Review to Ren et al.: Inferring on-glacier meteorology from physical modeling and remote sensing, submitted to the journal The Cryosphere in 2026

Responses to reviewers for Ren et al, submitted to *The Cryosphere* in 2026

We would like to thank both reviewers for their constructive comments. Based on these comments, we have reran all analyses with a wider air temperature prior and air emissivity scheme. The main results are similar and the main conclusions have not changed. Accordingly, we have carefully revised the manuscript based on these new analyses. The reviewers' comments are marked in black and responses in blue. The line numbers in the reviewers' comments refer to the previously submitted manuscript, Fig. RX refers to the figure in this rebuttal and Fig. X and SX refer to the figures in main text and supplementary of the previously submitted manuscript, respectively. Italic sentences are taken from the revised manuscript to clarify our response.

Review 2: Douglas Brinkerhoff

This article describes the application of a Bayesian scheme for inferring temperature and precipitation biases in downscaled ERA5-land such that when used in conjunction with a surface energy balance model, the resulting predicted SMB and albedo are consistent with some independent satellite derived products. This method is applied to 4 glaciers (2 in the Alps and 2 in high-mountain Asia), and the incorporation of these additional constraints is shown to improve model fidelity to a held-out set of stake-based SMB measurements. In general, I find this manuscript to be a step in the right direction via its more-careful-than-most treatment of uncertainties, and its incorporation of novel observational datasets. However, it also suffers a little bit from confusing descriptions in a few places, and also from what is in my view an insufficient acknowledgment of some of the obstacles that stand in the way of the broad generalization of the methods presented here.

We thank the reviewer for their very detailed, constructive review and insightful comments. Based on those, we have added a number of statements to acknowledge the limitations of our method. We have also adopted the most suggestions to improve the clarity and rigor of our manuscript, and provided a detailed justification and explanation for a few points that the reviewer concerns.

1 Line-by-line comments

L95 In the broader context of this paper, I am not sure that I can co-sign the assertion that mass/energy balance models of the class described here are more easily transferable – the paper itself discusses the

need to do a surprising amount of hand-tuning and subjective modification to inversion procedures, even after incorporating additional constraints.

Thank you for this perspective, which has led us to clarify our stance. The aim of our study is to gain a more realistic representation of meteorological fields (air temperature and precipitation) over mountain glaciers at high spatiotemporal resolution. In the broader context of the paper we would like to state that fully physical models (more broadly) are more transferable and can provide deeper understanding of glacier change compared with temperature-index models that rely on empirical relationships between glacier mass balance and meteorological variables (e.g. air temperature and precipitation). We have revised this statement to make it clear, i.e. "*Compared with temperature-index glacier models that rely on the empirical relationship between glacier mass balance and meteorological variables, fully physical energy- and mass-balance models…*"

Regarding our inference work, it already operates semi-automatically and is promising to be a full automation since there are no arbitrary parameterizations. We are also actively working toward this direction to further enhance its user-friendliness and large-scale regional applications.

P3.2.2 This continuity approach (ASTER SMB) to surface balance inference should be viewed as exceptionally error prone. First, the actual uncertainties in each of the remotely sensed constituents to the calculation (surface elevation change, surface velocity, ice thickness) are all not well constrained. In particular, they all have spatial autocorrelation that does not appear in reported errors (with the exception of Hugonnet's dH/dt). More challengingly, using them here requires the computation of numerical derivatives of the computed flux, which amplifies the error in the underlying products. While the authors do a credible job of showing that this product does a reasonable job of reproducing observations for some of their glaciers, I think that this product should be presented with much more skepticism and a fairer accounting of its limitations.

Thanks for this explanation and concern about the limitation of ASTER SMB. We have added these details in this section to directly show the limitation of ASTER SMB, i.e., "*Indeed, each mentioned remotely sensed input used for calculating ASTER SMB flux, including surface elevation change, surface velocity, ice thickness datasets, has its own uncertainty and could lead to very high uncertainty in the final SMB through error propagation and amplification, which inevitably limit the reliability of our inference to some specific glaciers. To explore the sensitivity of the inferred glacier meteorology (Section 5.2) to the quality of ASTER SMB…* "

We have also extensively discussed the impact of SMB accuracy on inference in section 5.2, where we have compared inference between using ASTER and high resolution (Pléiades/TanDEM-X) SMB datasets (Fig. 10). If we take Parlung No.4 Glacier as an example, we see that the large underestimated mass loss at the glacier termini (3-6 m w.e. a^{-1}, Fig. S2) lead to overestimation in inferred precipitation (Fig. 7l). We also state that it is necessary to properly evaluate the uncertainty of ASTER SMB before processing inference.

L230 The asterisk operator should probably be a × instead, here and elsewhere.

Revised as suggested.

L221–239 This section would be more helpful if the downscaling methods that are currently relegated to a reference had a very short (like one sentence) description accompanying them.

Good suggestion. We have improved this statement, i.e., *"Following Machguth et al. (2009) and McCarthy et al. (2022, 2025), we use a spline interpolation with neighboring ERA5-Land grids to statistically downscale these variables are statistically downscaled to a set of locations across the glaciers where we run the T&C model (purple dots in Fig. 1)* "

L231 If the goal is to develop a method that is spatially transferable, then using different methods of downscaling for different glaciers is less than ideal. I think it would be better to either apply a precipitation lapse rate in all cases or in no cases.

Thanks for this thoughtful suggestion. We agree that maintaining consistency in applying or omitting a precipitation gradient is generally ideal for spatial transferability. However, our choice is constrained by the interaction between the precipitation gradient and the upper bound of the precipitation correction prior (k_p). To keep the Bayesian inference framework straightforward and robust, we set a maximum upper bound ($k_p = 3$). If we omit the precipitation gradient entirely, the required upper bound at certain glaciers would need to be exceptionally high, making it difficult to define a consistent prior. Conversely, if we apply a precipitation gradient to small glaciers that occupy only a single ERA5-Land grid, unrealistic gradients are often extrapolated since too many neighboring grids are off-glacier regions and are unable to well represent precipitation gradients on- or near to the glacier in question, adding additional obstacles to the inference. We therefore adopt a grid-dependent threshold: the lapse rate is only applied to glaciers covering more than two ERA5-Land grids (i.e., the Aletsch Glacier in this study). This rule-based approach preserves the transferability of the overarching inference framework while accounting for spatial scale.

We have clarified this rationale and added more details in the revised manuscript, i.e. "*Since large glaciers with wide elevation ranges are expected to exhibit gradients in precipitation, we apply a lapse rate to glaciers covering more than two ERA5-Land grids. Consequently, a static precipitation gradient derived from a 5×5 neighboring grid of ERA5-Land (0.719 mm yr^{-1} m^{-1}) is applied to the Aletsch Glacier, while no precipitation lapse rate is used for the smaller glaciers.*"

Eq. 2 This equation is incomplete without either (1) an explicit (mathematical) statement that θ lives within bounds, or (2, preferred) retention of the prior term.

We specify in the text related to Eq. 2 that θ follows a uniform distribution (which is bounded) and that hence the prior term $p(\theta)$ is a constant. For this case, the proportionality statement in Eq. 2 is correct and complete, and we prefer to show Eq. 2 without $p(\theta)$ for simplicity and clarity.

L276 It's not clear that MCMC is actually intractable here - especially if one were to start with a good initial guess, they would probably yield a significantly higher sample density in the vicinity of the MAP point. Brute force methods are good when they can be applied (like here), but it's worth mentioning that this will not be feasible in the event that the size of the parameter space were to increase (or if hierarchical priors were adopted to enforce spatial correlation between locations on a glacier).

Thanks for your insightful comment. We agree that with only two inference parameters (air temperature and precipitation) and a well-chosen initial guess, an MCMC approach would likely be tractable and yield a high sample density near the MAP point; yet, such an approach usually needs several chains and a few thousand samples per chain for convergence, even in apparently simple cases, which needs substantially more computational budget than allowed for our study.

Our choice of a brute-force Monte Carlo approach with a limited number of 1,000 realizations is motivated by its implementation simplicity and ease of parallelization across independent locations on clusters of HPC (High Performance Computation), as well as predictable computation time. However, we entirely agree that this brute-force approach is strictly bound to our current low-dimensional setup. It would immediately fail to scale if we increased the parameter size or introduced hierarchical priors to account for spatial correlation between locations. Following your advice, we have revised the text to remove the claim that MCMC is generally intractable here, and we have explicitly added a discussion regarding the scaling limitations of our direct sampling approach. Adapting an MCMC approach will be worthwhile for future applications, but requires careful testing to ensure convergence of solutions.

The new statements read: "*A Markov Chain Monte Carlo implementation as in, e.g. Rounce et al. (2020), is a common way to perform Bayesian inference, but the computation cost is relatively high since it usually needs several chains and a few thousand samples per chain for convergence and it is hard to implement in parallel calculation on HPC (High Performance Computation) due to its dependency on each simulation. Given the computational cost of the full energy balance model (2–3 minutes per run) and ease of parallelization, we instead approximate the Bayesian posterior solution with a "low-budget brute-force" Monte Carlo approach, using 1000 total realizations per location on the glacier (N_{MC}). This brute-force method is viable here because of the low-dimensional inference space and independence at each location; it would lose its effectiveness and justification for larger inference spaces or if hierarchical priors were adopted to enforce spatial correlation. We draw a limited 1000 realizations (see Section 3.4.3) from the prior distribution (see Section 3.4.1)...*"

Eq. 7 It's worth being a bit more rigorous in the presentation of this equation, and to describe in what ways θinf erred might end up being a poor approximation to the true posterior mean.

Thank you for this suggestion. In general, $\theta_{inferred}$ might end up being a poor approximation to the true posterior mean if the likelihood surface was multi-modal or extremely peaked, and/or the resulting posterior distribution is highly skewed. As we additionally monitor effective sample size, we become aware of potentially critical cases where statistics might become less reliable.

Accordingly, we have added this statement to the manuscript, "*We note that the inference might end up being a poor approximation to the true posterior mean if the likelihood surface is multi-modal or extremely peaked, and/or the resulting posterior distribution was highly skewed.* "

Eq. 8 It's also worth noting that very small ESS (i.e. size approximately 1) is the characteristic of the circumstance in which none of the sampled parameters provide a good fit to observations, and so the one that is the least bad absorbs all of the probability mass. Perhaps comment on whether that is occurring here?

We thank the reviewer for this warning, but wish to clarify that an ESS around 1 could also mean that a single sampled parameter fits much better than all others (whether it is close to the truth or not), and vice versa, the case that all sampled parameters are bad could produce a relatively high ESS in the case that all perform similarly bad. A small ESS just warns that there are huge differences in likelihood values of the top performing samples, and that larger ensemble sizes would be needed to characterize the steep likelihood fronts better.

We do observe very small ESS at some locations (Fig. R1), and we explicitly checked our results to decide if a realistic parameter set was found (small ESS but high likelihood) or whether the result was physically implausible. If the difference between modelled (corrected by posterior mean) and satellite target (glacier albedo or SMB) exceeded the uncertainty of satellite observation, we classified the inference as poor approximation and removed it from further analysis.

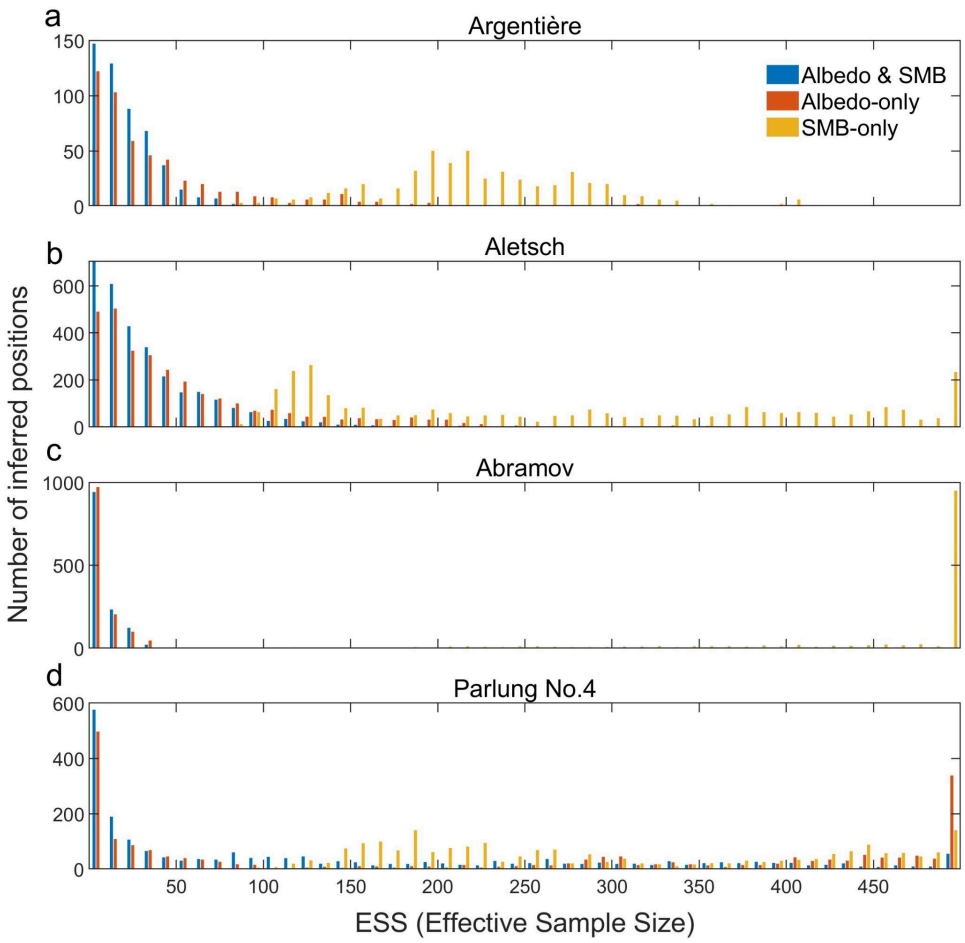


Figure R1. Histogram of Effective Sample Size (ESS) of all inferred positions on four tested glaciers. The last bar in each panel indicates the sum of ESS >500.

Accordingly, we have added a discussion to clarify this point: "*Very small ESS values, on the other hand, reveal that only very few correction factor values are able to fit the data sufficiently well. While very small ESS values would affect a reliable statistical analysis (e.g., investigating the shape of the posterior distribution in further detail, which is beyond the scope of this study), they identify scenarios of high identifiability and are hence of interest to us. We specifically check whether a realistic parameter set was found (small ESS but high likelihood) or whether the inferred value is physically implausible: if the*

difference between modelled (corrected by posterior mean) and satellite target (glacier albedo or SMB) exceeds the uncertainty of satellite observations, we classify the inference as poor approximation and remove it from further analysis.

L306 I am totally lost when it comes to this pre-inference prior choosing business. Where does the modelled SMB used in making this distinction come from? Is this some sort of peculiar form of empirical Bayes?

This step can be seen as similar to a localized form of empirical Bayes; however, we are not searching for unknown hyperparameters here, as is usually the case for empirical Bayes approaches. The modelled SMB used in this step is generated by driving our T&C model with downscaled ERA5-Land forcing, which is then compared against ASTER SMB to narrow down the parameter range before the formal Monte Carlo simulations. The main purpose is to reduce computational cost and focus our efforts on physically-plausible parameter regions for posterior estimation. In an ideal infinite-computation scenario, one could use completely independent unconstrained uniform priors. However, in our case, a global unconstrained search could cause the whole 1,000 realizations to be wasted on parameter combinations that yield instantly unreal glacier states (e.g., too much or too small glacier mass changes). Therefore, we adopt this empirical pre-screening as a pragmatic strategy to bring physics knowledge into the prior constraint.

Accordingly, we have modified the statement to make this clear, i.e., "*Therefore, we explicitly adopt an empirical Bayes to narrow down priors (Cremona et al., 2025, Fig. 3). We run T&C model with a very wide parameter combination, i.e.,* ΔTa *is from −10 to +10 °C with 1 °C interval and* k_p *is from 0.5 to 3 with 0.2 interval, and then determine the priors by physically interpreting the discrepancy between this modelled SMB (*Mod_{SMB}*) and satellite observation (*Sat_{SMB}*)....* "

Eqs. 11–15 Don't use the asterisk.

Revised them as suggested.

L355–359 I don't follow this spatial extrapolation reasoning. Please try to describe more clearly.

Sorry for this unclear statement. The goal of our inference framework is to obtain glacier meteorology at high resolution (100 m × 100 m, grey grids in Fig. R1). We find that the maximum computational cost is 1000 Monte Carlo simulations and it is unfeasible to process all these simulations at each 100 m grid cell. We therefore process Monte Carlo simulations at a 500 m grid (purple grid in Fig. R1), but still process

meteorological inference at 100 m grid by assigning nearby ensemble simulations, assuming that the downscaled meteorological forcing ensemble would not be materially different within 500 m. In this way, the computational cost is significantly reduced but the spatial resolution of inferred meteorology is still 100 m.

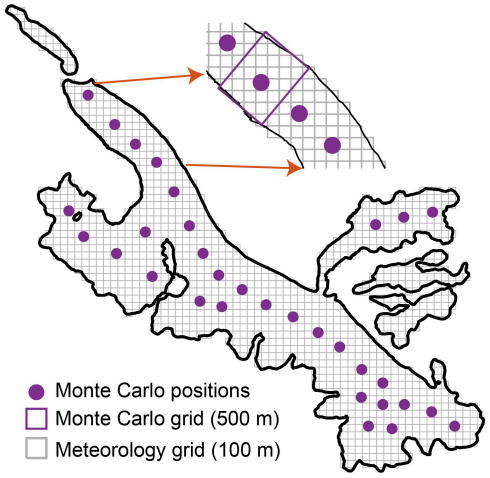


Figure R1. Schematic diagram mapping a spatially extrapolated Monte Carlo simulation grid to expected meteorological grids

We have added Fig. R1 into Supplementary (Fig. S3) and improved the explanation in the manuscript and added more details at the beginning of this section, "*Our goal is to obtain glacier meteorology at high resolution (100 m × 100 m, grey grids in Fig. S3), but it is unfeasible to run all 1000 Monte Carlo simulations at each 100 m grid due to large computational cost. To balance computational efficiency and spatial representativeness,...*"

L378 and elsewhere I am not sure that I understand the value that the cloud frequency comparison provides, since this is not strongly indicative of either of the inferred values. I suppose if there were never any clouds, you could confidently say that the precipitation was zero, but I don't think that's the case in any of your domains.

In our case, cloud frequency is used to indicate the relative variability of precipitation in space, i.e., where precipitation is more and where it is less over glaciers. We agree that cloud frequency is not strongly indicative of precipitation, but more cloud generally indicates more water vapor and precipitation, i.e., a positive correlation between cloud frequency and precipitation. This is a simplified proxy to qualitatively evaluate precipitation spatial patterns. A fully quantitative evaluation can only be performed once other high-resolution data over glaciers are available.

We have added more details to show the reason of cloud frequency more clearly, i.e. "*In general, more frequent clouds indicate more precipitation, such that comparing their respective patterns could provide a first-order, qualitative evaluation of the spatial variability of inferred precipitation. A fully quantitative evaluation can only be performed once other high-resolution data over glaciers are available.*"

Fig. 4, row 2, col 1 The concentration of samples at the edge of the plot like this is a classic indication of prior mis-specification, and a very dangerous side-effect of using uniform-in-a-box priors – the parameter wants to be outside this bound, but the prior doesn't allow it. Is a temperature correction of 1.99 actually infinitely more probably than a correction of 2.01? For the other plots - these exceptionally concentrated posteriors might imply that the data uniquely constrains a single model - or it might imply that all models fit badly.

Thanks for this insightful comment. To understand the effect of uniform-box prior on inference, we set up another two Monte Carlo simulations: one with 1000 simulations with a slightly wider Ta prior [-5°C, +4°C], and the other one with 2000 simulations with a largely wider one [-5°C, +10°C] at an accumulation location of Aletsch Glacier, and then compare to the narrow Ta prior currently used [-5°C, +2°C] (Fig. R2). The result shows that a large wide Ta prior generates a much wider posterior distribution than the narrow Ta prior (Fig. R1a and c) as expected, but in fact the structure and location of the most probable region changes only slightly, and consistent with that, the improvement of modelled albedo with a higher Ta correction (+2°C to +3°C) is limited compared to the original Ta correction around +2°C (Fig. R3).

According to the evaluation with in-situ air temperature observations (Shaw et al., 2025), downscaled ERA5-Land air temperature shows a 2 to 4 °C cold bias at only 3 of 115 sites. Fig. 1a and 1c show that 1000 simulations with a slightly wider Ta prior [-5°C, +4°C] can generate more complete converge posterior distribution, we therefore rerun all MC simulations with this new Ta prior [-5°C, +4°C] to ensure enough search space for the Ta correction factor. In our case, we leverage the existing knowledge to establish prior distribution, which can be improved with more accurate knowledge in the future. We also add the discussion about this limitation, "*Both priors can be further improved once more accurate knowledge is available.*".

In addition, we have added a filter to remove those poor inferences where the difference between modelled and observed SMB exceeds the uncertainty of satellite observation (see response to your comment on Eq. 7 before), further reducing the negative effect of prior mis-specification like this one.

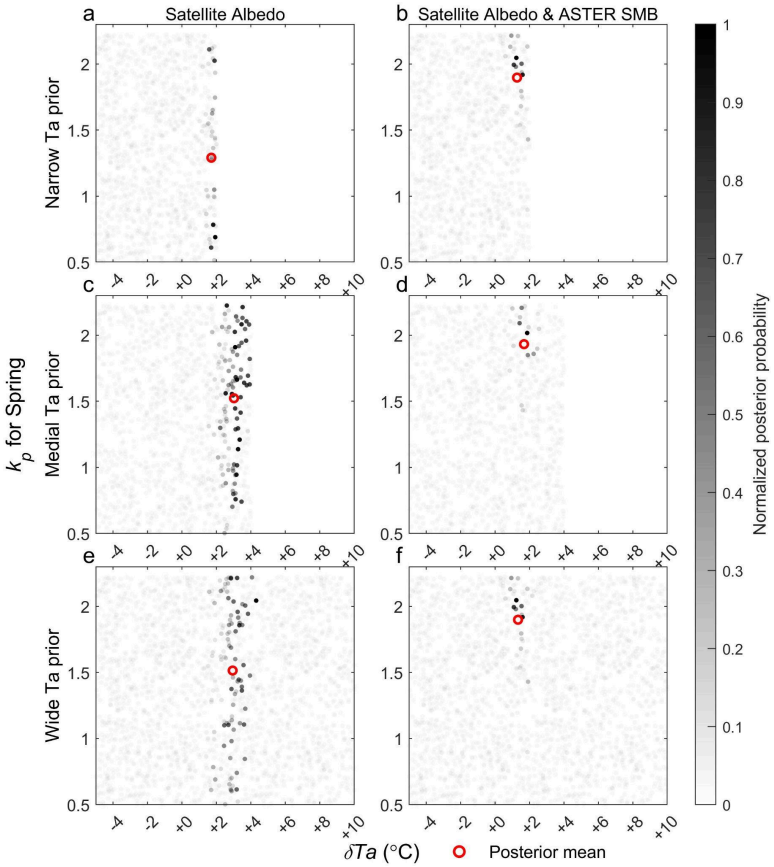


Figure R2. Comparison of Monte Carlo posterior samples between narrow [-5°C, +2°C] and wide [-5°C, +10°C]Ta priors at one accumulation location of Aletsch Glacier (marked P_{Al_acc} in Fig. 1d). Color coded is their normalized posterior probability for two different combinations. Red points are inferred (posterior mean) bias-correctors and black/grey points are 1000 Monte Carlo simulations.

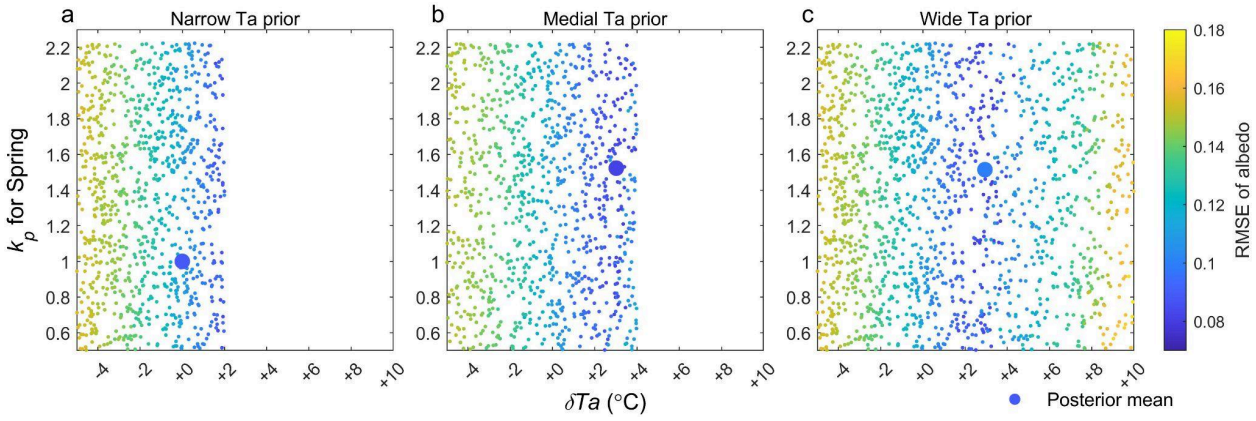


Figure R3. Same as Fig. R1 but for the RMSE between modelled and observed glacier albedo.

To check whether those concentrated posteriors indicate good constraints or not, we analyze the match between simulations to observations (Fig. R4). The result shows that our inferences are located in a zone that generally matches both albedo and SMB well, indicating good constraints by our inference framework.

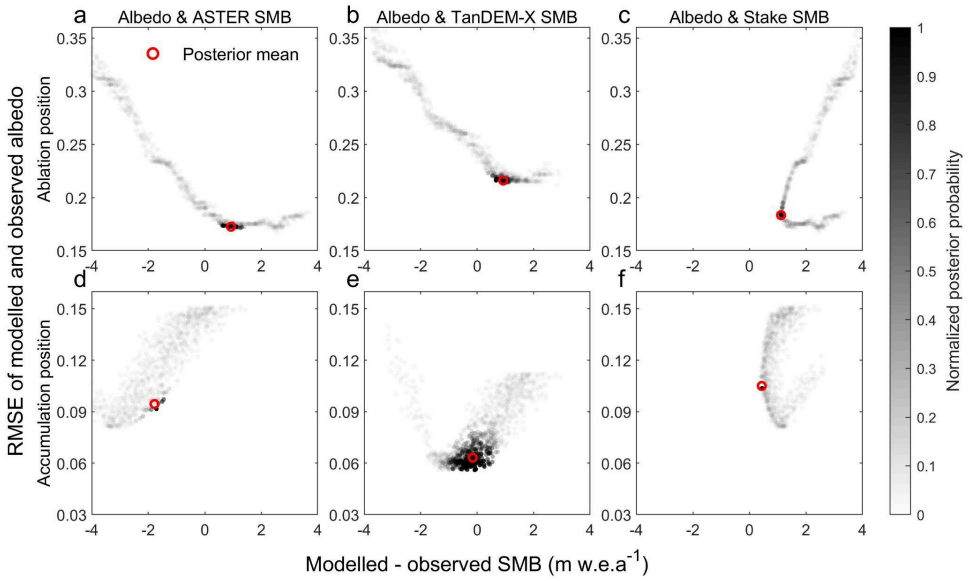


Figure R4. Two examples of Monte Carlo simulation against satellite/stake observations on Aletsch Glacier (same as Fig. 4). Red points are inferred results and black/grey points are 1000 Monte Carlo simulations. Color coded is their normalized posterior probability. In particular, the x axis of panel c and f are RMSE between modelled and stake SMBs since the stake SMBs are multitemporal observations.

Accordingly, we have added Fig. R4 into Supplementary and discussion about the impact of prior establishment in section 3.4.1, i.e., " *...which thus would result in an additive factor for air temperature ΔTa in the interval [-5, +4]. Although this prior may be too narrow for some specific regions, it is an inevitable compromise due to the limitation imposed by computational cost*".

Fig. 5 I don't understand why the gray dots are different between panels. Wouldn't it be better to use the same gray dots and just change the methods?

The gray dots in Fig. 5a-c are modelled SMB driven by downscaled meteorological forcing and are different in each panel because their periods are different. To ensure fair comparison, the periods of modelled SMB driven by downscaled meteorological forcing should be the same as observed SMB datasets, which are different in this case. The Stake SMB could be single-year or multi-year observations depending on data availability, while ASTER SMB is the average over 2015-2019 and Pléiades SMB over 2012-2022.

We have added an explanation in the caption of Fig. 5 to make this clear, i.e. "*The aggregation periods in panels a, b and c are different and correspond to the observation period of stake/satellite SMB*"

Fig. 7 This plot is pretty incomprehensible to me. There are a large number of axes, colors, and unlabeled numbers. It would be better to figure out how to present this information in a more targeted (and larger) format.

We agree that this could be made clearer, and have improved it by splitting air temperature and precipitation, labelling all numbers and changing the color (Fig. R5).

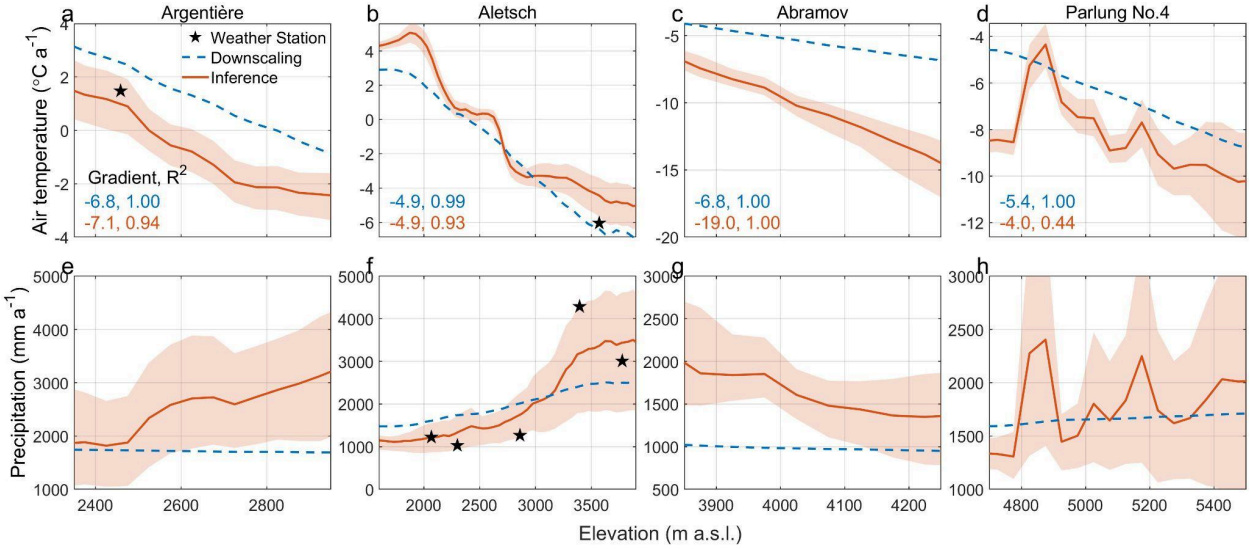


Figure R5**.** Downscaled and inferred multi-year average air temperature (a-d) and precipitation (e-h) by satellite albedo and ASTER SMB datasets as a function of elevation at the four tested glaciers during 2015-2019. The pentagrams indicate air temperature/precipitation of the station observations. Blue numbers indicate air temperature gradient (℃ km^{-1}) and R^2 for downscaled ERA5-Land, while red numbers correspond to inferred air temperature. The shaded area represents the uncertainty of inferred meteorological variables in each elevation bin. The dataset used for aggregation is the same as in Fig. 7 and S3.

L427 This line about spatial variability in precipitation correction factor is sort of unsatisfying. First, the associated plot is relegated to the supplement. Second (and more importantly), there's no discussion of what these precipitation patterns might mean. In the supplemental figure, we see theses factors varying wildly across pretty short spatial scales, while there also doesn't appear to be any temporal correlation between the same locations. Why? If the climate model/downscaling method is just missing a piece of physics, then shouldn't this be represented in these recovered biases? The patterns here seem to me much

more likely to be a result of mis-specification of either the physics or the likelihood. I recognize that hierarchical or spatially tied modelling is beyond the scope of the present work, but the inferred correction factors would be much more defensible if they had some modicum of agreement with one another.

Thank you for these insightful comments on the precipitation correction. We agree that this figure should be included in the main manuscript and that some necessary explanations and discussions are missing here. We attribute varied correction factors to the high spatial variability in target datasets. As the black rectangle shows (Fig. R6e, j and h), for example, strong spatial variabilities happen between SMB/albedo datasets and correction factors. Several factors can explain their high heterogeneity: for SMB, they could be blowing snow, avalanches (Argentière Glacier, Kneib et al., 2024) and errors of satellite observations, while for albedo there are presence of debris, crevasses and shadow due to terrain inside the glacier outline (red rectangles in Fig. R6s and t). As a comparison, precipitation correction factors on Aletsch Glacier are more spatially autocorrelated due to relatively homogeneous SMB and albedo. In addition, as the ablation region of Parlung Nr.4 shows (green rectangle in Fig. R6), an inappropriate prior can also result in poor inference (see the response to your comment for L524–527).

Accordingly, we have added Fig. R5 in the main text (now as Fig. 6) and pointed it out in the Result Section, "*However, some unrealistic patterns occur in specific regions due to the limitations of satellite datasets (rectangles in Fig. 6 and see discussions in Section 5.4)*" and added discussion in the Section 5.4, "*Generally consistent spatial patterns of satellite targets and inferred correction factors indicate that these strong spatial gradients are mostly caused by strong spatial heterogeneities in the SMB and albedo data. For SMB, these heterogeneities could be due to blowing snow, avalanches (Argentière Glacier, Kneib et al., 2024) and errors of satellite observations (green rectangle in Fig. 6), while for albedo the high heterogeneity can be related to presence of debris, crevasses and shadows due to terrain inside glacier outlines (red rectangles in Fig. 6). As a comparison, relatively homogeneous SMB and albedo on Aletsch Glacier infer a more spatially autocorrelated precipitation correction factor, suggesting that our inference framework is reliable as long as good quality satellite datasets are available.*"

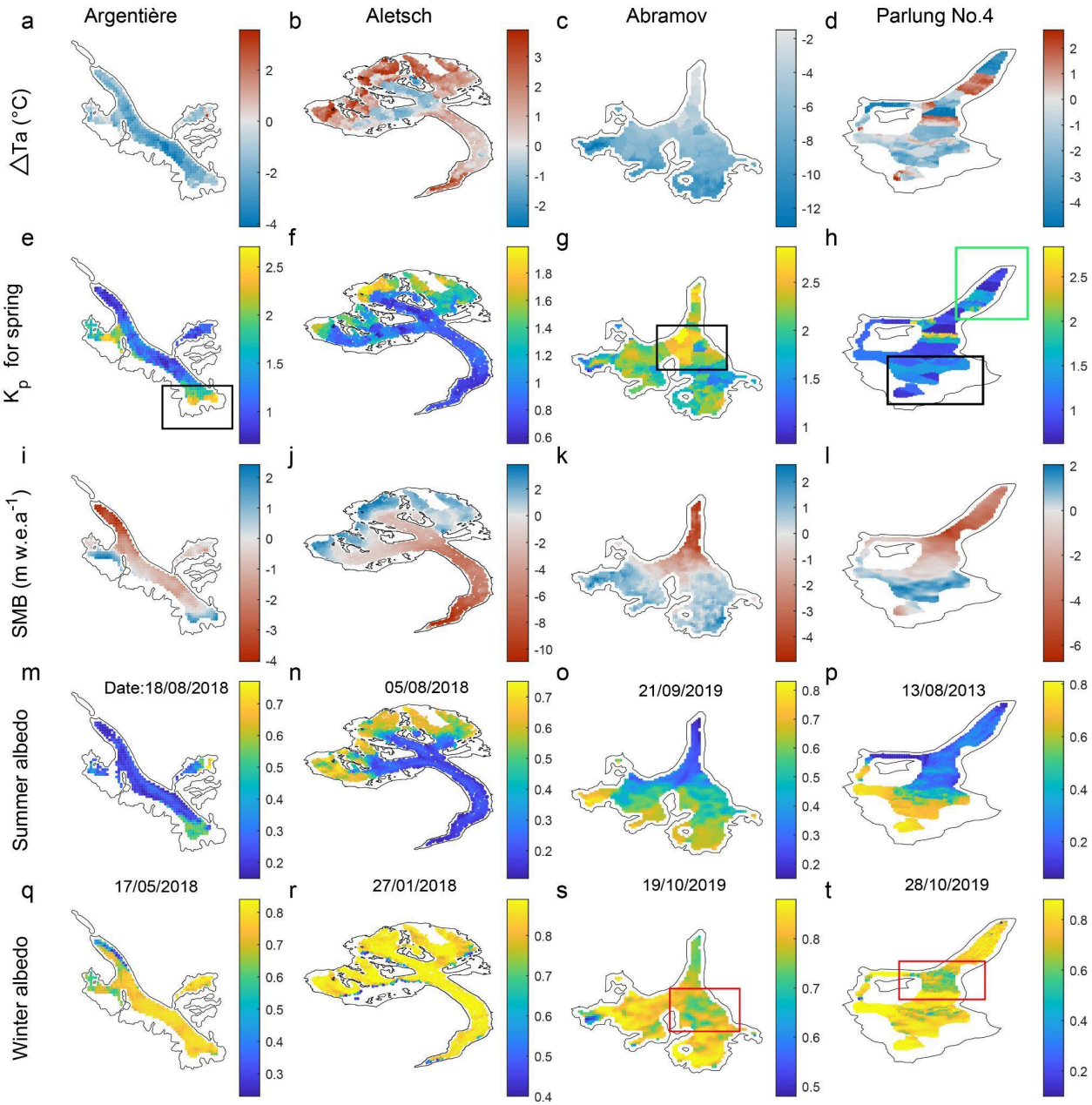


Figure R6. Inferred correction factors for air temperature (ΔTa, a-d) and spring precipitation (k_p, e-f) against ASTER SMB (i-l), satellite firn/ice albedo (m-p) and snow albedo (q-t) on four tested glaciers.

L484 It would be nice if these were reported alongside a notion of relative error. Does a 1m RMSE correspond to a 10% or 100% error relative to the stake measurement (recognizing that this gets tricky when mass balance is near zero).

We have computed the modelled error relative to stake SMB (Fig. R7), but removing mass balance records which lie in the [-1, 1] m w.e. interval. Overall, the relative errors of modelled SMB with

inference are about 25% for annual or summer mass balance across six tested glaciers. As you correctly mentioned, that relative error gets tricky near neutral mass balance, so we think that it is better to keep the original figure , but add this statement in the manuscript: "*RMSE values are reduced to <1.5 m w.e. for annual mass balance (i.e. 25% relative error) with improvements of 20-60%, and <1 m w.e. for seasonal mass balance with 40-50% improvement across most glaciers.*"

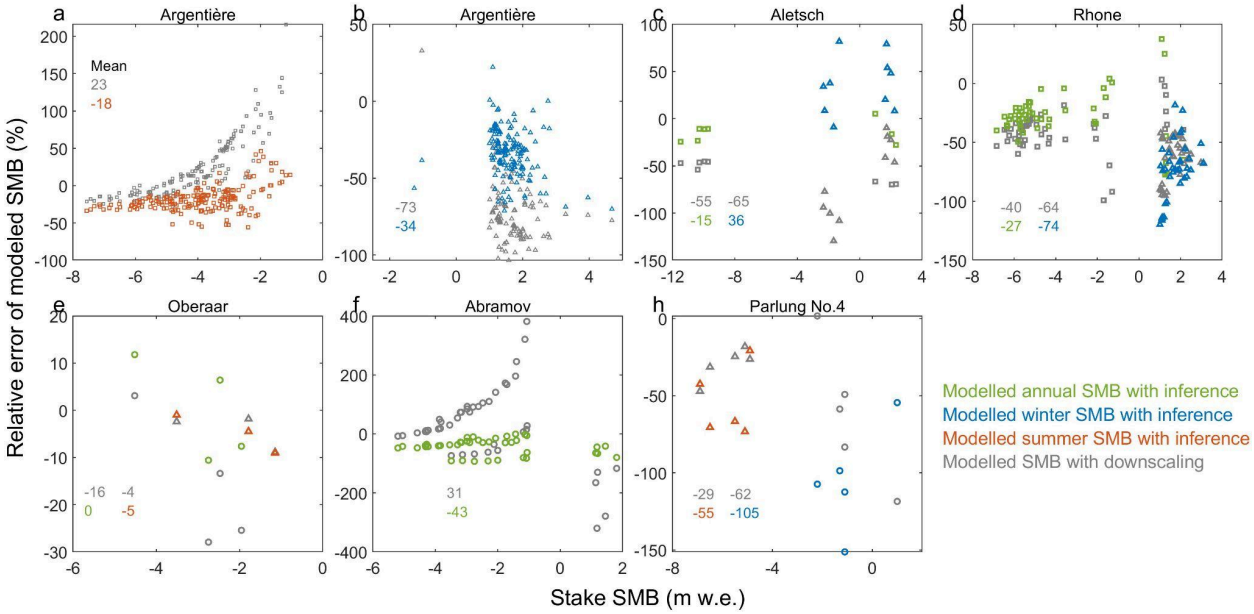


Figure R7. Relative error of annual/seasonal mass balance forced by downscaled and inferred meteorology against in-situ stake surface mass balance on the six studied glaciers. Each point represents an individual stake mass balance measurement.

L524–527 I am having trouble parsing what this line means – how do these observed products make a prior inappropriate? Is the prior wrong, or is the likelihood model associated with this product wrong?

This is a prior wrong related to the error of ASTER SMB. As mentioned before, we use ASTER SMB to narrow down the prior, meaning that the range of prior depends on satellite SMB. In the case of Parlung No.4 Glacier, mass loss underestimation by the satellite SMB will lead to a more positive range of precipitation prior, i.e. a wrong prior. Prior knowledge plays a critical role in Bayesian inference. To reduce the negative impact of uncertainty of ASTER SMB (which, as noted by the reviewer, is particularly uncertain), some filters should be performed to remove the dataset with large errors before proceeding to inference.

Accordingly, we have added more details at the beginning "*...large underestimation for Parlung No.4 Glacier (3-6 m w.e. a^{-1}, Fig. S2). Since we use ASTER SMB to narrow down the prior range (see section*

3.4.1), these underestimations of mass balance by the ASTER data can lead to inappropriate prior ranges…." and discussion at the end of this paragraph "*Therefore, a necessary check on the reliability of satellite SMB is recommended before undertaking inference to reduce the negative impact of errors in the SMB dataset.*"

L572 It's worth mentioning that adjusting your prior in response to a perceived inadequacy in the data product is a little bit of a Bayesian crime and that having to make these sorts of judgement calls limits global applicability of these methods.

Thank you for raising this point, as it will be important in many practical studies. In general, inadequacy of data products has to be judged before performing any type of calibration or inference, to avoid compensation of these (known or perceived) errors and resulting lack of generalization skill in prediction. There are several ways to confront this problem: if possible and justifiable, correct the data set itself; if the exact outcome of errors is unknown but statistical properties can be derived, assumptions about error distributions should inform the likelihood definition. In fact, we have incorporated uncertainty of albedo (σ_{alb}) and SMB (σ_{smb}). Adjusting priors to make up for erroneous observations would generally not be a good idea; here, however, we are inferring *corrections* and hence expert knowledge about the expected sign and extent of these corrections is a valid source of information for prior definition; especially in a resources-limited setting. As an outlook towards applicability of our proposed approach, we have added the following statement in the section of outlook:

"*In general, data quality will strongly affect the plausibility of the inference results, and we recommend case study-specific pre-processing analysis to reveal potentially needed modifications of our proposed approach, concerning e.g. the likelihood and prior definition.*"

References:

Cremona, A., Huss, M., Landmann, J. M., Schwaizer, G., Paul, F., and Farinotti, D.: Constraining sub-seasonal glacier mass balance in the Swiss Alps using Sentinel-2-derived snow-cover observations, J. Glaciol., 71, 25, 2025.

Machguth, H., Paul, F., Kotlarski, S., and Hoelzle, M.: Calculating distributed glacier mass balance for the Swiss Alps from regional climate model output: A methodical description and interpretation of the results, J. Geophys. Res. Atmos., 114, 2009.

McCarthy, M.: A MATLAB toolbox for statistical downscaling of climate reanalysis data (Version 1.0.0). GitHub, 2025.

McCarthy, M., Meier, F., Fatichi, S., Stocker, B. D., Shaw, T. E., Miles, E., and Pellicciotti, F.: Glacier contributions to river discharge during the current Chilean megadrought, Earths Future, 10, 2022 002852, 2022

Rounce, D. R., Khurana, T., Short, M. B., Hock, R., Shean, D. E., and Brinkerhoff, D. J.: Quantifying parameter uncertainty in a large-scale glacier evolution model using Bayesian inference: application to High Mountain Asia, J. Glaciol., 66, 175–187, 2020.

Kneib, M., Dehecq, A., Gilbert, A., Basset, A., Miles, E. S., Jouvet, G., Jourdain, B., Ducasse, E., Beraud, L., Rabatel, A., Mouginot, J., Carcanade, G., Laarman, O., Brun, F., and Six, D.: Distributed surface mass balance of an avalanche-fed glacier, The Cryosphere, 18, 5965–5983, 2024.