refer to **Table 3** for further information about the configured experiments.

Exp. ID	Permafrost extent				Active layer thickness	
	Gruber (2012)	Chadburn et al. (2017)	Obu et al. (2019)	Ran et al. (2022)	Aalto et al. (2018)	Ran (2022)
1	16.0	15.7	14.7	15.7	8.3	7.0
2	15.0	15.3	14.3	15.3	7.3	6.3
3	14.0	14.0	14.3	14.0	8.0	8.7
4	13.0	13.0	14.0	13.0	7.0	8.3
5	11.0	11.3	12.7	11.7	11.3	10.7
6	6.7	7.0	8.7	7.0	11.3	11.3
7	5.0	5.0	8.7	6.0	11.0	10.3
8	2.0	2.0	4.7	2.0	5.7	5.3
9	4.7	3.7	3.3	3.7	7.7	7.3
10	9.0	9.3	7.7	9.3	9.3	9.0
11	11.7	11.7	8.3	11.3	8.7	10.3
12	9.3	8.3	5.3	9.3	8.7	8.7
13	4.3	4.7	1.7	3.3	6.7	7.0
14	8.7	9.0	6.3	8.3	10.0	10.3
15	4.7	5.0	7.7	5.0	10.3	10.3
16	1.0	1.0	3.7	1.0	4.7	5.0

530   Notably, each of these datasets is characterized by significant uncertainty due to the imperfect representation of permafrost processes and upscaling issues, and cannot be used solely to assess the quality of permafrost zonation, as highlighted by Gruber (2012): "While the dataset presented here can be used as a reference for model evaluation, it does then by no means represent reliable ground truth." Another source of uncertainty with these products is implicitly related to our limited knowledge of adequate permafrost representation in ESMs/LSMs (*i.e.,* structural uncertainty),

535   from which the modelled air temperature is used as an input to most of these gridded permafrost products. Further, extending these products into the future (*e.g.,* Aalto et al., 2018; Chadburn et al., 2017) is not free of uncertainty in air temperature, which depends on the quality of permafrost representation in ESMs/LSMs, including scenario uncertainty, noting that some of the latest ESMs/LSMs still fail to properly represent permafrost dynamics (Burke et al., 2020; Lawrence and Slater, 2005).

3.2 Hydrologic evaluation

This section presents the simulations of streamflow and the main water-balance components and their implications. This analysis is vital due to the interconnection of permafrost status and surface/subsurface hydrologic connectivity (Connon et al., 2014), which could be reflected in the resultant streamflow and the partitioning of the water-balance components. **Table S16** (in the Supplement) summarizes the streamflow performance metrics for all 16 experiments at the four selected gauge stations over the period 2000-2016. The variability ranges for the NSE, KGE, and PBias are relatively small. For instance, PBias variability ranges between 2.5 and 3.1% for each gauge station across all model experiments. However, gauges 10AA001 and 10BE001 have the highest PBias (underestimation) of 21.9 to 24.8%, which is deemed a satisfactory result as per Moriasi et al. (2007) performance ratings. Another example of the limited variability across all experiments can be found for the NSE and KGE, which vary across all stations by 0.14 and 0.16, respectively, noting that the minimum/maximum values are 0.62/0.76 and 0.63/0.79, yielding a 'good' performance rating as per Moriasi et al. (2007). On the other hand, NSE_{Log} varied significantly, with 10AA001 and 10BE001 gauges having negative lower bounds (−0.53 and −0.56) and positive upper bounds (0.64 and 0.61). Further, the other two stations (*i.e.,* 10AA001 and 10ED002) had variability ranges that were 'relatively less' broad for the NSE_{Log} (0.19/0.79 and 0.19/0.77, respectively), making the performance rating vary from 'unsatisfactory' to 'very good' as per Moriasi et al. (2007).

Such results highlight that the overall streamflow is not sensitive to permafrost status, except for the low flows, which is a small component of LRB discharge (**Fig. 10**). The analysis of the basin-average water balance also accentuates the limited impact of permafrost status (represented by the 16 configured experiments) on the partitioning of surface fluxes (evapotranspiration (ET) and total runoff), which is not the case for the basin-average soil moisture storage or, to be more specific, the partitioning of frozen/liquid soil water content (see **Fig. S22** in the Supplement).

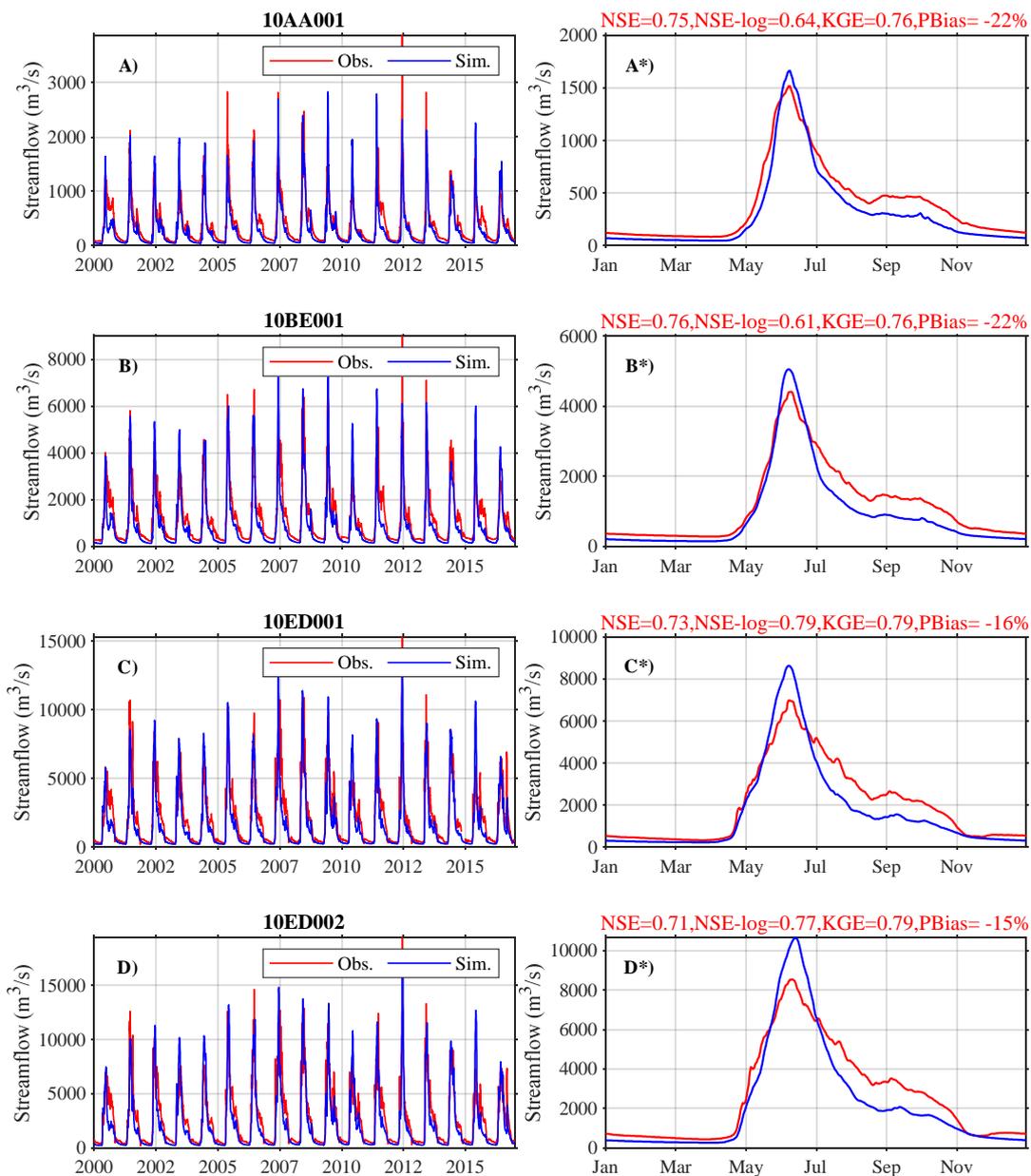



Fig. 10. Daily measured and simulated streamflow for the best performing experiment in terms of streamflow (ID 12) for the 2000-2016 period at four different gauge stations in the LRB: A) 10AA001 (headwaters), B) 10BE001 (intermediate), C) 10ED001 (intermediate), and D) 10ED002 (LRB outlet). For each station, the long-term mean daily hydrograph over the year is provided in a separate subplot with (*) designation, along with performance metrics (in red font).

Lastly, **Table 7** provides the ranking of all experiments for each gauge station; following the same approach as for permafrost (**Section 3.1**), the ranking for the four performance metrics is averaged for each experiment and for each station. Two main remarks can be extracted from **Table 7**: 1) several experiments can reproduce the observed

570   streamflow at the four gauges with 'slightly' superior performance when compared to the original configuration (Exp. 1) and 2) Exp. 12 is the best-performing experiment for all gauges.

Table 7. Average ranking of experiments based on their performance against the observed streamflow at four stations for the 2000-2016 period. Refer to **Table 3** for further information about the configured experiments.

Exp. ID	10AA001	10BE001	10ED001	10ED002
1	10.5	11.3	5.0	5.5
2	13.5	13.8	6.5	6.3
3	12.5	13.8	7.3	7.0
4	11.5	13.0	6.3	6.3
5	9.3	10.8	9.5	9.3
6	8.5	9.0	10.3	9.5
7	10.3	10.3	12.5	12.0
8	4.5	4.0	8.0	8.5
9	6.8	3.5	7.5	8.3
10	11.8	11.5	13.3	13.0
11	5.0	4.3	4.0	4.0
12	1.3	1.0	1.8	2.0
13	6.0	5.0	9.3	9.5
14	10.3	10.0	13.5	13.0
15	9.0	10.0	12.3	12.3
16	5.5	5.0	9.3	9.8

575   **3.3   Overall Evaluation and Ranking**

Here, the analysis results for permafrost and hydrology are combined, aiming to answer the main research question; *is it possible to concurrently simulate permafrost dynamics and hydrology over a large domain using an H-LSM?* This can be achieved by comparing the ranking of the experiments based on the local-scale assessment (**Table 5**), the large-scale gridded products assessment (**Table 6**), and the streamflow assessment (**Table 7**). **Fig. 11** (and Supplement

580   **Table S17**) provides the average ranking for each criterion, using equal weights for each sub-criterion (*e.g.,* different gridded products, or sites/stations). As per **Fig. 11**, Exps. 10 and 14 have the highest rankings (and best performance) among the eight observational sites used in the study. However, these two experiments are ranked as the 10th and 9th, respectively, based on their performance against the gridded datasets. Further, the same two experiments performed the worst for the streamflow simulation, ranked as 16th and 15th, respectively. Similarly, the best-performing

585   experiment versus all gridded products is Exp. 16, which is deemed the worst experiment for ground observations (16th) and with a moderate ranking for streamflow (5th). In the same context, the best experiment regarding streamflow simulation (Exp. 12) ranked as 11th for the ground observations and 7th for the gridded permafrost products.

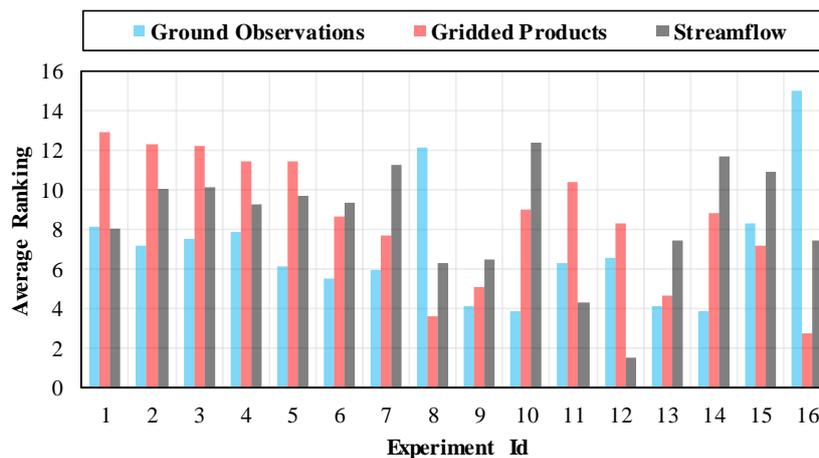



Fig. 11. Ranking of experiments based on their performance against permafrost ground observations, permafrost gridded products, and streamflow. For each criterion, the ranking is averaged over all of its items (*i.e.,* sites, products, and gauge stations).

This result highlights the difficulty that modellers may encounter when configuring models to simultaneously simulate permafrost and hydrology (streamflow). However, it does not provide a concrete answer to our main research question, as there might be a trade-off between local- and large-scale permafrost aspects and streamflow simulation quality that cannot be fully identified from our work due to the: 1) limited number of experimental configurations (constrained by time and computational resources), 2) limited spatial coverage of the available observational sites, 3) large uncertainty in the gridded permafrost products (as provided and due to the re-gridding/interpolation done for the current study), and 4) uncertainty in meteorological forcing variables (inherently and due to the spatial regridding and temporal disaggregation done for the current study). Still, we can conclude that the consideration of streamflow is beneficial when developing and diagnosing an H-LSM for a permafrost-dominated region. In other words, for study areas underlain by permafrost, it is advantageous to concurrently assess the spatio-temporal evolution of both hydrologic and thermal regimes under different data sources that provide complementary information. One way to address this modelling challenge is by building an ensemble of H-LSM setups that account for different hydrologic/thermal trajectories/responses, which can then be used for short-term hindcasting or long-term projection.

4   Summary and conclusions

Over the past five decades, significant effort has been directed at enhancing the realism, functionality, and fidelity of H-LSMs. This stems from incorporating state-of-the-art representation of processes and benefiting from the rapid advancement in remote sensing technology and ever-growing computational resources. However, the downside of this enhancement is the remarkable increase in H-LSM dimensionality and complexity, which complicates parameter identification. Moreover, simulating the dynamics of permafrost is further challenged by the limited availability of field observations, which, unlike other hydrologic states/fluxes (*e.g.,* snow water equivalent), cannot presently be remotely sensed. Further, the simulation of permafrost thermal regime has a substantial impact on the partitioning of energy/water fluxes, which if not well represented/constrained could produce deceptive trajectories for the present

and future climate and hydrology. In this article, we explored the challenges of configuring a MESH model that can simultaneously reproduce the observed permafrost and streamflow dynamics in the Liard River Basin (LRB) of Canada. The assessment included different hydrologic and permafrost-related variables, for which different data products were utilized.

Building upon previous research, the LRB setup was first extracted from an existing setup for the Mackenzie River Basin (MRB) (Elshamy et al., in preparation), which was calibrated/validated for permafrost and streamflow using a relatively short forcing dataset (GEM-CaPA: 2002-2016). However, in this work, a longer forcing dataset was used (W5E5 V2: 1979-2019) with reasonable temporal coverage for the available permafrost data. This dataset was also used for bias correcting and downscaling the future projections of different GCMs (ISIMIP3b protocol), which will be utilized in a climate change assessment for the LRB in a subsequent manuscript. The LRB model response was evaluated under W5E5 V2 to verify the applicability/compatibility of the estimated parameter values (based on GEM-CaPA). This step included comprehensive assessment of streamflow and water-balance states/fluxes, which showed modest/minor changes in the performance under W5E5 V2. However, the model did not perform well for permafrost and thus some parameters and configuration aspects were revisited to improve permafrost simulation and their impacts were then checked for hydrology. 15 additional model setups for the LRB were configured, where the most sensitive parameters were progressively perturbed. Each experiment was later evaluated against permafrost and streamflow observations.

In this study, ground observations from seven permafrost sites were utilized to evaluate simulated temperature envelopes (T_{min} and T_{max}) and active layer thickness (ALT) using RMSE, BIAS, and MAE metrics. To clarify the relative performance of all experiments, a ranking was performed based on the three performance measures for each site. The metric-based ranking underlined the difficulty of electing a single configuration with reasonable performance against ground observations for all sites simultaneously, which became more problematic for complex-canopy sites. The significant spatial heterogeneity of permafrost observations poses a further challenge as sites in close proximity (falling in the same model pixel) can have different profiles or even different permafrost states (even under the same vegetation) due to soil heterogeneity, which cannot be easily captured at large scales. For instance, ALT was monitored at Manners Creek (discontinuous permafrost) using two thaw tubes (92TT-4 and 92TT-5) located within 500 m with the same vegetation cover; however, the observations depicted a significant difference in ALT (0.5 m on average), highlighting the spatial heterogeneity of permafrost over short distances.

Despite the fact that ground observations represent the best available knowledge, their limited spatial coverage hinders any comprehensive assessment of permafrost dynamics over large scales. Permafrost mapping can aid in addressing the lack of a well-distributed observational network. In this study, we incorporated four gridded permafrost products (*i.e.,* Gruber (2012), Chadburn et al. (2017), Obu et al. (2019), and Ran et al. (2022)) that provide estimates of permafrost extent (PE) and the associated permafrost area (PA), and two products (*i.e.,* (Aalto et al. (2018) and Ran et al. (2022)) that offer ALT maps. These datasets were re-gridded to the same resolution as the LRB model for a more straightforward comparison and interpretation of simulation results. However, there is no overwhelming consensus among the gridded products on the spatial pattern of PE and ALT. Similarly, the derived PAs from these products

were highly divergent, varying between 64,000 and 112,000 km^2. The same remark is valid for the ALT products
650   which had persistent difference of 0.25-0.5 m over the basin, despite depicting similar spatial patterns. This is
understandable (and predictable) because diverse methods are available to map permafrost based on different
philosophies, definitions, and assumptions, and thus the level of agreement between them can be relatively low.
However, we considered them as having equal weights to the ground observations in our assessment because the latter
sites are not well distributed across the LRB.

655   To shed light upon the spatial correlation between experiments and gridded products, a grid-to-grid statistical analysis
was performed using Spearman's rho, the difference (or error if those products are considered quasi-observed), and
RMSE for PE and ALT. Remarkably, two model experiments (Exps. 8 and 16) had a high correlation and low error
with respect to all gridded PE datasets, except for Obu et al. (2019) for which another experiment (Exp. 9) yielded
better performance. The same analysis was performed for the ALT, which showed that no single experiment had a
660   positive correlation to the two datasets; two model experiments (Exp. 8 and 16) had the best performance as per their
lowest difference (−0.02 to −0.18 m) and RMSE (+1.5 m) values. However, we reiterate that these products cannot be
used alone (without any supplementary ground observations) to evaluate permafrost dynamics.

The study also examined the associated implications on the generated streamflow at four different gauges and the
partitioning of main water-balance states/fluxes. The performance of streamflow was assessed for each experiment
665   using four error metrics: NSE, NSE$_{Log}$, KGE, and PBias over 2000-2016. The study highlighted a small variability
range in the NSE, KGE and PBias for all stations and experiments; for instance, PBias variability was within 2.5 to
3.1%. Conversely, low flows were more sensitive to perturbations of the LRB setup as NSE$_{Log}$ varied between 0.19-
0.77 at the outlet, highlighting the impact of permafrost status on the low flow regime. The metric-based ranking
showed that Exp. 12 had the best replicability of the observed streamflow at all four stations. For water balance, a
670   small-to-negligible impact on the partitioning of surface fluxes (ET and total runoff) was observed, which was not the
case for the soil storage where the partitioning of total soil water into ice/liquid water contents showed significant
variability on an annual basis. This remark is in line with the changes occurring in the thermal regime, represented by
the changes in the soil temperature. However, these conclusions could be biased by the weak coupling of permafrost
and hydrology in the current generation of LSMs, fundamentally due to the absent/limited lateral migration of heat
675   and water within model tiles and grids.

The assessment of the different experiments highlighted the significant influence of the amount, and extent of
accumulated snow (ZSNL parameter) and the timing of reaching full cover. The reduction of ZSNL value across the
experiments resulted in a drastic decline in PE (PA ~77,000 km^2) and increased the simulated high and low flows at
the LRB outlet by 350 m^3/s (equivalent to 4.1% of the long-term mean daily discharge) and 85 m^3/s (15.5%),
680   respectively. Further, the chosen hydrological year for spin-up had an intermediate impact on the simulated
permafrost; employing a warmer year was shown to lessen PE throughout the basin, leading to a PA reduction of
~30,000 km^2, which on the other hand had negligible impacts on streamflow simulation – an average increase of 50
m^3/s (0.6%) and 10 m^3/s (1.8%) for high and low flows at the basin outlet, respectively. The influence of organic soil
configuration was dependent on the value of ZSNL parameter (*i.e.,* those parameters interacted); mineral soil with

685 organic content with lower ZSNL values (*i.e.,* soil is the major insulator) reduced PA by 20,000 km^2, increased high flows by 500 m^3/s (5.9%) and reduced low flows by 80 m^3/s (14.5%); fully organic soil with higher ZSNLs (*i.e.,* both soil and snow provide insulation) reduced PA by 2,700 km^2, increased high flows by 130 m^3/s (1.5%) and reduced low flows by 50 m^3/s (9.1%). The impact of the depth and vertical distribution of organic soil was modest for the simulated PA/PE and streamflow.

690 Lastly, we investigated the collective experiments' ranking based on the local-scale (ground observations), large-scale (gridded products), and streamflow performance to explore the possibility of having a single H-LSM setup that can concurrently (and adequately) simulate permafrost dynamics and hydrology for the LRB. The best-performing experiments against ground observations had inferior rankings versus gridded products and the worst performance against streamflow. The same is the case for the best-performing experiment against gridded products, which had the

695 worst ranking for ground observations and a moderate ranking for streamflow. Likewise, the best experiment regarding the streamflow simulation had an average ranking against ground observations and gridded permafrost products. These results highlight the complexities and challenges that modellers may encounter when configuring models to simultaneously simulate permafrost and hydrology over large domains characterized by data scarcity. It is noteworthy that the fundamental issue with permafrost studies in large regions is the deep uncertainty in the present/historical

700 status of permafrost, which in turn complicates model development and evaluation. Accounting for such ignorance requires a careful interpretation of model results.

Although the outcomes of this study were specific to the MESH H-LSM and limited to the selected study area, evaluation sites, forcing dataset, gauge stations, gridded datasets, and methods, this study contributes practical information to advance modelling applications, particularly those related to permafrost. The study highlights the

705 importance of considering streamflow when developing and diagnosing an H-LSM for a permafrost-dominated region. Further, the study highlights the discrepancies among the available gridded products of permafrost, which not only stem from the absence of a unifying mapping approach but also from the lack of a well-distributed and representative monitoring network; for example, the LRB permafrost observations are aligned with a specific pipeline project. While current remote sensing technologies cannot address the paucity of permafrost data over large domains, this will

710 hopefully be addressed soon. Finally, future research could be directed toward generalizing the outcomes of this study to other basins with more ground observations and examining the extension of this work to different H-LSMs with various complexities and under different forcing datasets.

Code and data availability statement
The source code for MESH (version r1813) is available from the MESH GitHub page (https://github.com/MESH-

715 Model/MESH-Releases/releases/tag/SA_MESH_1.4%2FSA_MESH_1.4.1813). The sources and access to observed permafrost data (temperature and geology) and climate forcing datasets are provided in **Section 2**.

Author contribution
MSA contributed to the conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing - original draft, and visualization. MEE contributed to the conceptualization, methodology, writing

720   - review & editing. HSW contributed to the conceptualization, methodology, writing - review & editing, supervision, and funding acquisition. SR contributed to the conceptualization, methodology, writing - review & editing, supervision, and funding acquisition.

Competing interests

The authors declare that they have no conflict of interest.

725   **Acknowledgement**

This research was financially supported by the Canada Excellence Research Chair (CERC) Program 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 thank ECCC for supporting the open-source code of MESH software as well as meteorological forcing. The authors would like Joe Melton for sharing the scripts used for temporal

730   disaggregation of climate datasets. The authors also thank Information & Communications Technology (ICT) at the University of Saskatchewan for providing continuous support in using the High-Performance Computing (HPC) Plato and Copernicus research clusters, as well as Compute Canada's Graham cluster.

References

Aalto, J., Karjalainen, O., Hjort, J. and Luoto, M.: Statistical Forecasting of Current and Future Circum-Arctic Ground Temperatures and Active Layer Thickness, Geophys. Res. Lett., 45(10), 4889–4898, doi:10.1029/2018GL078007, 2018.

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

Abdelhamed, M. S., Elshamy, M. E., Wheater, H. S. and Razavi, S.: Hydrologic-land surface modelling of the Canadian sporadic-discontinuous permafrost: initialization and uncertainty propagation, Hydrol. Process., 36(3), 1–22, doi:10.1002/hyp.14509, 2022a.

Abdelhamed, M. S., Elshamy, M. E., Wheater, H. S., and Razavi, S.: Challenges in hydrologic-land surface modelling of permafrost signatures - Impacts of parameterization on model fidelity under the uncertainty of forcing. ESS Open Archive, under review, doi:10.1002/essoar.10510317.1, 2022b.

Abdelhamed, M. S., Elshamy, M. E., Wheater, H. S., and Razavi, S.: The fate of permafrost thawing in the Canadian Subarctic, in preparation.

Alexeev, V. A., Nicolsky, D. J., Romanovsky, V. E. and Lawrence, D. M.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, Geophys. Res. Lett., 34(9), doi:10.1029/2007GL029536, 2007.

Bahrami, A., Goïta, K., Magagi, R., Davison, B., Razavi, S., Elshamy, M. and Princz, D.: Data assimilation of satellite-based terrestrial water storage changes into a hydrology land-surface model, J. Hydrol., (May), 125744, doi:10.1016/j.jhydrol.2020.125744, 2020.

Berry, P., Yassin, F., Belcher, K. and Lindenschmidt, K. E.: An economic assessment of local farm multi-purpose surface water retention systems under future climate uncertainty, Sustain., 9(3), doi:10.3390/su9030456, 2017.

Brown, J., Ferrians Jr, O. J., Heginbottom, J., and Melnikov, E.: Circum-arctic map of permafrost and ground ice conditions, National Snow and Ice Data Center, available at: https://nsidc.org/ data/GGD318/versions/2 (last access: 2 September 2022), 1998.

Budhathoki, S., Rokaya, P. and Lindenschmidt, K.-E.: Improved modelling of a Prairie catchment using a progressive two-stage calibration strategy with in situ soil moisture and streamflow data, Hydrol. Res., 505–520, doi:10.2166/nh.2020.109, 2020.

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

Burn, D. H., Abdul Aziz, O. I. and Pietroniro, A.: A Comparison of Trends in Hydrological Variables for Two Watersheds in the Mackenzie River Basin, Can. Water Resour. J., 29(4), 283–298, doi:10.4296/cwrj283, 2004a.

Burn, D. H., Cunderlik, J. M. and Pietroniro, A.: Hydrological trends and variability in the Liard River basin, Hydrol. Sci. J., 49(1), 53–68, doi:10.1623/hysj.49.1.53.53994, 2004b.

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

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G. and Westermann, S.: An observation-based constraint on permafrost loss as a function of global warming, Nat. Clim. Chang., 7(5), 340–344, doi:10.1038/nclimate3262, 2017.

Connon, R. F., Quinton, W. L., Craig, J. R. and Hayashi, M.: Changing hydrologic connectivity due to permafrost thaw in the lower Liard River valley, NWT, Canada, Hydrol. Process., 28(14), 4163–4178, doi:10.1002/hyp.10206, 2014.

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

Cucchi, M., P. Weedon, G., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H. and Buontempo, C.: WFDE5: Bias-adjusted ERA5 reanalysis data for impact studies, Earth Syst. Sci. Data, 12(3), 2097–2120, doi:10.5194/essd-12-2097-2020, 2020.

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

Deacu, D., Fortin, V., Klyszejko, E., Spence, C. and Blanken, P. D.: Predicting the net basin supply to the Great Lakes with a hydrometeorological model, J. Hydrometeorol., 13(6), 1739–1759, doi:10.1175/JHM-D-11-0151.1, 2012.

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

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

Elshamy, M., Princz, D., Sapriza-Azuri, G., Abdelhamed, M. S., Pietroniro, A., Wheater, H. S. and Razavi, S.: On the configuration and initialization of a large-scale hydrological land surface model to represent permafrost, Hydrol. Earth Syst. Sci., 24(1), 349–379, doi:10.5194/hess-24-349-2020, 2020.

Elshamy, M., Loukili, Y., Pomeroy, J. W., Pietroniro, A., Richard, D. and Princz, D.: Physically based cold regions river flood prediction in data-sparse regions: The Yukon River Basin flow forecasting system, in Journal of Flood Risk Management, pp. 1–19., 2022.

Elshamy, M. E., Pietroniro, A., Pomeroy, J. W., Abdelhamed, M. S., and Wheater, H. S.: Development and validation of a large-scale hydrological land surface model for the Mackenzie River Basin, in preparation.

van Everdingen, R. O.: Multi-language glossary of permafrost and related ground-ice terms, Calgary, Alberta, CANADA T2N 1N4., 1998.

Farquharson, L. M., Romanovsky, V. E., Cable, W. L., Walker, D. A., Kokelj, S. V. and Nicolsky, D.: Climate Change Drives Widespread and Rapid Thermokarst Development in Very Cold Permafrost in the Canadian High Arctic, Geophys. Res. Lett., 46(12), 6681–6689, doi:10.1029/2019GL082187, 2019.

Farr, T. G. and Kobrick, M.: Shuttle radar topography mission produces a wealth of data, Eos (Washington. DC)., 81(48), 583–585, doi:10.1029/EO081i048p00583, 2000.

Gruber, S.: Derivation and analysis of a high-resolution estimate of global permafrost zonation, Cryosphere, 6(1), 221–233, doi:10.5194/tc-6-221-2012, 2012.

Hachem, O., Allard, M. and Duguay, C.: Using the MODIS Land Surface Temperature Product for Mapping Permafrost: An Application to Northern Quebec and Labrador, Canada, Permafr. Periglac. Process., 20, 407–416, doi:10.1002/ppp, 2009.

Haghnegahdar, A., Tolson, B. A., Davison, B., Seglenieks, F. R., Klyszejko, E., Soulis, E. D., Fortin, V. and Matott, L. S.: Calibrating environment Canada's MESH modelling system over the Great Lakes Basin, Atmos. - Ocean, 52(4), 281–293, doi:10.1080/07055900.2014.939131, 2014.

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

Harris, C., Arenson, L. U., Christiansen, H. H., Etzelmüller, B., Frauenfelder, R., Gruber, S., Haeberli, W., Hauck, C., Hölzle, M., Humlum, O., Isaksen, K., Kääb, A., Kern-Lütschg, M. A., Lehning, M., Matsuoka, N., Murton, J. B., Nötzli, J., Phillips, M., Ross, N., Seppälä, M., Springman, S. M. and Vonder Mühll, D.: Permafrost and climate in

Europe: Monitoring and modelling thermal, geomorphological and geotechnical responses, Earth-Science Rev., 92(3–4), 117–171, doi:10.1016/j.earscirev.2008.12.002, 2009.

Hayashi, M., Quinton, W. L., Pietroniro, A. and Gibson, J. J.: Hydrologic functions of wetlands in a discontinuous permafrost basin indicated by isotopic and chemical signatures, J. Hydrol., 296(1–4), 81–97, doi:10.1016/j.jhydrol.2004.03.020, 2004.

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

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

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J. N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146(730), 1999–2049, doi:10.1002/qj.3803, 2020.

Hulley, G. and Hook, S.:MOD21A2 MODIS/Terra Land Surface Temperature/3-Band Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2022-10-17 from https://doi.org/10.5067/MODIS/MOD21A2.006, 2017.

Husain, S. Z., Alavi, N., Bélair, S., Carrera, M., Zhang, S., Fortin, V., Abrahamowicz, M. and Gauthier, N.: The Multibudget Soil, Vegetation, and Snow (SVS) Scheme for Land Surface Parameterization: Offline Warm Season Evaluation, J. Hydrometeorol., 17(8), 2293–2313, doi:10.1175/JHM-D-15-0228.1, 2016.

Jacques, J. M. S. and Sauchyn, D. J.: Increasing winter baseflow and mean annual streamflow from possible permafrost thawing in the Northwest Territories, Canada, Geophys. Res. Lett., 36(1), 1–6, doi:10.1029/2008GL035822, 2009.

Ji, H., Nan, Z., Hu, J., Zhao, Y. and Zhang, Y.: On the Spin-Up Strategy for Spatial Modeling of Permafrost Dynamics: A Case Study on the Qinghai-Tibet Plateau, J. Adv. Model. Earth Syst., 14(3), doi:10.1029/2021ms002750, 2022.

Jones, P. W.: First- and second-order conservative remapping schemes for grids in spherical coordinates, Mon. Weather Rev., 127(9), 2204–2210, doi:10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, 1999.

Kouwen, N., Soulis, R., Seglenieks, F., Bingeman, A. and Davison, B.: An Introduction to WATFLOOD and WATCLASS. [online] Available from: http://www.watflood.ca/, 1993a.

Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J. and Harrington, R. A.: Grouped response units for distributed hydrologic modelling, J. Water Resour. Plan. Manag., 119(3), 289–305, 1993b.

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

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0), Geosci. Model Dev. Discuss., 12(7), 3055–3070, doi:10.5194/gmd-12-3055-2019, 2019.

Lawrence, D. M. and Slater, A. G.: A projection of severe near-surface permafrost degradation during the 21st century, Geophys. Res. Lett., 32(24), 1–5, doi:10.1029/2005GL025080, 2005.

Lawrence, D. M. and Slater, A. G.: Incorporating organic soil into a global climate model, Clim. Dyn., 30(2–3), 145–160, doi:10.1007/s00382-007-0278-1, 2008.

Letts, M. G., Comer, N. T., Roulet, N. T., Skarupa, M. R. and Verseghy, D. L.: Parametrization of peatland hydraulic properties for the Canadian land surface scheme, Atmos. - Ocean, 38(1), 141–160, doi:10.1080/07055900.2000.9649643, 2000.

Louie, P. Y. T., Hogg, W. D., MacKay, M. D., Zhang, X. and Hopkinson, R. F.: The water balance climatology of the Mackenzie basin with reference to the 1994/95 water year, Atmosphere-Ocean, 40(2), 159–180, doi:10.3137/ao.400206, 2002.

Luo, Y., Arnold, J., Allen, P. and Chen, X.: Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China, Hydrol. Earth Syst. Sci., 16(4), 1259–1267, doi:10.5194/hess-16-1259-2012, 2012.

MacDonald, M. K., Pomeroy, J. W. and Pietroniro, A.: Parameterizing redistribution and sublimation of blowing snow for hydrological models: tests in a mountainous subarctic catchment, Hydrol. Process., 23, 2570–2583, doi:DOI: 10.1002/hyp.7356, 2009.

MacKay, M. D., Seglenieks, F., Verseghy, D., Soulis, E. D., Snelgrove, K. R., Walker, A. and Szeto, K.: Modeling Mackenzie basin surface water balance during CAGES with the Canadian Regional Climate Model, J. Hydrometeorol., 4(4), 748–767, doi:10.1175/1525-7541(2003)004<0748:MMBSWB>2.0.CO;2, 2003.

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

Manabe, S.: Climate and the Ocean Circulation 1. The Atmospheric Circulation and the Hydrology of the Earth's Surface, Mon. Weather Rev., 97(1), 739–774, doi:10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2, 1969.

Mekonnen, M. A., Wheater, H. S., Ireson, A. M., Spence, C., Davison, B. and Pietroniro, A.: Towards an improved land surface scheme for prairie landscapes, J. Hydrol., 511, 105–116, doi:10.1016/j.jhydrol.2014.01.020, 2014.

Melton, J. R. and Arora, V. K.: Competition between plant functional types in the Canadian Terrestrial Ecosystem Model (CTEM) v. 2.0, Geosci. Model Dev., 9(1), 323–361, doi:10.5194/gmd-9-323-2016, 2016.

Melton, J. R., Arora, V. K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E., Teckentrup, L. and Melton, J.: CLASSIC v1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM)-Part 1: Model framework and site-level performance, Geosci. Model Dev., doi:10.5194/gmd-2019-329, 2019.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, Trans. ASABE, 50(3), 885–900, doi:10.13031/2013.23153, 2007.

Morse, P. D., Wolfe, S. A., Kokelj, S. V. and Gaanderse, A. J. R.: The Occurrence and Thermal Disequilibrium State of Permafrost in Forest Ecotopes of the Great Slave Region, Northwest Territories, Canada, Permafr. Periglac. Process., 27(2), 145–162, doi:10.1002/ppp.1858, 2016.

Muskett, R. R. and Romanovsky, V. E.: Alaskan Permafrost groundwater storage changes derived from GRACE and ground measurements, Remote Sens., 3(2), 378–397, doi:10.3390/rs3020378, 2011.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., Khomutov, A., Kääb, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T., Yamkhin, J. and Zou, D.: Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale, Earth-Science Rev., 193(March), 299–316, doi:10.1016/j.earscirev.2019.04.023, 2019.

Park, H., Walsh, J., Fedorov, A. N., Sherstiukov, A. B., Iijima, Y. and Ohata, T.: The influence of climate and hydrological variables on opposite anomaly in active-layer thickness between Eurasian and North American watersheds, Cryosph., 7(2), 631–645, doi:10.5194/tc-7-631-2013, 2013.

Pastick, N. J., Jorgenson, M. T., Wylie, B. K., Rose, J. R., Rigge, M. and Walvoord, M. A.: Spatial variability and landscape controls of near-surface permafrost within the Alaskan Yukon River Basin, J. Geophys. Res. Biogeosciences, 119, 1244–1265, doi:doi:10.1002/2013JG002594., 2014.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E., Caldwell, R., Evora, N. and Pellerin, P.: Development of the MESH modelling system for hydrological ensemble forecasting of the

Laurentian Great Lakes at the regional scale, Hydrol. Earth Syst. Sci., 11(4), 1279–1294, doi:10.5194/hess-11-1279-2007, 2007.

Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for climate models, Int. J. Climatol., 23(5), 479–510, doi:10.1002/joc.893, 2003.

Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J. and Carey, S. K.: The cold regions hydrological model : a platform for basing process representation and model structure on physical evidence, Hydrol. Process., 21, 2650–2667, doi:10.1002/hyp, 2007.

Prentice, I. C., Liang, X., Medlyn, B. E. and Wang, Y. P.: Reliable, robust and realistic: The three R's of next-generation land surface modelling, Atmos. Chem. Phys. Discuss., 15(17), 5987–6005, doi:10.5194/acpd-14-24811-2014, 2015.

Quinton, W. L. and Marsh, P.: A conceptual framework for runoff generation in a permafrost environment, Hydrol. Process., 13(16 SPEC. ISS.), 2563–2581, doi:10.1002/(SICI)1099-1085(199911)13:16<2563::AID-HYP942>3.0.CO;2-D, 1999.

Quinton, W. L., Hayashi, M. and Chasmer, L. E.: Peatland hydrology of discontinuous permafrost in the northwest territories: Overview and synthesis, Can. Water Resour. J., 34(4), 311–328, doi:10.4296/cwrj3404311, 2009.

Ran, Y., Li, X., Cheng, G., Che, J., Aalto, J., Karjalainen, O., Hjort, J., Luoto, M., Jin, H., Obu, J., Hori, M., Yu, Q. and Chang, X.: New high-resolution estimates of the permafrost thermal state and hydrothermal conditions over the Northern Hemisphere, Earth Syst. Sci. Data, 14(2), 865–884, doi:10.5194/essd-14-865-2022, 2022.

Razavi, S. and Gupta, H. V.: A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application, Water Resour. Res., 52, 440–455, doi:10.1002/ 2015WR017559, 2016.

Riseborough, D., Shiklomanov, N., Etzelmuller, B., Gruber, S., Marchenko, S., Johansson, M. and Åkerman, H. J.: Recent Advances in Permafrost Modelling, Permafr. Periglac. Process., 19(January), 279–292, doi:10.1002/ppp, 2008.

Romanovsky, V. E., Sazonova, T. S., Balobaev, V. T., Shender, N. I. and Sergueev, D. O.: Past and recent changes in air and permafrost temperatures in eastern Siberia, Glob. Planet. Change, 56(3–4), 399–413, doi:10.1016/j.gloplacha.2006.07.022, 2007.

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

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., Denning, A. S., Mooney, H. A., Nobre, C. A., Sato, N., Field, C. B. and Henderson-Sellers, A.: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, Science (80-.)., 275(5299), 502–509, doi:10.1126/science.275.5299.502, 1997.

Shrestha, R. R., Cannon, A. J., Schnorbus, M. A. and Alford, H.: Climatic controls on future hydrologic changes in a subarctic river basin in Canada, J. Hydrometeorol., 20(9), 1757–1778, doi:10.1175/JHM-D-18-0262.1, 2019.

Smith, M. W. and Riseborough, D. W.: Permafrost Monitoring and Detection of Climate Change, Permafr. Periglac. Process., 7, 301–309, 1996.

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

Smith, S. L., Throop, J. and Lewkowicz, A. G.: Recent changes in climate and permafrost temperatures at forested and polar desert sites in Northern Canada, Can. J. Earth Sci., 49(8), 914–924, doi:10.1139/E2012-019, 2012.

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

Tapley, B. D., Bettadpur, S., Watkins, M. and Reigber, C.: The gravity recovery and climate experiment: Mission overview and early results, Geophys. Res. Lett., 31(9), 1–4, doi:10.1029/2004GL019920, 2004.

Thorne, R.: Uncertainty in the impacts of projected climate change on the hydrology of a subarctic environment: Liard River Basin, Hydrol. Earth Syst. Sci., 15(5), 1483–1492, doi:10.5194/hess-15-1483-2011, 2011.

Verseghy, D.: Canadian Land Surface Scheme for GCMS I. Soil model, Int. J. Climatol., 11, 111–133, doi:10.1002/joc.3370110202, 1991.

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

Verseghy, D.: CLASS – The Canadian land surface scheme (version 3.6) - technical documentation., 2012.

Walvoord, M. A. and Kurylyk, B. L.: Hydrologic Impacts of Thawing Permafrost—A Review, Vadose Zo. J., 15(6), 1–20, doi:10.2136/vzj2016.01.0010, 2016.

Westermann, S., Duguay, C. R., Grosse, G. and Kääb, A.: Remote sensing of permafrost and frozen ground, in Remote Sensing of the Cryosphere, edited by M. Tedesco, pp. 307–344, JohnWiley & Sons, Ltd, New York., 2015.

Wheater, H. S., Pomeroy, J. W., Pietroniro, A., Davison, B., Elshamy, M., Yassin, F., Rokaya, P., Fayad, A., Tesemma, Z., Princz, D., Loukili, Y., DeBeer, C. M., Ireson, A. M., Razavi, S., Lindenschmidt, K., Elshorbagy, A., MacDonald, M., Abdelhamed, M., Haghnegahdar, A. and Bahrami, A.: Advances in modelling large river basins in cold regions with Modélisation Environnementale Communautaire - Surface and Hydrology (MESH), the Canadian hydrological land surface scheme, Hydrol. Process., 36(4), 1–24, doi:10.1002/hyp.14557, 2022.

Woo, M. K.: Permafrost hydrology, Springer Science & Business Media., 2012.

Woo, M. K. and Thorne, R.: Snowmelt contribution to discharge from a large mountainous catchment in subarctic Canada, Hydrol. Process., 20(10), 2129–2139, doi:10.1002/hyp.6205, 2006.

Woo, M. K., Lewkowicz, A. G. and Rouse, W. R.: Response of the canadian permafrost environment to climatic change, Phys. Geogr., 13(4), 287–317, doi:10.1080/02723646.1992.10642459, 1992.

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

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G. and Wheater, H.: Representation and improved parameterization of reservoir operation in hydrological and land-surface models, Hydrol. Earth Syst. Sci., 23(9), 3735–3764, doi:10.5194/hess-23-3735-2019, 2019.

Zhang, T., Frauenfeld, O. W., Serreze, M. C., Etringer, A., Oelke, C., McCreight, J., Barry, R. G., Gilichinsky, D., Yang, D., Ye, H., Ling, F. and Chudinova, S.: Spatial and temporal variability in active layer thickness over the Russian Arctic drainage basin, J. Geophys. Res. D Atmos., 110(16), 1–14, doi:10.1029/2004JD005642, 2005.

Zhao, L., Wu, Q., Marchenko, S. S. and Sharkhuu, N.: Thermal state of permafrost and active layer in central Asia during the international polar year, Permafr. Periglac. Process., 21(2), 198–207, doi:10.1002/ppp.688, 2010.

Zhao, S., Cheng, W., Yuan, Y., Fan, Z., Zhang, J. and Zhou, C.: Global permafrost simulation and prediction from 2010 to 2100 under different climate scenarios, Environ. Model. Softw., 149, 105307, doi:10.1016/j.envsoft.2022.105307, 2022.