Response to Reviewer #1

Manuscript: Modeling the Distribution of Mountain Permafrost in Chile

Alexander Brenning et al.

March 31, 2026

1 General response

We thank Anonymous Referee #1 for the careful and constructive review of our manuscript. The comments helped us clarify several aspects of the modeling framework, its interpretation, and the presentation of the results. In particular, we revised the manuscript to better emphasize the relationship of the present approach to earlier studies (Boeckli et al., 2012; Azócar et al., 2017), clarified the role of empirical calibration in the interpretation of climatic predictors (MAAT), expanded the discussion of model interpretation and threshold choice, and quantified uncertainties of the Permafrost Favorability Index (PFI). We also improved the description of the response variable, the use of borehole observations, and the classification procedure for rock glaciers.

In the detailed responses below, **RC** denotes *Reviewer Comment* and **AR** denotes *Authors' Response*. Changes made to the manuscript are summarized after each response where relevant.

2 General comments

> **RC1.1:** *The modelling framework here closely follows the Boeckli et al. (2012) method as applied to Chile by Azócar et al. (2017), with relatively modest refinements. I think the paper would benefit from being more clear about this and instead focusing on what was learned from scaling the approach across 38 degrees of latitude. The suppressed 'cold anomaly' at 37–44° S is a case in point — that's an interesting result that deserves more attention.*

AR: We thank the reviewer for this constructive comment. We agree that the methodological framework of this study builds primarily on the approach developed by Azócar et al. (2017), which itself adapted the statistical framework originally proposed by Boeckli et al. (2012). The main novelty of the present study therefore lies less in methodological innovation and more in applying this approach across the full latitudinal extent of mainland Chile (~18–56° S), which comes with its own challenges.

Following the reviewer's suggestion, we have revised the Introduction and Discussion to make this clearer. In particular, we now highlight how the shape-constrained GAM helped suppress an otherwise implausible "cold anomaly" predicted in south-central Chile (37–44° S) when unconstrained smoothing functions were used.

The region in question is characterized by comparatively low mountain elevations and a transition from periglacial-dominated to glacier-dominated environments. As a result, rock glaciers are sparse and the available permafrost evidence is limited. Without constraints, the model would produce an unrealistic local cooling pattern in this data-poor region. A monotonicity constraint prevented such artifacts and improved the geomorphological plausibility of the predictions. In addition, we now explicitly highlight south-central Chile (37–44° S) as a region where the scarcity of permafrost observations leads to greater model uncertainty.

Changes in manuscript: We revised the Introduction to clarify the relationship between the present modeling framework and earlier studies (Boeckli et al., 2012; Azócar et al., 2017). In addition, the discussion of the suppressed cold anomaly in south-central Chile has been expanded in Section 5.2 to highlight how the shape-constrained GAM improves model robustness when extrapolating across large latitudinal gradients. We now note that future sampling and monitoring efforts should prioritize such data-poor regions in order to better constrain model predictions and improve the representation of permafrost conditions across the full latitudinal extent of the Andes.

> **RC1.2a:** *The PFI approach assumes that rock glacier activity status reflects present-day permafrost conditions as characterised by current MAAT. But rock glaciers integrate climate signals over centuries to millennia...*

AR: We thank the reviewer for highlighting this relevant conceptual point. We fully agree that rock glaciers and permafrost integrate climatic conditions over long time scales and therefore may not be in equilibrium with present-day climate.

However, the statistical model used in this study is empirically calibrated, meaning that the relationship between predictors and permafrost evidence is estimated directly from the data. Any systematic offset between the MAAT reference period (1979–2019) and the climatic conditions that governed the formation or long-term persistence of rock glaciers is therefore implicitly captured by the fitted predictor–response relationships. In other words, the model does *not* assume that present-day MAAT represents the equilibrium temperature threshold for permafrost formation; rather, it estimates the statistical association between MAAT and the observed distribution of permafrost indicators.

In addition, the model includes latitude as a predictor, which allows the statistical relationships to vary along the climatic gradient spanning ~38° of latitude. This captures spatially varying offsets between present-day climate and the long-term climatic conditions reflected by geomorphological indicators. In practice, latitude can absorb part of the large-scale climatic variability associated with differences in seasonality, snow regimes, and long-term climatic history across the Andes.

Finally, our validation using borehole data (Table 3) demonstrates that our model, along with our recommended interpretation of PFI, is well calibrated and not systematically off in one direction or another.

Changes in manuscript: We agree that rock glaciers may persist under climatic conditions warmer than those under which they formed, which is partly addressed through the aforementioned adjustments. This is already addressed in the manuscript (Introduction: L45-48), and we add a note to Section 3 (Methods: Model adjustments) highlighting the mentioned empirical, latitudinally variable adjustment.

> **RC1.2b:** *An intact rock glacier sitting at +1 °C MAAT today may have formed during the Little Ice Age or earlier and persist due to thermal inertia. The model labels it*

> *'permafrost present' and associates it with current warm temperatures – which could push predictions too warm at marginal temperatures. Azócar et al. (2017) noted this: they found substantial proportions of intact rock glaciers at positive MAAT, particularly in the southern watersheds (31–32° S).*

AR: We agree that intact rock glaciers may persist under climatic conditions warmer than those under which they formed, owing to the long response times of such ice-rich landforms.

In the present study this issue is addressed in two ways. First, following Azócar et al. (2017), we used rock glacier root-zone locations rather than toe locations as indicators of permafrost conditions, because these better represent the climatic conditions under which permafrost is preserved. Second, where only toe locations were available, we applied the altitudinal offset correction of 89 m proposed by Azócar et al. (2017) to approximate root-zone elevations.

More generally, the statistical model is empirically calibrated, meaning that the relationships between predictors and permafrost indicators are estimated directly from the observations rather than interpreted as equilibrium temperature thresholds. Systematic offsets between present-day MAAT and the climatic conditions reflected by long-lived geomorphological indicators are therefore absorbed in the fitted predictor–response relationships (see detailed response to Comment 1.2a, including reference to borehole-based validation).

Changes in manuscript: We added a sentence in the *Model adjustments* subsection (in Section 3 on Methods) clarifying that the empirically calibrated model implicitly captures potential offsets between the MAAT reference period and the longer time scales represented by geomorphological indicators, and that these adjustments may vary along the latitudinal climatic gradient.

> **RC1.2c:** *The CHELSA data represent a 1979–2019 mean, during which the study region warmed on the order of a degree – comparable to the altitudinal offset applied to correct for rock glacier creep. So the spatial displacement is carefully corrected, but a temporal displacement of similar or larger magnitude doesn't seem to be addressed.*

AR: We thank the reviewer for raising this point. We agree that the CHELSA MAAT data represent recent climatic conditions (1979–2019) and that warming during this period may introduce a potential mismatch between present-day climate predictors and the longer time scales represented by geomorphological indicators such as rock glaciers.

However, it is important to note that the altitudinal offset correction of 89 m applied in the model does not represent a climatic adjustment but rather a geomorphological correction accounting for the downslope displacement of rock glacier toes relative to their root zones (Azócar et al., 2017). The purpose of this adjustment is therefore to achieve consistency with other inventories used in this study, for which we were able to extract root-zone locations and elevations. This reason is already given in L184-185.

Changes in manuscript: See Comment 1.2a, and L184-185 of the initial submission.

> **RC1.2d:** *I'm curious why a partial remedy that's already cited in the paper wasn't explored: InSAR-derived rock glacier kinematics from Cusicanqui et al. (2025) could distinguish truly active rock glaciers from inactive ones. Even a discussion of whether restricting the training set to rock glaciers with confirmed present-day movement would sharpen the response variable would be useful.*

AR: We thank the reviewer for this helpful suggestion. Remote sensing approaches based on InSAR-derived kinematics indeed provide valuable information on the present-day activity of rock glaciers and represent an important recent development in rock glacier research. The analysis conducted in the present study was completed prior to the publication of Cusicanqui et al. (2025) and before the release of the RGIK guidelines for rock glacier inventories (Brardinoni et al., 2026), which provide important recommendations for standardized mapping and classification.

Incorporating kinematic information derived from InSAR would require a dedicated processing workflow and consistent temporal coverage across the entire ~38° latitudinal extent of the study area. Such an analysis was beyond the scope of the present project, which relied primarily on geomorphological interpretation of high-resolution imagery and existing inventories to compile a consistent dataset of rock glacier indicators.

We agree that integrating remotely sensed kinematic information represents a promising direction for improving the classification of rock glacier activity and refining permafrost distribution models. We therefore added a short discussion of this possibility and cite both Cusicanqui et al. (2025) and the recently published RGIK guidelines.

Changes in manuscript: We added references to Cusicanqui et al. (2025) and to the RGIK guidelines for rock glacier inventories (Brardinoni et al., 2026) and briefly discuss the potential of InSAR-based kinematic analyses for improving the classification of rock glacier activity in future studies (Section 5.1).

> **RC1.3a:** *The model is trained on rock glacier inventories plus 238 in-situ observations, and then validated against 80 borehole sites. Are these 80 boreholes a subset of the 238 in-situ observations, or are they entirely independent? This isn't clear to me from the text. If there's overlap, that would affect how the performance metrics should be interpreted. Can the authors clarify?*

AR: Thank you for pointing out that this was not sufficiently clear in the manuscript. The 80 borehole sites used in Table 3 are a subset of the 238 in-situ observations. The full in-situ dataset comprises boreholes, test pits, and ground surface temperature measurements, and all of these data, together with the rock glacier evidence, were used to define the response variable of the model and to train the model.

We agree that the borehole-based evaluation is therefore not fully independent from the training data—and we didn't make such claim. However, this comparison was not intended as the primary validation of model generalization. The main model assessment is based on spatial cross-validation, which evaluates predictive performance on withheld spatial blocks and is specifically designed to assess spatial generalization and transferability.

The borehole subset was evaluated separately because boreholes provide the most reliable direct evidence of permafrost presence or absence among the in-situ observations. Since they are only a fraction of our data set, there is no risk that our model was overfit to these data points; we are therefore confident that our assessment is not over-optimistic.

Changes in manuscript: We revised the description of the response variable in Section 3.2 to state explicitly that it is derived from all geomorphological and in-situ evidence described in Section 3.1.1. We also clarified in the caption of Table 3 that the borehole-based evaluation uses a subset —the most reliable subset— of the training data (less than 0.8 %, therefore practically independent although we continue to avoid this word).

RC1.3b: *Assuming the validation is independent, the model at PFI* ≥ 0.75 *is very conservative: when it detects permafrost, it's usually right (88% positive predictive value), but it only catches 40% of actual permafrost sites. For a planning tool, it might be preferable to have some false alarms rather than miss 60% of permafrost occurrence.*

Lowering the threshold to PFI ≥ 0.50 *improves sensitivity to 83% but specificity drops to 33%. There doesn't seem to be a threshold that works well on both counts, which suggests the model may need more information to discriminate – whether that's additional predictors, or addressing the temporal mismatch discussed above.*

AR: We agree that the threshold of PFI ≥ 0.75 yields a relatively conservative classification, with high specificity and positive predictive value but comparatively low sensitivity. This behavior is consistent with the interpretation of high PFI values as indicating conditions that are clearly favorable for permafrost occurrence.

Nevertheless, we interpret the PFI primarily as a continuous favorability index rather than as a fixed binary classifier. The threshold of 0.75 was chosen to delineate areas with conditions that are clearly favorable for permafrost occurrence, but it is not intended as a universally recommending decision threshold for all planning contexts.

Depending on the intended application, lower thresholds may well be appropriate if a more precautionary interpretation is desired and a higher false-positive rate is acceptable. Our intention is therefore not to prescribe a single operational cutoff, but to provide a transparent continuous index from which different thresholds can be chosen according to context.

We also note that threshold-dependent classification statistics are not directly comparable across studies because they depend strongly on the definition of the response variable and the selected cutoff. The AUROC obtained here, especially under spatial cross-validation, indicates useful discriminatory ability and spatial generalization across a large and heterogeneous study domain.

At the same time, we agree that incorporating additional predictors describing snow cover, substrate properties, and other local controls would likely improve discrimination in marginal permafrost settings. We already discuss these limitations and future directions in the manuscript, and point the user of the PFI to the possibility of taking such small-scale factors into account.

Changes in manuscript: We revised Section 4.3 to clarify that PFI ≥ 0.75 is used here to identify conditions favorable for permafrost occurrence, while the continuous PFI can support alternative thresholds depending on the application. We also added wording to make clear that threshold-dependent classification statistics should be interpreted in the context of the intended use of the model.

RC1.4: *The model uses only three predictors: MAAT, centred PISR, and latitude. The authors acknowledge that local factors like snow cover and ground material aren't included but it would be helpful to understand why additional predictors that are available at high resolution weren't tried. Snow cover duration proxies can be derived from remote sensing. Precipitation was a significant predictor in Boeckli et al. (2012) for the Alps, and even in drier settings snow albedo affects the surface energy balance — worth exploring given that PISR is already a key predictor.*

The study domain spans from near-zero precipitation in the Atacama to several thousand mm/yr in western Patagonia. In the hyper-arid north, snow is essentially absent and *permafrost is primarily a radiation–temperature problem. Further south, snow insulation, duration, and melt infiltration become increasingly important controls on the ground thermal regime. The relative importance of the processes driving permafrost distribution shifts fundamentally across this domain, and with only three additive predictors — none of which capture snow processes — it's not clear the model can express that.*

AR: We thank the reviewer for this thoughtful comment. The selection of predictors was guided primarily by the need to ensure consistent spatial coverage and robustness across the entire study domain, which spans nearly 38° of latitude and encompasses very different climatic regimes.

While additional predictors such as snow cover duration, precipitation, or substrate characteristics may locally influence permafrost occurrence, incorporating them consistently at the spatial scale of mainland Chile presents several challenges. In particular, snow cover duration products derived from remote sensing typically have substantially coarser spatial resolution (e.g., MODIS, ESA CCI: 500 m) than the 30 m target resolution used in this study. Higher-resolution products are available only locally for a small part of our study region and after our analyses were finalized (e.g., a study by Dietz et al., 2025, in *Remote Sensing*). Terrain-dependent downscaling of coarse-resolution products would have been beyond the scope of this work.

This being said, snow cover duration and related snow processes are strongly influenced by temperature and radiation, which are already represented in the model through MAAT and potential incoming solar radiation (PISR). At the regional scale considered here, PISR also captures terrain-related controls on shading and solar exposure that strongly influence snow persistence.

Regarding precipitation, we agree that this variable can be important in some mountain regions. However, across the Chilean Andes the dominant precipitation gradient is primarily latitudinal, ranging from hyper-arid conditions in the north to very high precipitation in western Patagonia. This large-scale gradient is therefore largely represented by the latitude predictor included in the model. At the same time, spatially consistent high-resolution precipitation datasets covering the entire study area remain limited, and in general weather stations are extremely scarce throughout the Chilean Andes.

For these reasons we deliberately adopted a parsimonious predictor set that captures the main first-order climatic controls while minimizing the risk of introducing spatial artefacts or inconsistencies across the large and climatically heterogeneous study domain. As noted in the manuscript, incorporating additional predictors describing snow cover, substrate properties, or microclimatic effects represents an important direction for future work once suitable datasets become available.

Changes in manuscript: We clarified in Section 5.2 that the selection of predictors reflects a deliberate focus on spatially consistent variables available across the entire study domain, and that additional predictors such as snow cover duration or substrate characteristics may improve future regional models once suitable datasets are available.

RC1.5: *The in-situ evidence is heavily concentrated between 26–28° S (181 of 238 observations), with only 6 south of 34° S, and the rock glacier inventories have gaps at 24–26° S and 36.5–43.5° S. The model is extrapolating into large regions with little ground truth, yet the PFI maps cover all of mainland Chile without spatial uncertainty quantification. For a planning tool product, this seems worth flagging – at minimum, which regions are extrapolations vs. validated predictions?*

AR: We agree that the spatial distribution of permafrost evidence used in this study is uneven. This pattern reflects the availability of field observations rather than the design of the modeling study.

However, it is also important to note that parts of the study area where observational evidence is sparse —particularly in south-central Chile (36–44° S)— are characterized by comparatively low elevations. As a result, the potential domain of mountain permafrost in this region is limited, which reduces the practical implications of sparse observations.

To assess how well the model generalizes beyond the immediate vicinity of observations, we evaluated model performance using spatial cross-validation, which withholds geographically clustered subsets of the data and therefore provides an assessment of predictive performance in fairly large unseen regions.

In response to the reviewer's suggestion, we have now also obtained model-derived uncertainty estimates using the delta method, an approximation recommended by Wood (2017) for situations with large numbers of predictions. Uncertainties (standard deviation) on the PFI scale are mostly (75 percent) below 0.03; only 1 percent are greater than 0.077. Mean uncertainty is lowest between ~27 and 40°S and highest south of about 51°S. This confirms the high accuracy of PFI estimates and complements our previous model assessments.

Finally, as noted in the revised Discussion, regions with sparse observational evidence should be interpreted cautiously and represent priority areas for future sampling and monitoring efforts.

Changes in manuscript: (1) Information on uncertainty estimates from the delta method has been added to Section 4.1.

> **RC1.6:** *I'd like to see a single map showing the entire model domain (18–56° S) with the PFI output. Figure 1 is useful for the detail but it's hard to get a sense of the full extent of the predictions — particularly how the model behaves in the south, where training data are sparse and conditions are very different from the core study area. Even at a coarse scale this would help the reader judge the spatial coverage. The boreholes and rock glaciers could be additionally added to this map, so the reader quickly grasps the setup. That said, Figure 1 is a really informative graphic and should be kept.*

AR: We thank the reviewer for this suggestion, based on which we include an additional map in the revised manuscript. Since this is challenging due to Chile's size and geometry, in the initial submission we chose to illustrate the results using a combination of latitudinal summary plots (Figures 1 and 3) and regional example maps. In particular, the two maps in the appendix (Figures A1 and A2) for representative areas in the northern Andes (27° S) and the southern Andes (54° S) together with the map in Figure 5 illustrate the range of environmental conditions covered by the model. Numerous additional regional maps are also included in the technical report associated with the project (DGA, 2022a), which we will point out in the revised manuscript.

Changes in manuscript: We added a national-scale map and added a sentence to the manuscript pointing readers to the publicly available dataset and the technical report, which contains numerous additional regional maps illustrating the model results across Chile.

> **RC1.7:** *The activity classification of 10,517 rock glaciers is based on visual interpretation of ESRI World Imagery – a composite mosaic with spatially variable and unspecified acquisition dates, not a time series. Several criteria in Table 1 seem to imply dynamic*

assessment (e.g. "fresh" debris) that would be difficult from a single undated image. What date range does the imagery represent? This adds to the temporal mismatch discussed in Main Comment 2. Also, visually classifying this many rock glaciers is a large manual undertaking and intact vs. relict can be quite subjective. It's not clear how this was done in practice – how many operators, was reliability tested, were existing classifications re-assessed by the same person(s)? Some clarification would be helpful on this task.

AR: We thank the reviewer for raising these important points regarding the classification procedure. Rock glaciers were classified through visual interpretation of high-resolution satellite imagery using the ESRI World Imagery available in 2022, which provided the most recent consistent imagery available across the study area at the time, features a resolution of 1 m or better (mostly Maxar WorldView–2 / GeoEye-2).

The identification and classification of rock glaciers relied on geomorphological, geomorphometric, and environmental criteria visible in the imagery, including the morphology of the rock glacier front, the presence of ridges and furrows, surface texture and tone, and indicators of surface stability or collapse, as summarized in Table 1. These criteria follow commonly used approaches in rock glacier inventories (Roer & Nyenhuis, 2007) and were adapted to the conditions of the Andes (Azocar et al., 2017).

The classification work was carried out by two operators over a period of approximately two months. Where activity status information was available from existing inventories, these classifications were systematically re-evaluated using the same criteria to ensure consistency across the dataset. In addition, ambiguous cases were reviewed jointly by the operators in order to minimize subjective differences in interpretation. Due to this collaborative and interactive process it was not possible to quantitatively measure operator agreement or disagreement.

We agree that visual classification of rock glacier activity involves a degree of subjectivity, which is an inherent limitation of many rock glacier inventories based on remote sensing. To acknowledge this limitation and highlight recent methodological advances, we now also refer to emerging remote-sensing approaches and recent guidelines for rock glacier inventories.

Changes in manuscript: We added a short description of the imagery source, the classification workflow, and the number of operators involved in the interpretation in Section 3.1.1. We also added references to recent methodological developments and guidelines for rock glacier inventories.

3 Technical details

RC: *1. l.74: "prevalant" → "prevalent."*

RC: *2. l.93: "potential incoming solar radation" → "radiation."*

RC: *3. l.311: "significanctly" → "significantly."*

AR: Thanks. Changed as requested.

RC: *4. l.200: "concurvity" is technically correct but likely unfamiliar to most readers. A brief description might help.*

AR: Since the term “concurvity” may not be familiar to all readers, we added a short clarification explaining that it refers to the nonlinear analogue of multicollinearity in generalized additive models.

Changes in manuscript: A short explanatory phrase was added at the first occurrence of the term “concurvity”.

> **RC:** *5. CHELSA downscaling achieves* $R^2_{adj} = 0.893$ *with residual SD of 1.09 °C. This seems non-negligible relative to the MAAT range critical for permafrost (−4 to 0 °C). How much does this propagate into PFI uncertainty?*

AR: We thank the reviewer for this important point. We agree that a residual SD of 1.09 °C is not negligible relative to the temperature range relevant for marginal permafrost conditions. This statistic, however, refers to the accuracy of the downscaling model used to transfer CHELSA MAAT from its native resolution to the 30 m target grid, rather than to the total physical uncertainty of MAAT itself.

We did not propagate this uncertainty explicitly into a separate PFI uncertainty estimate. The predictive performance assessed by spatial cross-validation and the evaluation using in-situ evidences already reflects the combined influence of uncertainty in predictors, ambiguity in the response data, and model misspecification.

Changes in manuscript: We added a short note in Section 3.1.2 clarifying that the MAAT downscaling error propagates into the PFI model and is included in its performance estimates.

> **RC:** *6. Comparison with Gruber (2012): the authors note that the global PZI is “too restrictive” in the Chilean Andes but don’t really explore why. Is it resolution, input data, or the modelling approach? A more nuanced comparison would be helpful.*

AR: We agree that differences between the global Permafrost Zonation Index (PZI) of Gruber (2012) and the regional model presented here deserve clarification. However, it is difficult to attribute the differences to specific sources of uncertainty without a dedicated analysis of the global model framework.

More generally, the two models differ substantially in spatial scale and modeling approach. The PZI is a global product representing coarse-scale permafrost patterns using globally available input data, whereas the model presented here is a regional empirical model calibrated specifically for the Chilean Andes using local permafrost evidence and higher-resolution terrain data. Differences between the two maps are therefore expected.

At the same time, the evaluation against in-situ observations presented in this study (and not available for PZI modeling) clearly indicates that the regional PFI model does not systematically overestimate permafrost occurrence. Rather, the classification results suggest that the threshold used to identify favorable conditions for permafrost is relatively conservative—we report specific results in the Discussion, comparing PZI to our field and borehole evidence. The conservative bias of PZI was previously also noted by Azocar et al. (2017), although without reference to such in-situ data.

Changes in manuscript: We revised the discussion comparing the PFI results with the global Permafrost Zonation Index (PZI) of Gruber (2012) to clarify that the PFI model is empirically calibrated and evaluated using local permafrost evidence. We also added a reference to Azócar et al. (2017), who previously noted that the global PZI appears comparatively restrictive in the Chilean Andes.

RC: *7. Code and data availability: only the PFI raster is shared. The R code, training data, and rock glacier inventory should also be made available. Consider Zenodo or similar.*

AR: We agree that public access to data and code is important for reproducibility. The permafrost index generated in this study has therefore been released through the official digital repository of the Dirección General de Aguas (DGA) under an open-data licence, and a copy is being shared via Zenodo. However, the underlying datasets used to train and evaluate the model (including the rock glacier inventory and the compilation of geomorphological and in-situ permafrost evidence), as well as the modeling code, were developed within a project funded by the DGA, which retains the rights to distribute these materials. The authors are therefore not authorized to redistribute these input datasets or the code independently. Access requests may be directed to the DGA.

Changes in manuscript: We revised the Data and Code Availability statement to clarify that the PFI raster is publicly available through the DGA repository and that the underlying datasets and modeling code are subject to distribution restrictions imposed by the funding agency; a copy of the PFI raster is being deposited in a research data platform.