Reply to referee 3

We appreciate the time and effort spent on the revision of the manuscript "Simulating the permafrost thermal regime in the Northwestern Antarctic Peninsula from 1950 to 2100". The comments and suggestions were very well received, and we have subsequently revised and improved the text and work presented. In the following, we provide point-by point replies to all issues raised. The reviewer comment appears in bold font, our replies in normal font and changes to the text which will be implemented in the revised version of the manuscript are in italics.

On behalf of all authors,

Joana Baptista

<u>Reply to the major comments:</u>

Point 1: I understand the practical reason for selecting relatively deep boreholes in bedrock and thereby reducing site heterogeneity. Nevertheless, this choice also limits the environments represented by the study. I would like the authors to explain how far these five sites can reasonably represent the ice-free areas of the northwestern Antarctic Peninsula. In particular, are unconsolidated marine terraces, glacial deposits, colluvium, or ice-rich sediments largely absent from the selected network? If so, this should be stated clearly before the point simulations are extended to coastal landscapes or used to discuss thermokarst, slope instability, and infrastructure exposure.

Reply: With the exception of the very shallow boreholes in Deception Island (not analysed here), all PERMANTAR boreholes are on bedrock, and the network does not include unconsolidated marine terraces, glacial deposits, colluvium, or ice-rich sediments. This was a design option aiming at standardizing the setting and assessing the regional variability, rather than the effect of site specific conditions. While our boreholes represent a significant part of the regional setting, the results have to be interpreted considering that they are for bedrock and cannot be directly applied to unconsolidated materials. Regarding the characteristics of observatories, a description is provided both on the methods and appendix A also with photographs.

Concerning the statement that the simulations were run for bedrock sites, we clarified it in the conclusions section.

Conclusions *"Moreover, the application of reanalysis data and CMIP6 projections from ACCESS-CM2 as forcing, allowed modelling the evolution of the recent past and future permafrost conditions on bedrock, providing an overview of the long-term ground thermal regime in the Northwestern Antarctic Peninsula.*

(...)

Our results for bedrock sites indicate that the future projected warming across the Northwestern Antarctic Peninsula, independently of the CMIP6 scenario will very likely lead to the thawing of permafrost from coastal areas and low elevation sites. Although extrapolation to non-bedrock settings is subject to limitations, our results provide an important indication of regional warming trends that are likely to influence permafrost conditions across a range of ice-free environments. Such substantial changes in the ground thermal regime may modify surface and subsurface hydrology, biogeochemical fluxes and geomorphological processes, with implications for the sensitive terrestrial (and potentially also to nearshore) ecosystems. They may also increase the vulnerability of human infrastructure (research stations, roads, and airstrips) to degradation. Similarly to what is recorded in comparable regions in the Arctic, in non-consolidated deposits, thermokarst and landslide activity will likely increase. In the steep rocky slopes, so typical of the coastal areas of the Northwestern Antarctic Peninsula, as bedrock warms and thaw propagates along joints, rockfalls and landslides may become more frequent, increasing hazards in areas that are experiencing growing tourist visitation during the austral summer."

Point 2: It's difficult to reconstruct the observational basis for each simulation from the information distributed between the main text and the appendices. A single site-characteristics table would be helpful. It should include logger type and accuracy, monitored depths, observation periods and gaps, slope aspect and inclination, surface and subsurface material, vegetation or organic cover, and available information on local snow conditions.

It is also unclear which air-temperature observations were used to correct ERA5 at Amsler, where the local air-temperature logger appears to have been installed only in 2023.

Reply: We complemented the Table 1 with information on the site description and observational period. Information on the dataloggers will remain available on the appendix as their inclusion in Table 1 would compromise the readability due to the various modifications made over time. Additionally, the origin of the air temperature used to corrected ERA5 is also explained on the section Appendix A: Description of PERMANTAR monitoring sites.

"Air-temperature observations from 2016 to 2017 are available from a meteorological station that worked intermittently from 2012 to 2017."

Point 3: As I understand the workflow, at least some subsurface parameters and the snowfall multipliers were selected by comparing simulations with the same ground-temperature records that were later used to report model performance. If it is correct, those statistics describe goodness of fit after calibration rather than performance against independent data, and they are likely to be optimistic.

Please separate calibration and validation periods, or apply a suitable temporal cross-validation procedure. Performance metrics should be presented independently for both periods. The duration and number of observations used at each site and depth should also be reported.

Reply: We acknowledge that the separation of the calibration and evaluation periods can improve the interpretation of the procedure. To address this, the year 2021 was now used for calibration of the subsurface parameters and snowfall multiplier, while the period from the beginning of the observational records through 2020 was used for independent model evaluation. We have also specified in the revised manuscript (Table 1 and Section *3.2.2 Model calibration and evaluation*), for each observatory, the calibration and evaluation periods, the observations used, and the depths considered. Separate tables reporting calibration and evaluation statistics have been added accordingly (Section *4.1 Sensitivity analysis to subsurface parameterization and validation of the surface energy model*).

Point 4: I could not determine the full parameter space explored during calibration. Please report the tested ranges, increments, and selection criterion for mineral, organic, water, and ice fractions and for thermal conductivity. At line 155, the manuscript states that the sum of water and ice contents remains constant, but the value at which it is held constant is not defined.

It would be useful to distinguish parameters supported by direct observations from parameters inferred through fitting. Were porosity, fracture density, water or ice content, and related properties documented during drilling or measured on cores? If not, the adopted values should be treated as uncertain parameters rather than physically validated site properties. A sensitivity or ensemble analysis is required to assess parameter equifinality.

Reply: With the changes implemented following previous comments, the methods were rewritten explaining the parameters adjusted during the calibration phase (mineral, water and ice content and snowfall multiplication factor). Thermal conductivity, as stated was defined based on the measurements performed on the rock cores collected from the boreholes. The sum of water and ice was derived from the volumetric content. Fracture density was observed during the core analysis.

Point 5: What is the basis for applying the same geothermal heat flux of 50 mW m^{-2} at all five sites, and how deep is the lower model boundary? Is this value supported by regional measurements or previous studies? Since the principal results concern temperature at 20 m, sensitivity to geothermal heat flux, domain depth and initial temperature profile should be evaluated.

Reply: We re-run the simulations using geothermal heat flux values from An et al. (2015) for each observatory. These is now described on the methods section. The results showed minor differences.

"Measured ground surface temperature from the PERMANTAR monitoring sites was used as upper boundary input while a geothermal heat flux value obtained from (An et al., 2015) *and applied as the lower boundary. Geothermal heat flux values for King Sejong, Papagal and SKO Meteo were 73 mWm^{-2}, for Cierva were 61 mWm^{-2}, and for Amsler were 58 mWm^{-2}.*"

Point 6: The boreholes are only 4–15 m deep, whereas the main projections concern MAGT at 20 m. At several sites, the calculated temperature at 20 m is therefore well outside the observed depth interval. The associated uncertainty should be explicitly quantified.

Positive MAGT at 20 m should not automatically be equated with complete permafrost loss. Please define the diagnostic criterion used to determine the presence or absence of permafrost. The manuscript should distinguish between: positive MAGT at a specified depth; loss of permafrost at that depth; active-layer deepening; and complete disappearance of the permafrost body.

Regional statements about widespread permafrost loss should be revised accordingly.

Reply: We aimed to estimate the MAGT at a depth of zero annual amplitude and in most observatories, this is not reached at 10 m depth. The estimates at 10 m depth complement the analysis and allowed a comparison to the records in some of the observatories. We extended the explanation in the methods.

"In the simulation's output we provide estimates of ground temperatures for a profile of 20 m depth, with vertical resolution varying with depth: 0.05 m from the surface to 1 m, 0.1 m from 1 to 2 m, 0.2 m from 2 to 5 m, and 0.5 m below 5 m depth. These are used to calculate the mean annual ground surface temperature (MAGST) and MAGT at 10 and 20 m depth. MAGT at 10 m depth provides an overview of near-surface thermal changes and their potential implications for surface and subsurface hydrology, biogeochemical fluxes, and geomorphological processes. MAGT at 20 m depth provides insight into long-term ground temperature conditions near the depth of zero annual amplitude."

Sentences establishing a direct relationship between positive MAGT and permafrost degradation were removed.

Point 7: Snow is interpreted as a major control on differences among sites, but the snowfall multiplier was calibrated indirectly against ground temperature. Are any field observations of snow depth, snow-cover duration, or snow water equivalent available? If not, independent remote-sensing or photographic evidence should be considered.

The assumption of a constant multiplier of 0.3 or 0.5 throughout the future simulation also requires justification. Future changes in precipitation phase, wind redistribution, and snow duration may alter the relationship between large-scale precipitation and local accumulation. At minimum, the sensitivity of the projections to the snowfall multiplier should be shown.

Reply: The snowfall multiplier was calibrated by comparing estimated and measured ground surface temperatures and assessing the resulting snow insulating effect. The selected value therefore corresponds to the multiplier that provided the best agreement between modelled and observed ground surface temperatures.

Section *3.2.2 Model calibration and evaluation:*

"In the second phase of the calibration, the best configuration of volumetric content was used in a ground class (GROUND_freeW_seb_snow, Westermann et al., 2023) in which the phase change of water again occurs at 0 ºC and the water and ice contents remain constant. The surface energy balance scheme is applied to simulate energy exchange processes between the atmosphere and the model's first grid cell (Baptista et al. 2025; Westermann et al. 2023). For this, the simulations were forced with bias-corrected ERA5 reanalysis (see Sect. 3.3.1)

(…)

When the ground is snow-covered, the ground class is stacked below a snow class (SNOW_crocus_bucketW_seb, (Westermann et al., 2023) *that represents the seasonal snow microphysics scheme following the Crocus model* (Vionnet et al., 2012; Zweigel et al., 2021). *The snow model represents transient snow density changes due to compaction and wind drift, as well as meltwater infiltration and refreezing. Due to the micro topography at each monitoring site and to the spatial resolution of ERA5 variables, an over- or underestimation of snow depth can be produced by the model. Previous studies reported that ERA5 often overestimates snow cover in coastal Antarctic areas* (Obu et al., 2020)*, due to unrepresented topography and ice-free vs glacier surfaces and wind redistribution effects. Therefore, ERA5-derived snowfall Ps was corrected for each site using a snowfall multiplication factor which allows a phenomenological increase or reduction of the simulated snow depth based on the intensity of the insulating effect and duration of the snow cover* (Baptista et al., 2025; Martin et al., 2019). *As for the first phase, the validation of the surface energy model and snowfall multiplication factor were performed through the comparison between the estimated and measured ground temperatures for 2021 at 4 depths and the analysis of the four statistical measures (bias, r, p-value, MAE and RMSE)."*

Section *4.1 Sensitivity analysis to subsurface parameterization and validation of the surface energy model:*

"In the sensitivity analysis of surface parameters, on the second phase of calibration when the simulations were forced ERA5 reanalysis data, snowfall multiplication factor values were determined by comparing measured and estimated ground temperatures at a depth of 0.4 m (Table 5). Initial simulations using ERA5 snowfall data resulted in an overestimation of snow accumulation resulting in excess insulation. Consequently, snow multiplication fractions were reduced across all observatories, ranging from 30% to 50% of the original ERA5 snowfall input (Table 5)."

Snow depth measurements are not available at the PERMANTAR observatories. Remote-sensing approaches were previously explored within work conducted for the PERMANTAR network. However, they present substantial limitations in the Northwestern Antarctic Peninsula. Multispectral imagery is frequently affected by persistent cloud cover, resulting in limited temporal coverage. SAR data provide information on wet snow being their applicability largely restricted to the summer season, when most of the study areas are snow free. Furthermore, the spatial resolution of available satellite products would introduce considerable uncertainty at the observatory scale, where microtopography and wind redistribution strongly control snow distribution and, consequently, its insulating effect.

Concerning the future projections, as stated in the manuscript, we applied the model configuration defined through the sensitivity analysis, while acknowledging the limitations associated with its application under future climatic conditions.

Section: *3.2.2. Model calibration and evaluation*

"The selected configuration was subsequently applied in the long-term simulation of past and future evolution of permafrost temperatures, forced by CMIP6 model and ERA5. However, its application under future climatic conditions involves uncertainty since CMIP6-projected climate in the Antarctic Peninsula may differ from the historical conditions used for calibration. Therefore, the model configuration should be understood as the best available representation of the present-day site conditions rather than as a fully constrained description of future conditions."

Point 8: The phrase "adequate data to force the CCM" should be defined more precisely. Please state which required variables, temporal resolutions, ensemble members, and experiments were available for each CMIP6 model.

ACCESS-CM2 is selected mainly because its MAAT trend and variability resemble ERA5 at Papagal. The evaluation should include all five sites and the forcing variables most relevant to the surface energy balance, especially precipitation, radiation, and wind.

Projections based on one GCM cannot represent the full climate-model uncertainty.

Reply: To clarify, a description of the variables required, and their temporal resolution was added to the manuscript.

Section *3.3.2 CMIP6 scenarios and models:*

"The INM-CM4-8, the IPSL-CM6A-LR and the ACCESS-CM2 models from CMIP6 were evaluated, as they provide the required monthly variables to force the CCM: (a) near surface air temperature, (b) surface downwelling longwave radiation, (c) surface downwelling shortwave radiation, (d) precipitation, (e) near surface specific humidity, (f) near surface wind speed, (g) toa incident shortwave radiation, and (h) surface air pressure."

Detailing the variables, temporal resolutions, ensemble members, and experiments available for all CMIP6 models considered during the screening process is not feasible considering the existence of least 58 CMIP6 models.

Regarding the evaluation of ACCESS-CM2, Papagal was selected because it is representative of the regional climatic conditions of the study area and provides a suitable observational record for model evaluation. Observations of precipitation, radiation, and wind are not available, and extending the evaluation would not necessarily result in a more robust model selection and could introduce inconsistencies in the comparison.

"To evaluate the ability of the downscaled CMIP6 models to represent climate conditions in the northwestern Antarctic Peninsula, we compared the simulated MAATs at the Papagal observatory location with ERA5 reanalysis data for the reference period 1950–2020 (Fig. 2 and Table 3). Papagal observatory was selected as the reference site because it is located in an exposed location that is representative of regional climate conditions and provides a long-term, high-quality observational record."

We acknowledge that projections based on a single GCM do not capture the full range of climate-model uncertainty. The objective of this study, however, was to investigate the sensitivity of permafrost using a climate forcing that best reproduces the observed temperature characteristics at the study area. A section acknowledging the limitation associated to the forcing has been added to the discussion.

Section *5.1 Evaluation of the forcing data:*

"In the present study, following the evaluation of each CMIP6 model against the bias-corrected ERA5 air temperature for the period 1950-2020, ACCESS-CM2 was ultimately selected as the most suitable forcing model, as it showed the best overall agreement with ERA5, combining a realistic long-term temperature trend with a good representation of interannual variability. Future studies should simulate permafrost evolution for an ensemble of different CMIP models to account for the considerable inter-model spread in future climate conditions."

Point 9: I found the construction of the submonthly forcing difficult to follow. When a historical ERA5 month is sampled as the carrier, are all meteorological variables taken together from the same month and year, or are variables sampled separately? The former is important for retaining covariance among temperature, humidity, precipitation, wind, and radiation.

The random carrier sequence is itself a source of uncertainty. I would expect different admissible sequences to produce some spread in annual ground temperatures and in the estimated timing of threshold crossings. Several realisations should be generated and propagated through the analysis, with the random seed and sampling procedure documented.

Reply: The same month was used for all forcing variables to conserve the covariance. The sentence was rewritten for clarity. Additionally, in the discussion a new section was added referring the limitations associated with the forcing data.

"m_i^{CMIP} *is equal to* M_i^{CMIP} *during the overlap period, while it is set to the running mean of 10-year periods for the remainder of the time series, so that short-term fluctuations can be separated from multi-annual trends (which are accounted for by the first three terms on the right-hand side). To conserve the covariance between the different model forcing variables, they were all derived from the same month. With this procedure, we ensure that the corrected CMIP6 data features the same monthly mean and standard deviation as the ERA5 data for the overlap period. As "carrier" for the synthesized time series, we randomly sample individual months from the ERA5 time series for each month of the year, which provides somewhat realistic weather conditions on below-monthly timescale. Finally, the carrier data are again corrected on the basis of monthly averages with the corrected CMIP6 time series by adding the offset."*

Point 10: I could not find a description of the statistical method used to estimate temperature trends or the uncertainty reported after the ± sign. Please state the regression method, confidence level and significance criterion.

The four historical periods appear to have been selected visually. Either apply an objective change-point or segmented-regression method, or describe these intervals explicitly as qualitative periods. Decadal rates calculated for the short 2015–2022 interval should be interpreted cautiously.

Reply: In the methods section, an explanation of the statistical method used to calculate temperature trends and uncertainty was added. Regarding the intervals, they are now described as qualitative periods.

Section *3.4 Simulation Long-term simulation of the past and future evolution of permafrost temperatures:*

"The analysis of past permafrost temperature evolution focuses on the period 1950-2020, whereas the analysis of future projections begins in 2020. The temperature trends were estimated though the calculation of a least-squares linear regression. The associated uncertainty was obtained from the standard error of the slope."

Point 11: I agree that elevation, snow persistence, wind exposure, sea-ice decline, and changes in the Amundsen Sea Low are plausible explanations for the simulated contrasts. However, the present experiments do not isolate their individual effects. Unless additional factorial simulations or atmospheric indices are analysed, the language should be changed from causal statements to cautious interpretations such as "may be associated with" or "is consistent with".

Reply: Corrected.

Example in Section *5.3 Future evolution of permafrost temperature (2020 - 2100):*

"*In contrast to the SSP1-2.6 scenario, SSP2-4.5 displays a distinct spatial pattern in MAAT warming rates, with increased warming at the more southerly observatories which may be related to the sea ice loss on the Bellingshausen Sea and consequent deepening of the ASL that strengthens north-northwesterly flow towards the Antarctic Peninsula, favoring advection of warm, maritime air and raising surface air temperatures, especially during the winter* (Abram et al., 2010; Clem et al., 2017; Hosking et al., 2013; Turner et al., 2013)*"*

Point 12: Please cross-check the abstract, Results, Discussion, Table 5, and Conclusions. There are inconsistencies in the reported years when MAGT becomes positive and in the use of "warming" versus "warming rate". For example, values such as 0.59, 2.59, and 5.05 °C appear to represent total changes by 2100 in some places but are described as rates elsewhere. The statement that positive MAGT occurs at most sites by 2045 under SSP5-8.5 is also not consistent with the site-specific results. The title of Sect. 4.3 refers to active-layer thickness, but active-layer results are not presented.

Reply: The inconsistencies were corrected as the warming rates now refer as decadal.

<u>Reply to the minor comments:</u>

At first use, spell out CryoGrid Community Model (CCM), and then use the abbreviation consistently.

Reply: Corrected.

The manuscript uses "Recommended Concentration Pathways"; this should be corrected.

Reply: Corrected.

Please standardise units and notation, including °C decade^{-1}, W m^{-1} K^{-1}, and m a.s.l.

Reply: Corrected.

TTOP needs a precise definition, including how it is diagnosed from the model output and how it is interpreted at sites without confirmed permafrost within the borehole.

Reply: TTOP was removed from the analysis.

Replace "biochemical fluxes" with "biogeochemical fluxes".

Reply: Corrected.

Increase the font and panel size in Figs. 2, 3, 5, and D1. Several time series and legends are difficult to read.

Reply: Corrected.

References

Abram, N. J., Thomas, E. R., McConnell, J. R., Mulvaney, R., Bracegirdle, T. J., Sime, L. C., and Aristarain, A. J.: Ice core evidence for a 20th century decline of sea ice in the Bellingshausen Sea, Antarctica, J. Geophys. Res. Atmos., 115, 2010JD014644, https://doi.org/10.1029/2010JD014644, 2010.

An, M., Wiens, D. A., Zhao, Y., Feng, M., Nyblade, A., Kanao, M., Li, Y., Maggi, A., and Lévêque, J.: Temperature, lithosphere-asthenosphere boundary, and heat flux beneath the Antarctic Plate inferred from seismic velocities, J. Geophys. Res. Solid Earth, 120, 8720–8742, https://doi.org/10.1002/2015JB011917, 2015.

Baptista, J., Brito Guapo Teles Vieira, G., Manuel De Carvalho Soares Correia, A., Lee, H., and Westermann, S.: Modelling the evolution of permafrost temperatures and active layer thickness in King George Island, Antarctica, since 1950, The Cryosphere, 19, 3459–3476, https://doi.org/10.5194/tc-19-3459-2025, 2025.

Clem, K. R., Renwick, J. A., and McGregor, J.: Large-Scale Forcing of the Amundsen Sea Low and Its Influence on Sea Ice and West Antarctic Temperature, J. Clim., 30, 8405–8424, https://doi.org/10.1175/JCLI-D-16-0891.1, 2017.

Hosking, J. S., Orr, A., Marshall, G. J., Turner, J., and Phillips, T.: The Influence of the Amundsen–Bellingshausen Seas Low on the Climate of West Antarctica and Its Representation in Coupled Climate Model Simulations, J. Clim., 26, 6633–6648, https://doi.org/10.1175/JCLI-D-12-00813.1, 2013.

Martin, L., Nitzbon, J., Aas, K. S., Etzelmüller, B., Kristiansen, H., and Westermann, S.: Stability Conditions of Peat Plateaus and Palsas in Northern Norway, J. Geophys. Res. Earth Surf., 124, 705–719, https://doi.org/10.1029/2018JF004945, 2019.

Obu, J., Westermann, S., Vieira, G., Abramov, A., Balks, M., Bartsch, A., Hrbáček, F., Kääb, A., and Ramos, M.: Pan-Antarctic map of near-surface permafrost temperatures at 1 km^2 scale, The Cryosphere, https://doi.org/10.5194/tc-2019-148, 2020.

Turner, J., Phillips, T., Hosking, J. S., Marshall, G. J., and Orr, A.: The Amundsen Sea low, Int. J. Climatol., 33, 1818–1829, https://doi.org/10.1002/joc.3558, 2013.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geosci. Model Dev., 5, 773–791, https://doi.org/10.5194/gmd-5-773-2012, 2012.

Westermann, S., Ingeman-Nielsen, T., Scheer, J., Aalstad, K., Aga, J., Chaudhary, N., Etzelmüller, B., Filhol, S., Kääb, A., Renette, C., Schmidt, L. S., Schuler, T. V., Zweigel, R. B., Martin, L., Morard, S., Ben-Asher, M., Angelopoulos, M., Boike, J., Groenke, B., Miesner, F., Nitzbon, J., Overduin, P., Stuenzi, S. M., and Langer, M.: The CryoGrid community model (version 1.0) – a multi-physics toolbox for climate-driven simulations in the terrestrial cryosphere, Geosci. Model Dev., 16, 2607–2647, https://doi.org/10.5194/gmd-16-2607-2023, 2023.

Zweigel, R. B., Westermann, S., Nitzbon, J., Langer, M., Boike, J., Etzelmüller, B., and Vikhamar Schuler, T.: Simulating Snow Redistribution and its Effect on Ground Surface Temperature at a High-Arctic Site on Svalbard, J. Geophys. Res. Earth Surf., 126, e2020JF005673, https://doi.org/10.1029/2020JF005673, 2021.