



# *Supplement of*

## Future permafrost degradation under climate change in a headwater catchment of central Siberia: quantitative assessment with a mechanistic modelling approach

Thibault Xavier et al.

*Correspondence to:* Laurent Orgogozo (laurent.orgogozo@get.omp.eu)

The copyright of individual parts of the supplement might differ from the article licence.

#### **S1: Estimating soil surface temperature from external conditions**

The CMIP6 climate scenarios used in this work provide access to air temperature and precipitation projections for the next century. However, permaFoam is a solver dedicated to water and heat transfer within the soil, and the boundary conditions at the soil surface must be provided. The soil sur-

- 5 face is separated from the air by a vegetation layer (in Kulingdakan, mainly lichens and mosses), which is covered by a snow layer in winter. In order to carry out numerical simulations to quantify the impact of climate change on the thermal and hydrological regime of Kulingdakan soil, a dedicated procedure has to be set up to estimate the soil surface temperature from external conditions. The methodology should be based on the variables for which in-situ measurements are available:
- 10 A measurement campaign in Kulingdakan between August 2003 and September 2005 provided daily measurements of the soil temperature at different depths: the top of the moss layer, the top of the organic soil horizon (duff), the top of the mineral horizon and 10 cm and 20 cm within the mineral soil horizon.
- The following measurements from the Tura weather station between 1999 and 2014 were used: 15 the daily snow depth, air temperature (including min./max.) and precipitation (undifferentiated rain or snow). The town of Tura is located 5 km away from the Kulingdakan watershed. The air temperature measured at Tura was compared with the temperature at the top of the moss layer during the summer months (June to August), when there is no snow separating the moss layer from the ambient air. The variations of the two signals are similar, with an average difference of 1.9°C between open-place/sub-20 canopy temperature measurement (Zellweger et al., 2019; Haesen et al., 2021). Therefore, the meteorological data from the Tura weather station are used as an approximation of the meteorological conditions in the Kulingdakan watershed.

The proposed empirical transfer function, based on parametrical calibrations, is not directly transferable to other study sites. The methodology is solely aimed at translating the range of possible 25 future climatic conditions into soil surface signals on the slopes of the Kulingdakan watershed, in order to build proper surface boundary conditions for performing the targeted cryohydrogeological simula-

tions. The model output compared to the available data and climate projections can be found in Figures 3 and 4. The soil surface temperature estimator produced a signal with an average error limited to 1.4°C and  $1.6^{\circ}$ C for the North Aspect Slope (NAS) and South Aspect Slope (SAS), respectively (L<sup>1</sup> norm).

30 In the following, we describe the details of this methodology, which is organised in two steps: a model for the evolution of the snow cover and an estimator of the soil surface temperature.

## **S1.1: Construction of local scenarios of future climate changes**

To construct the scenarios of the future air temperature and precipitation until 2100, we use the following approach: first, a synthetic year of the daily air temperature and daily precipitation is comput-35 ed by averaging the measured daily values of each calendar day over the years of available data (1999– 2014), and then we add to this synthetic representation of current climatic conditions the changes in air temperature / precipitation projected for East Siberian region by the CMIP6 models (Iturbide et al., 2021). The multi-annually averaged daily data describing the climatic forcings of the synthetic year representing current climatic conditions are presented in Figure S1.



40

**Figure S1: Daily air temperature and precipitation of a synthetic year representing current climatic conditions (day by day averages of observation data from 1999 to 2014).**

#### **S1.2: Estimation of snowpack evolution**

As empirical models of snowpack evolution, temperature index models make it possible to 45 simulate snowpack dynamics using only a limited number of variables. A more detailed and physicsbased snow model would require the use of additional variables beyond the air temperature and precipitation and is therefore beyond the scope of this work. Since only the air temperature and precipitation data are available for simulating the snowpack of the Kulingdakan watershed, we use a temperature index model. Additionally, only snow depth observations are available in this area, not the snow water 50 equivalent, which is the main output of a temperature index model. Thus, to enable the training of the temperature index model, we estimated the snow water equivalent (SWE) from snow depth measurements; we use a data assimilation approach.

Here, we base our approach on the formulation of Hock (2003) and express the variation of the snow water equivalent mass per day ΔSWE [mm.day<sup>-1</sup>] as a result of the melt *M* [mm.day<sup>-1</sup>] and accu-55 mulation through snow precipitation  $P$  [mm.day<sup>-1</sup>], as shown in Eq. (A.1):

$$
\Delta SWE = P - M \tag{S1}
$$

The accumulation term is estimated using the snow precipitation, which is considered to be the precipitation recorded by the weather station when the air temperature is below  $0^{\circ}$ C. The melt term is zero when the air temperature is below  $0^{\circ}$ C and positive when the air temperature is strictly above  $0^{\circ}$ C. It is estimated using a degree-day factor (DDF) and the air temperature, with  $M=DDF^*T_{air}$ . The DDF needs 60 to be calibrated to represent the observed melt during the thaw season. In the case of Kulingdakan, a high value of the DDF is found to be optimal  $(15 \text{ mm}.day^{-1} \text{K}^{-1})$  due to the abrupt thaw event at the end of the snow season, when the entire snow cover is usually melted within two weeks. Evaporation, sublimation and wind transport effects during the winter season are neglected. Due to the continental climate, strong diurnal temperature variations can be observed at the Tura station, and thaw events may be 65 reported even when the daily mean temperature is negative. Since the extremum temperatures reached each day are available in the weather data, we include them in the model by considering that the air temperature varies during the day from  $T_{min}$  to  $T_{max}$  as a piecewise linear function of time (linear between  $T_{min}$  and  $T_{mean}$ , and between  $T_{mean}$  and  $T_{max}$ ) while respecting the daily mean temperature. Taking

this temperature variation into account, Eq. (S1) is modified so that both melting and accumulation can 70 occur on the same day, leading to Eq. (S2). Precipitation is assumed to be equally distributed throughout the day but is considered as snow only when  $T(t) > 0$ °C:

$$
\Delta SWE = \int_{t=0}^{t=24h} \left( P(t) - M(t) \right) dt \tag{S2}
$$

In order to calibrate this empirical model to the Tura weather station data, the SWE must be estimated from the available snow depth measurements. For this purpose, we use the Multiple Snow Data Assimi-75 lation System (MuSA) of Alonso-Gonzalez et al. (2022). MuSA is an ensemble-based data assimilation toolbox designed to fuse observations with the mechanistic model called the Flexible Snow Model (FSM2; Essery, 2015). The simulations are fed by ERA5 land global reanalysis data (Hersbach et al., 2020), perturbing the precipitation and the temperature in order to fit the output snow depth to local data. The objective SWE is obtained from the posterior mean of the ensembles (Fig. 3a).

80 Considering the evolution of the snowpack over the 12 winters for which data are available (Fig. 3a), the utilised temperature index model predicts the extent of the snow period well, while the maximum snow water equivalent obtained in the snowpack is predicted with an accuracy of 24% based on the L2 error norm (respectively 22% for the L1 norm, 5% minimum error, 37% maximum error). The projection of the snowpack dynamics to the year 2100 (Fig. 3b) shows that the increase in temperature 85 leads to a shortening of the snow season (1 month shorter in SSP5-8.5), while the increase in precipitation leads to a higher accumulation rate in the winter, leading to an increase in the maximum SWE of up to  $26\%$  (+41 mm, SSP5-8.5).

#### **S1.3: Estimation of soil surface temperature**

In order to estimate the soil surface temperature, an empirical method based on multiple regression is 90 implemented. The soil surface temperature is estimated by two different approaches, depending on whether snow covers the moss layer during winter (hereafter referred to as the 'cold season') or during the rest of the year (referred to as the 'warm season').

During the warm season, the soil surface is only separated from the ambient air by the moss layer. Moreover, liquid precipitation enhances the heat transfer from external air and soil surface tempera-

95 ture variations that follow external air variations with a significantly shorter response time than in the presence of snow. Therefore, for the warm season, we use a direct estimator of the soil surface temperature in the form of first-order multiple regression based on the air temperature and precipitation.

During the cold season, both the snow layer and the moss layer isolate the soil surface, and conduction effects with a slower time response occur when heat fluxes are affected by the snow thickness. 100 In order to mimic this behaviour while maintaining a simple, empirical and data-based approach, we used a multiple regression method based on the time derivative of the soil surface temperature as a function of the SWE and the temperature difference with the air temperature. We used a first-order regression for each variable, since higher-order tests did not produce a better soil surface temperature estimation. The cold season and warm season models are activated below or above the 0°C air temperature 105 threshold. In order to ensure the smoothness of the signal, an interpolation between the two models was used when the temperature was close to  $0^{\circ}C$  ( $\pm 0.5^{\circ}C$ ). Figure 4 presents the comparison between the measurements and the estimates obtained by the described empirical procedure for the two years of soil temperature data acquisition.

Using only two full years of data makes the statistical assessment of the empirical estimation 110 accuracy difficult. However, the soil surface temperature obtained with the present model consistently follows the dynamics observed in the field, while the extremum temperature remains in a similar range for both the NAS ([-18.0°C; 13.0°C] for field measurements, [-18.6°C; 12.5°C] for model output) and the SAS ([-15.9°C; 14.5°C] for field measurements, [-17.9°C; 12.5°C] for model output).



**Figure S2: (a) Schematic diagram of the numerical domain geometry and main variables' boundary conditions. (b) Representation of the first soil column metre for the NAS and SAS.** 

- 120 The calculation setup follows the procedure used in previous work (Orgogozo et al., 2019) and is described below. The procedure consists of eight calculations, corresponding to the four SSP scenarios applied to the two slopes of the Kulingdakan watershed. The calculations are performed between 2014 and 2100 based on the climate scenario and estimated soil surface temperature (see Supplementary Material A and Figs. 3 and 4). An additional 30 years of calculations were carried out with a repetition
- 125 of the final conditions corresponding to the year 2100 in order to assess the thermohydric equilibrium state of the first few metres of soil after the simulated climate change sequence. The calculations were performed using OpenFOAM version v2212 and the permaFoam solver (January 2023 version).

#### **S2.1: Geometry and mesh information**

130 The geometry used to represent a slope of the Kulingdakan watershed is a 2D parallelogram (Fig. S2) covering 2.5 km in the x-direction and with 10 m of thickness (z-direction), with an inclination of 18.5%. The mesh is constructed using a uniform mesh size regarding the x-direction and a nonuniform cell thickness with a geometrical growth, with a finer mesh close to the soil surface where steep fronts occur. The mesh used for climate scenarios is  $2048x256$  ( $\sim$ 5.2 $*10<sup>5</sup>$ ) cells, with a cell length of

135 1.22 m and a cell thickness ranging from 2.53 mm close to the surface to 16.5 cm at the bottom of the domain.

This mesh has been chosen following a convergence study constituted by three refinement levels: 1024x128, 2048x256 and 4096x512. Current conditions, constructed as the mean year of the years between 1999 and 2014, are used for the mesh convergence study. The criterion for assessing numerical 140 convergence is active layer thickness evolution. This preliminary study showed that the 1024x128 mesh produces an unrealistic active layer thickness estimation (e.g. 12 cm for the NAS). The 4096x512 mesh, on the other hand, results in a small difference in the active layer thickness compared to the 2048x256 case (for the NAS, 65 cm and 64 cm, respectively; for the SAS, 98 cm and 100 cm, respectively): the differences in the computed active layer thickness between the medium and large mesh cases were 145 small, with maximum differences of 2.2% for the NAS and 1.3% for the SAS. Meanwhile, the use of the large mesh involves a heavy computational cost for the centennial scale, in this case requiring approximately four times more CPU hours compared to the medium (2048x256 cells) mesh case. Then, the medium mesh of  $2048x256 \left( \sim 5.2*10^5 \right)$  cells is adopted. The resolution is 1.2 m laterally and between 0.25 cm (top) and 16.5 cm (bottom) vertically, since we use a vertically graded mesh in order to 150 lower the computation time.

## **S2.2: Thermohydraulic properties**

The domain is represented by a porous medium constituted by either an organic or mineral soil matrix, filled with water (whether in the liquid or solid state) and air. The thermal properties used for the simulation are described in Table S1, while hydraulic properties are listed in Table S2; they are the 155 same as those used in Orgogozo et al. (2019). Note that the organic layer thickness is 11.6 cm for the NAS and 7.7 cm for the SAS (see Fig. S2).

	Organic ma-	Mineral ma-			
	trix	trix	Liquid water	Ice	Air
Heat capacity $[J.m^{-3}.K^{-1}]$	$2.51 \times 10^6$	$1.92 \times 10^6$	$4.18 \times 10^6$	$1.90 \times 10^6$	$1.23 \times 10^3$
Thermal conductivity $[J.m^{-1}.s^{-1}.K^{-1}]$	0.25	2.9	0.6	2.14	0.026
Latent heat of fusion ice / liquid water $[J.m^{-3}]$			$3.34 \times 10^8$		

**Table S1: Thermal properties used to represent soil in permaFoam simulations conducted in this work.**

160



**Table S2: Hydraulic properties used to represent soil in permaFoam simulations conducted in this work.**

## **S2.3: Boundary conditions, sink term and initial conditions**

To build the SSP scenarios, a representative year for meteorological forcings under current cli-165 matic conditions is built from weather data measured between 1999 and 2014. A multi-annual average is obtained for the year-round evolution of precipitation and air temperature along seasons by averaging these data for each day of the available measurement years. The air temperature and precipitation annual trends provided by SSP scenarios are then applied to this virtual, averaged year representative of current climatic conditions in order to build the atmospheric conditions up to the year 2100. The boundary 170 conditions for the simulations of the cryohydrogeological changes under climate change are briefly presented in Figure S2 and summarised below.

The top boundary conditions for the soil surface temperature are constructed from the atmospheric forcings, as described in the 'Material and Methods' section and in Supplementary Material A. Water fluxes at top boundary conditions are directly taken as the liquid precipitation (considered liquid

175 when  $T_{\text{air}} > 0$ °C), with switching boundary conditions for dealing with a water-saturated soil surface (Orgogozo et al., 2019). All other variables are subjected to zero-gradient boundary conditions.

For the upslope side of the domain, a zero-gradient boundary condition is applied to all the variables, except for the water pressure, which is described with the noRainFlux conditions, which impose a zero flux (Orgogozo et al., 2023).

180 For the downslope side of the domain, a zero-gradient boundary condition is applied to all the variables, except for the water pressure, which is described with static pressure head.

At the bottom of the domain, geothermal flux is imposed for the thermal equation (0.018W.m<sup>-2</sup>, Duchkov et al., 1997\*), as mentioned in section 2.3 of the main text. Water pressure is described by a noRainFlux boundary condition (no water flux), while the boundary conditions for other variables are 185 zero-gradient.

As seen in Eq. (1), water that is taken up by roots is represented by a sink term in the Richards equation. This volumetric term is calculated from the potential evapotranspiration (PET) rate distributed over the root layer thickness. The PET is calculated using the Hamon formula based on the air temperature (Hamon, 1963), which has previously been used in studies of forested boreal areas (Frolking, 190 1997). As mentioned earlier, the root layer thickness is 22 cm on the NAS (10 cm into the mineral horizon) and 68 cm on the SAS (60 cm into the mineral horizon).

The initial conditions are obtained by a spin-up procedure. The first guess corresponds to the state of the active layers extracted from previous calculations under early  $21<sup>st</sup>$  century conditions for the same site (2003–2012; see Orgogozo et al., 2019). In the absence of observations of the moisture at

195 depth, an initial value of 0.335 is chosen, resulting from the annual mean of the water content averaged on the north and south slope active layers obtained in these previous calculations. This state is then used as the seed for a spin-up performed by cycling simulations considering the representative year for mete-

orological forcings under current climatic conditions until convergence, i.e. after 10 years. The convergence criterion is defined by evaluating the change in the active layer thickness from one year to the

200 next. At the end of the ten-year spin-up, a variation of less than 0.2% on both slopes is achieved. Beyond the depth of the active layer, the soil water content remains practically equal to the value chosen at initialisation due to the very slow water flow under permanently frozen conditions.

\*Duchkov AD, Sokolova LS, Balobaev VT, Devyatkin VN, Kononov VI, Lysak SV.. Heat flow and 205 geothermal field in Siberia. Geologiya / Geofizika. 1997;38(11):1716-1729.

## **S2.4: High-performance computing methodology**

Despite the use of the 2D assumption and the excellent parallel performance of the permaFoam solver, carrying out a mechanistic permafrost dynamics simulation at the scale of the catchment over almost a century remains a particular computational challenge. This section outlines some elements of 210 the methodology and computing means used to meet this challenge.

The calculations are carried out on the IRENE JOLIOT-CURIE supercomputer operated by the French Alternative Energies and Atomic Energy Commission (CEA). This supercomputer offers, among other partitions, an AMD partition equipped with AMD Rome (Epyc) processors, with 64 computational cores each. OpenFOAM is used in this work only on CPUs with the MPI communication 215 protocol. Since the mesh domain is composed by 525k cells for each slope (sufficient for convergence, see Supplementary Material B), the number of MPI processes can be kept relatively low, with the use of 256 MPI processes for each case treated here. Like most of the fluid mechanics solvers based on finite volume discretisation, permaFoam exhibits a memory-bound nature in most of its operations, with a low arithmetic intensity. Therefore, we adapted the use of the supercomputing infrastructure by depopulat-220 ing the compute nodes by a factor of two (using only 32 computational cores out of the 64 cores available on each processor), thus largely broadening the bandwidth available for each MPI process. This operation reduced the computation time by almost a factor of two without requiring significantly higher CPU hour costs.

As a whole, the computational campaign required the use of 1.8 million CPU hours, generated 225 almost 2 TB of raw data and produced ~80k inodes, with a restitution time of approximately one month for each simulation (i.e. for one scenario and for one slope). Regarding the energy consumption associated with the performed simulations, we propose the following estimate. On the IRENE-ROME supercomputer, the power consumption is estimated to be 5.02734 W/core (personal communication from the operating team). Thus, the energy consumption of our 1.8-million-CPU-hour simulation campaign 230 could be roughly estimated to be 9 MWh. To compare this consumption with that of a typical fourperson household in the European Union, we propose to consider the final energy consumption in households (all end uses, including water heating, space heating and cooling, cooking and electrical appliances) available in the Eurostat database for the year 2021 (1584677 terajoules\*, equal to 440188055 MWh). Then, this total consumption may be divided by the census population in 2021 235 (445649015 inhabitants\*\*) to obtain an estimate of the average energy consumption per person and per year in the European Union (0.987 MWh/person/year). According to this estimate, the energy consumed by our numerical modelling process is equivalent to the energy required to power a typical four-person household for about 27 months.

\* https://doi.org/10.2908/NRG\_D\_HHQ

240 \*\* https://doi.org/10.2908/CENS\_21AG

#### **S3: Changes in the main variables according to the four climate projections**

The tables in the section of the Supplementary Material compile the changes in the main variables between present conditions and 2100 for the four climate scenarios considered in this paper (SSP1- 2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5) for the NAS (Table S3) and SAS (Table S4).



245 **Table S3: Summary of values obtained for current conditions and the four climate projections for 2100 used in this study for the NAS.**



**Table S4: Summary of values obtained for current conditions and the four climate projections for 2100 used in this study for the SAS.**

### **S4: A view of seasonal change of liquid water available for vegetation uptake**



250

**Figure S3: Liquid water available for vegetation uptake in years 2014, 2050 and 2100 according to the different climate projections. Computed by integrating the liquid water volumetric fraction along the surface layer above the 0°C isotherm.**

The multi-annual changes of the quantity of liquid water available for the root water uptake along a hy-255 drological cycle is assessed in Figure S3, showing the seasonal variation of the daily value of the inte-

gral of liquid water volumetric fraction above the 0°C isotherm for three different date : 2014, 2050 and 2100. As expected the stronger the climate warming is, the stronger changes in liquid water availability are. The contrast between the two slopes is also enhanced for the pathways with the most forcing, with a maximum of water availability in South aspected slope.