CMUG Extension Phase 2 Deliverable
Reference: D2.0a v2.1: Interim Progress Report for OWP5.1
Due date: March 2026
Submission date: April 2026
Version: 2.1.0

Climate Modelling User Group (CMUG)

Extension Phase 2 Deliverable D2.0a: OWP5.1 Machine Learning to advance climate model evaluation and process understanding – interim progress report

Centres providing input: DLR

Version	Date	Status
2.0	20 April 2026	first version
2.1	01 May 2026	submitted to ESA

Table of Contents

Summary ... **3**

1 OWP5.1.3 – Causal model evaluation for cloud regimes and land cover types ... **3**

1.1 Quantifying the causal effect of cloud controlling factors on marine stratocumulus clouds ... *4*

1.2 ICON-XPP data ... *4*

2 OWP5.1.4 Evaluation of CMIP6 models with the ESMValTool ... **5**

2.1 Evaluation of snow ... *6*

2.2 Evaluation of permafrost ... *8*

References ... **11**

CMUG Extension Phase 2 Deliverable

Reference:	**D2.0a v2.1: Interim Progress Report for OWP5.1**
Due date:	**March 2026**
Submission date:	**April 2026**
Version:	**2.1.0**

WP5.1: Machine Learning to advance climate model evaluation and process understanding – interim progress report v2

Summary

This report summarizes the progress made within the optional tasks of WP5.1 until March 2026. The main aim of WP5.1 is to develop and apply machine learning (ML) techniques for advanced climate model evaluation and process understanding with ESA CCI data. Progress is reported for the two optional tasks OWP5.1.3 "Causal model evaluation for cloud regimes and land cover types" and OWP5.1.4 "Evaluation of CMIP6 models with the ESMValTool" (OWP5.1.4).

Within OWP5.1.3, the method of causal analysis and quantification of the role of selected cloud controlling factors on cloud properties developed in WP5.1.2 (see section 1.1) is applied to output from the ICON-XPP model. This serves as a proof of concept for a process-oriented model evaluation of for selected cloud regimes

In OWP5.1.4, snow and permafrost from CMIP6 models are evaluated with CCI data using the Earth System Model Evaluation Tool (ESMValTool) further developed and extended within the CMUG task 4. Here, we introduce the selection of diagnostics implemented for these evaluation tasks and show first examples for the historical simulation from the CMIP6 model MIROC6.

1 OWP5.1.3 – Causal model evaluation for cloud regimes and land cover types

OWP5.1.3 is a continuation of WP5.1.2, in which ESA CCI data are used to (1) better understand the causal drivers for specific cloud regimes and land cover types and (2) to enhance regime-oriented causal model evaluation with causal discovery. Causal discovery belongs to unsupervised machine learning and aims to discover and quantify causal interdependencies and dynamical links inside a system, such as the Earth's climate (Runge et al., 2015; 2019). This approach goes beyond correlation-based measures by systematically excluding common driver effects and indirect links. Based on dataset-specific fingerprints, this approach offers a novel and objective pathway for process-oriented model evaluation (Nowack et al., 2020).

In OWP5.1.3, the same method is applied to output from the global climate model ICON-XPP. The resulting causal networks are then compared to the ones obtained from the observations in order to evaluate the models and estimate the sensitivity of cloud properties to the cloud controlling factors. As not all relevant cloud controlling factors are available from satellite observations, we will also include some additional datasets for controlling factors such as lower tropospheric stability (LTS) from ERA5. Specifically, the following two science questions are addressed:

- What are the causal relationships and networks in the real world as derived from ESA CCI data?
- Can the ICON-XPP climate model reproduce the causal networks between cloud properties and cloud controlling factors as derived from ESA CCI data for different cloud regimes? And for different land cover types?

1.1 Quantifying the causal effect of cloud controlling factors on marine stratocumulus clouds

Our incomplete understanding of clouds and their role in cloud–climate feedbacks leads to large uncertainties in climate projections. By applying causal discovery, an unsupervised machine learning method, and afterward causal effect estimation (supervised method) to daily satellite and reanalysis data, we systematically analyze causal links between cloud properties and selected cloud-controlling factors with the aim of identifying and quantifying the sensitivity of cloud properties to different factors. This will contribute to improving our understanding of how clouds react to changes in controlling factors as a consequence of climate change. Here, we focus on marine stratocumulus (Sc) clouds off the coast of South America. Specifically, we analyse dynamical and thermodynamical cloud-controlling factors to quantify the causal effects on macrophysical properties of marine Sc clouds. In agreement with previous studies, we find that sea surface temperature, lower tropospheric stability, surface sensible heat flux, and 10-m wind speed are the main drivers influencing these clouds. While the causal links between these factors and the cloud properties—total cloud cover, total cloud water path, and cloud optical depth—show similar behaviour, the cloud effective radius remains largely unexplained, suggesting that the background aerosol might play an important role. In contrast, the cloud-top pressure is influenced by all cloud-controlling factors investigated, except the surface sensible heat flux. Application of this approach to climate model output could be a step toward a process-oriented evaluation of the sensitivity of simulated clouds to climate change.

Results of this study are now published in Bock et al., 2026.

1.2 ICON-XPP data

To apply the causal inference framework to the ICON-XPP model output, it was first necessary to identify the appropriate output variables. These included the cloud controlling factors and five custom variables (total cloud fraction, total cloud water path (liquid + ice), cloud optical thickness, cloud particle effective radius, and cloud-top pressure). The objective is to execute a five-year simulation with daily output storage.

In order to apply the causal network derived from the ESACCI and reanalysis data (Bock et al., 2026), we need output from ICON-XPP of the cloud controlling factors (see Figure 1) and the five cloud properties total cloud fraction, total cloud water path (liquid + ice), cloud optical thickness, cloud particle effective radius, and cloud-top pressure. Since cloud optical thickness, cloud particle effective radius, and cloud-top pressure are not included in the default model output, these variables must be derived and integrated into the ICON-XPP source code.

Our main task at the moment is to modify the ICON-XPP code to derive these additional cloud properties. As a next step, a 5-year ICON-XPP simulation is performed to continue the causal inference study on daily output.

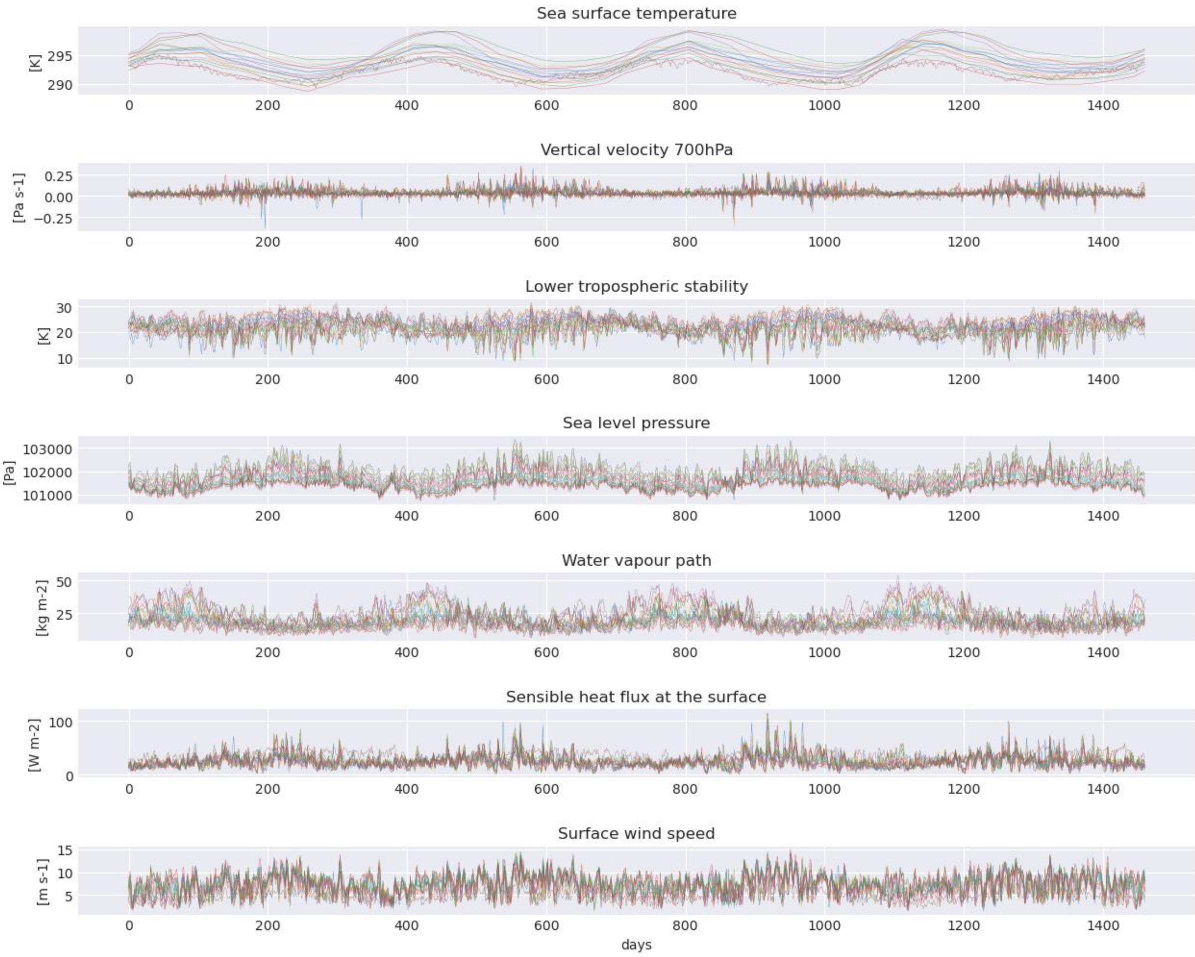

Figure 1 Timeseries of all cloud-controlling variables from a 5-year ICON-XPP simulation averaged over 5°x5° boxes over the Southeast Pacific (75°–95°W, 10°–30°S)

2 OWP5.1.4 Evaluation of CMIP6 models with the ESMValTool

Within OWP5.1.4, the CCIs SNOW and PERMAFROST implemented into ESMValTool as part of Task 4 will be applied to the CMIP6 model ensemble to address the science question:

- How well do CMIP6 models reproduce the observed variability and (if detectable in the observations) trends in the historical record of snow cover and permafrost?

Initially, the diagnostics for analysis of the CMIP6 ensemble have been defined and tested with a single CMIP6 model (MIROC6). In the following, examples of these diagnostics for this CMIP6 model are shown to illustrate the diagnostics. The snow and permafrost CCI datasets used for comparison with the CMIP6 models are summarized in

Table 1.

Table 1 CCI datasets used by ESMValTool for comparison with the CMIP6 models.

Dataset	*Variables*	*Time range*
ESACCI-SNOW (daily L3C, merged, version 2.0)	swe (snow water equivalent), scfg (snow cover fraction)	1979-2019 (swe), 1982-2018 (scfg)
ESACCI-PERMAFROST (NH + SH, version 5.0)	alt (active layer thickness), gtd (permafrost ground temperature), pfr (permafrost extent)	1997-2023

2.1 Evaluation of snow

Seasonal snow cover has important implications for the energy, moisture and gas fluxes between the land surface and the atmosphere. Because of the typically high albedo of snow, snow cover can cause an important feedback effect in a warming climate and is highly relevant to climate change.

Here, the focus of the CMIP6 model evaluation is on the snow water equivalent (swe), which is the depth of liquid water produced if a snowpack is melted and represents the total water content stored.

As an overview, polar stereographic maps of the snow water equivalent are used showing the geographical distribution and a proxy for the total mass of snow per grid cell. In the example shown in Figure 2, the 5-year average swe values for the Northern Hemisphere (NH) winter (December-January-February) are shown. Comparison of the MIROC6 model results (historical experiment) with the CCI data shows that the model is able to capture the basic features of the geographical pattern and amplitude but tends to produce snow reaching too far south.

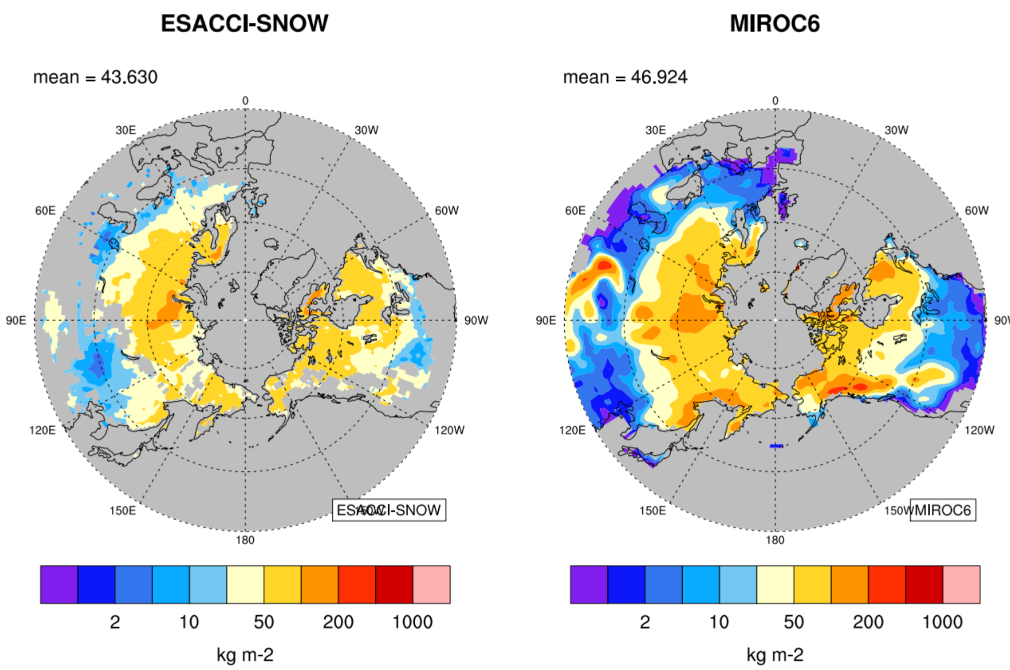

Figure 2 Comparison of the 5-year average (2009-2014) snow water equivalent (in kg m^{-2}) for December-January-February from ESACCI-SNOW (left) with the historical simulation from the CMIP6 model MIROC6 (right). Glaciated areas have been masked out.

Figure 3 shows a time series from 1979-2014 (overlapping period of CCI data and output from the MIROC6 historical experiment) of the total (sum) snow water equivalent in the Northern Hemisphere. As the model shows fewer snow-free areas than the CCI dataset (see Figure 2), the total NH swe amount calculated by the model is about 30-50% higher than in the CCI data. This diagnostic will also be used to estimate the interannual variability and possible trends.

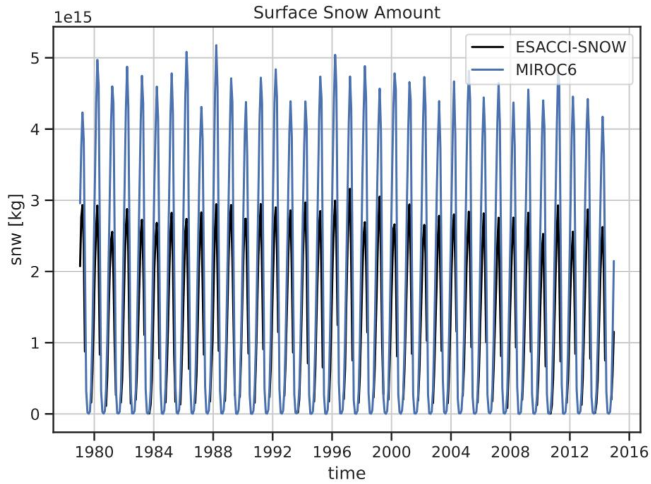

Figure 3 Time series of the global total snow water equivalent (in 10^{15} kg) in the Northern Hemisphere from ESACCI-SNOW (black line) in comparison with the historical simulation from the CMIP6 model MIROC6 (blue line).

The temporal and spatial variability of daily swe values can be investigated with probability density functions (PDFs) for selected regions. For better comparability, the model and CCI data have been regridded to the same 1°x1° latitude-longitude grid and glaciated areas are masked out. For the calculation of the PDF, only values larger than 10 kg m^{-2} are shown to focus on areas with a significant snow cover. In the example shown in Figure 4, the MIROC6 model shows a higher frequency of small swe values in the range 10-20 kg m^{-2} as well as an overestimation of swe values larger than about 60-70 kg m^{-2}. This is consistent with (1) small swe values reaching too far south in the model compared with CCI data (see Figure 2) and (2) the total sum of swe in the Northern Hemisphere being overestimated by 30-50% (see Figure 3).

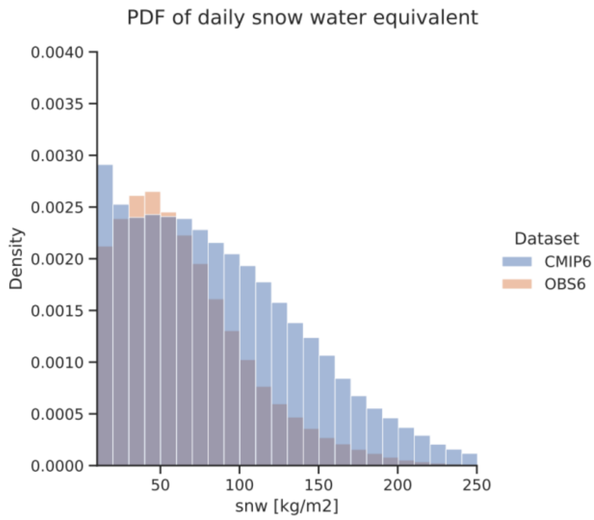

Figure 4 Probability density function (PDF) of the daily snow water equivalent (in kg m^{-2}) in the Northern Hemisphere over the years 2009-2014 from ESACCI-SNOW (orange, "OBS6") and the historical simulation from the CMIP6 model MIROC6 (blue, "CMIP6"). Data have been regridded to a regular 1°x1° latitude-longitude grid, glaciated areas have been masked out.

Another important and climate-relevant aspect of the snow cover is the seasonal cycle. As an example, Figure 5 shows the average seasonal cycle of the NH mean swe (in kg m^{-2}). The MIROC6 model is able to reproduce the basic features including the timing of the seasonal cycle with the maximum in March but tends to slightly overestimate the amplitude of the seasonal cycle by about 10-15%.

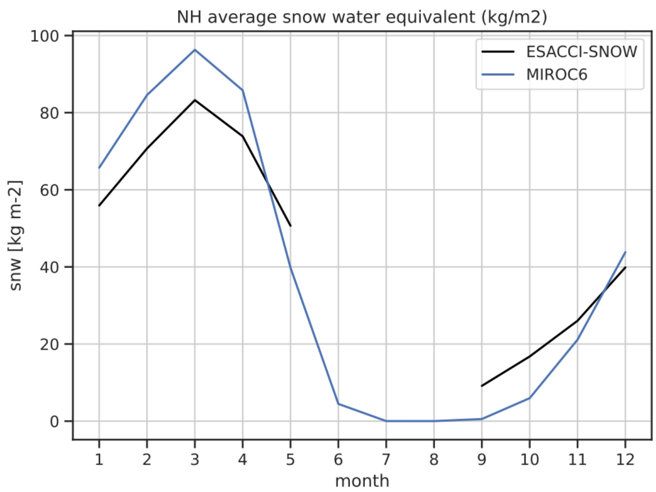

Figure 5 5-year mean (2009-2014) seasonal cycle of the average Northern Hemisphere snow water equivalent (in kg m^{-2}) from ESACCI-SNOW (black line) and the historical simulation from the CMIP6 model MIROC6 (blue line). Glaciated areas have been masked out.

2.2 Evaluation of permafrost

Permafrost is a particular thermal state below the land surface. Here, permafrost is defined as ground at or below the freezing point of water for two or more years. Changes in permafrost interact with ecosystems and climate on various spatial and temporal scales. For climate, the release of the greenhouse gases carbon dioxide and methane by thawing permafrost is of particular importance as this feedback can significantly accelerate global warming. Here, evaluation of the CMIP6 models focuses in particular on the geographical distribution and temporal changes in the permafrost extent.

As an example, Figure 6 shows a comparison of the 10-year average (2005-2014) permafrost extent in % from CCI data and the MIROC6 model. Because of the way permafrost is derived from the modelled soil temperatures following Burke et al. (2020), a model grid cell can be either fully covered by permafrost (100%) or not at all (0%). A finer distinction as possible in the CCI data is not possible at the moment. The MIROC6 model is able to reproduce the geographical distribution of the permafrost reasonably well. The overestimation of the average value is at least partly caused by the limitation to 100% and 0% permafrost extent in the model grid cells.

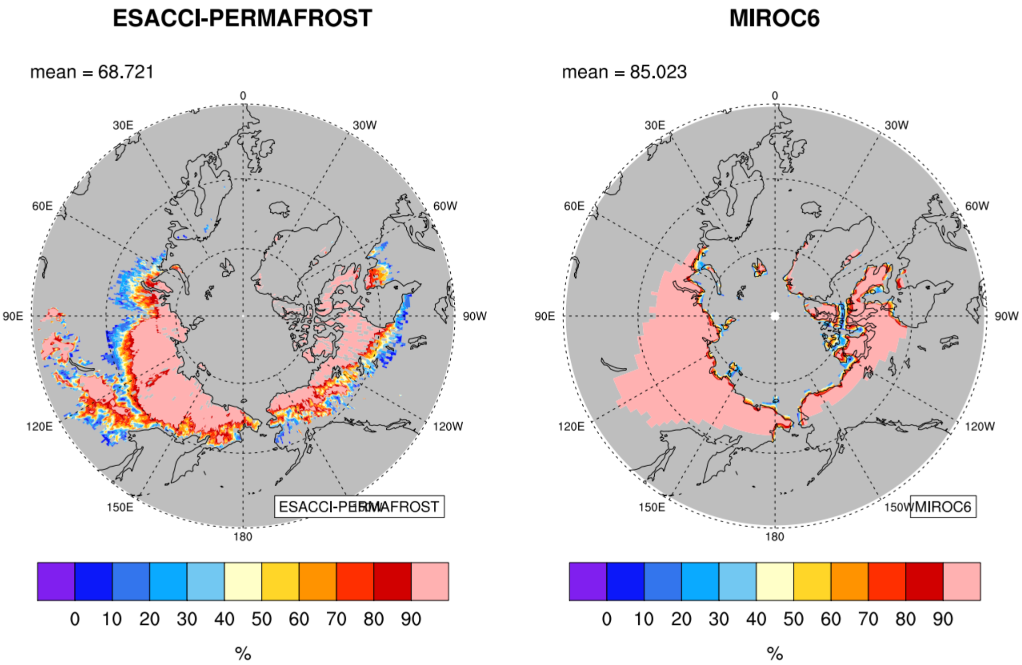

Figure 6 Comparison of the 10-year average (2005-2014) permafrost extent (in %) from ESACCI-PERMAFROST (left) with the historical simulation from the CMIP6 model MIROC6 (right). Glaciated areas have been masked out.

The temporal evolution of the permafrost extent can be investigated by calculating time series of the NH total (sum) permafrost extent (Figure 7). The CCI data shows a higher variability than MIROC6 alongside with a faster decrease over time.

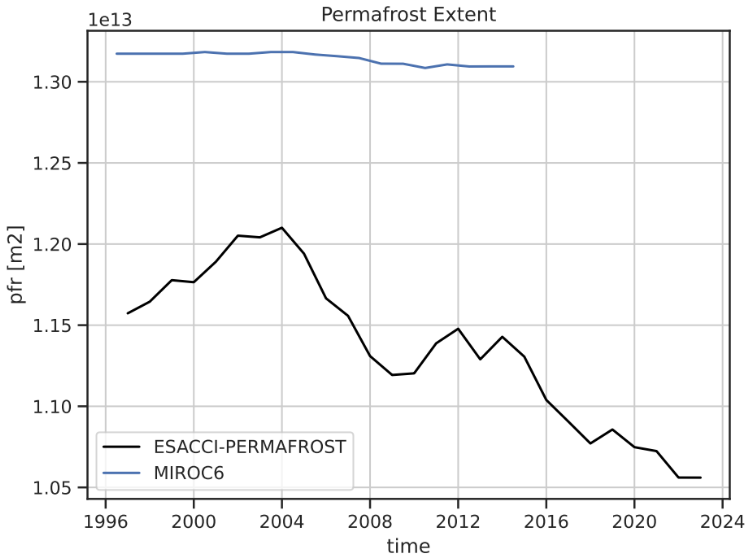

Figure 7 Time series of the Northern Hemisphere total permafrost extent (in 10^{13} m^2) from ESACCI-PERMAFROST (black line) and the CMIP6 model MIROC6 (blue line). Glaciated areas have been masked out.

As an additional diagnostic, a comparison of the multi-year annual average soil temperature at a depth of 2m has been implemented. An example of this is shown in Figure 8.

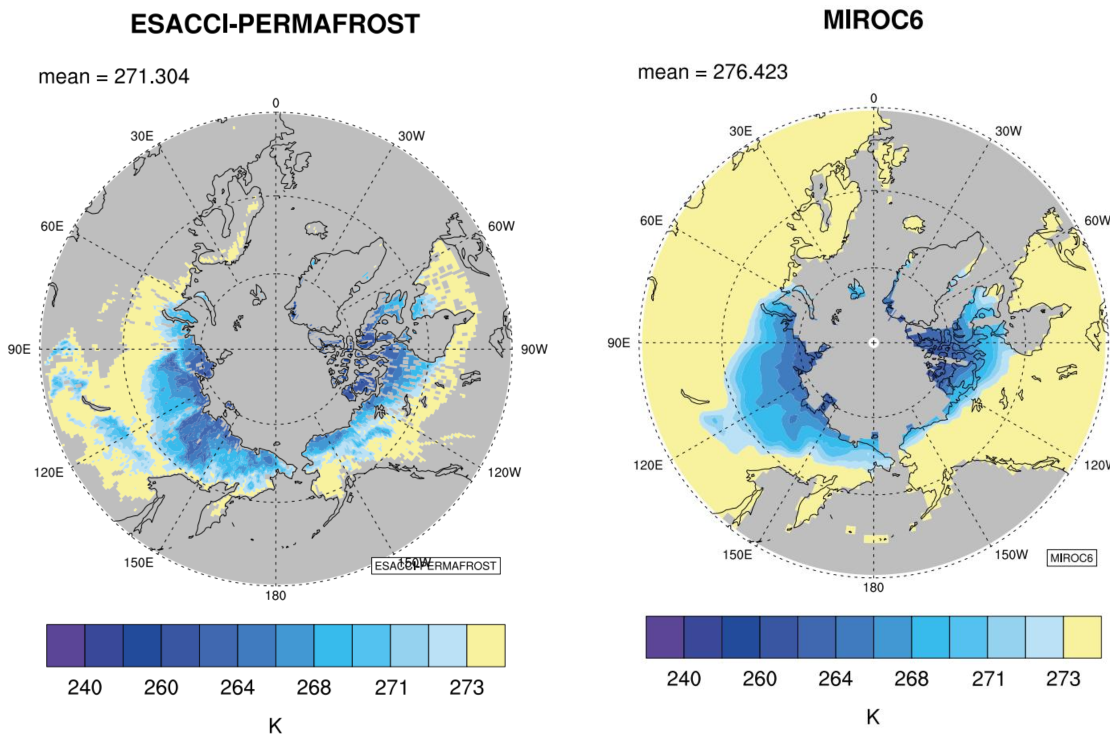

Figure 8 10-year average of the soil temperature at a depth of 2m in permafrost locations from ESACCI-PERMAFROST (left) in comparison with the historical simulation from the CMIP6 model MICRO6 (right). Glaciated areas have been masked out.

References

Bock, L., A. Lauer, and J. Runge, 2026: Quantifying the Causal Effect of Cloud-Controlling Factors on Marine Stratocumulus Clouds. *J. Atmos. Sci.*, **83**, 255–264, https://doi.org/10.1175/JAS-D-25-0153.1.

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

Nowack, P., Runge, J., Eyring, V., and Haigh, J. D.: Causal networks for climate model evaluation and constrained projections, *Nat Commun.*, **11**, 1415, https://doi.org/10.1038/s41467-020-15195-y, 2020.

Runge, J., Petoukhov, V., Donges, J. *et al.*: Identifying causal gateways and mediators in complex spatio-temporal systems, *Nat Commun.*, **6,** 8502, https://doi.org/10.1038/ncomms9502, 2015.

Runge, J., Bathiany, S., Bollt, E. et al.: Inferring causation from time series in Earth system sciences, *Nat. Commun.*, **10**, 2553, https://doi.org/10.1038/s41467-019-10105-3, 2019.