Response to Reviewer 1

This manuscript by Zhu et al. investigates a critical topic. However, in my opinion, there are fundamental issues with model validation, methodological justification, and presentation. If the author can address these issues, I would recommend publication with major revisions.

Weaknesses in Model Validation, Methodology, and Presentation

1. Inadequate Temporal Validation of Core Model Outputs (Figs. 2, 3, 4 & 5):

Figs. 2 & 3 (Scatter Plots): Presenting albedo, radiation, and mass balance as scatter plots obscures the model's performance in simulating the temporal evolution of these variables. Time-series plots are essential to reveal whether the model correctly captures the timing and duration of key events, such as the rapid albedo drop during heatwaves. A good correlation (R-value) can mask significant temporal lags or failures to replicate specific events.

Response: We fully agree. In the revised manuscript, we have replaced all scatter plots with time-series comparisons. Figure 2 now presents observed vs. modeled time series for albedo, incoming longwave radiation, and outgoing longwave radiation throughout the study period. These plots clearly demonstrate that the model captures both the magnitude and temporal evolution of these variables, including the rapid albedo drop during heatwaves. Corresponding methodological details have been expanded in Section 4.2.

Fig. 3 (Mass Balance): The scatter plots validate only the annual net result. To truly validate the melt process, the authors must present **observed vs. modeled cumulative mass balance over time** for each year. This would indicate whether the model accurately simulates the rate and timing of ablation. Furthermore, disaggregating into **observed vs. modeled ablation and accumulation separately** is critical, as a correct net balance can result from compensating errors in melt and snowfall.

Response: We agree that annual net balance is insufficient. We have now added observed versus modeled cumulative mass balance over time in Figure 2g, which directly validates the rate and timing of ablation across the study period. In addition,

we have separately validated ablation and accumulation using seasonal mass balance observations for June–September 2021 (Figure 2d) and glacier-wide mass balance for the three balance years (Figures 2f–g). This ensures that the correct net balance is not achieved through compensating errors.

Unclear Representativeness: The manuscript does not clearly state how the point measurements from a limited number of stakes (Fig 1d) are representative of the glacier-wide means presented in Figs. 4 and 5.

Response: We acknowledge this lack of clarity. In the revised Section 3.2, we now explicitly state that point mass balances were calculated at individual stakes following Yao et al. (2012), and glacier-wide mass balances were obtained by integrating these point measurements over the glacier area using an elevation-based interpolation method. This approach is standard in glacier mass balance studies and ensures that the point measurements are representative of the glacier-wide means. Please see lines 188–191.

Redundancy (Figs. 4 & 5): Fig. 4 shows only modeled monthly fluxes. If sub-annual observed data exist, this figure should be updated to include the observed cumulative mass balance overlaid on the modeled one, making it the primary validation figure. This would render the interannual bar chart in Fig. 5 redundant, but it would also provide valuable insights into the model's temporal performance.

Response: We have reorganized the presentation to eliminate redundancy and improve clarity. All validation comparisons between observations and simulations are now consolidated in Figure 2 (including time-series for albedo, radiation, point-scale mass balance, seasonal mass balance, glacier-wide mass balance, and cumulative mass balance). Figure 4 now focuses exclusively on the simulated energy and mass balance characteristics (monthly variations of energy components, mass components, interannual mass balance, and accumulation anomalies). The original Figure 5 (interannual bar chart) has been removed. This reorganization provides a clear separation between model validation (Figure 2) and model results (Figure 4).

2. Contradictory Evidence for Heatwave Identification (Fig. 7):

The central premise of the study is the impact of "extreme heatwaves." However, **Fig. 7b shows that air temperature during the identified heatwave periods (grey**

shading) does not appear anomalously high compared to other times in the same melt season. Some peaks outside the grey areas appear just as high or higher. This directly undermines the classification of these periods as thermally extreme events at the glacier surface and calls into question the applicability of the standardized heatwave index (S_HI) derived from a distant valley station (Tuotuhe). The authors should provide a rigorous justification for why these periods are classified as heatwaves despite the apparent lack of a strong temperature signal at the AWS.

Response: We appreciate this critical observation. To ensure robust and objective heatwave identification, we have taken the following steps:

1. Adopted a rigorous statistical method – We used the standardized heatwave index proposed by Shan et al. (2024), which uses a dynamic 30-year moving window, identifies warm spells as consecutive days exceeding the 90th percentile, and removes short-lived events.
2. Multi-source consistency – We applied the method to three independent temperature time series: (i) Tuotuohe meteorological station (in situ), (ii) ERA5 reanalysis at the Tuotuohe grid cell, and (iii) ERA5 reanalysis at the Sangqu Glacier grid cell. The three datasets identified slightly different heatwave periods. To resolve discrepancies, we defined the final heatwave events as the overlapping periods consistently identified by all three datasets. This yielded three periods: 1–8 June, 7–15 July, and 4–29 August 2022.
3. Reconciliation with visual inspection – A simple visual inspection of raw temperature peaks can be misleading. The standardized heatwave index method does not classify heatwaves based solely on instantaneous peaks, but on sustained periods of anomalously high temperatures relative to a dynamic climatological baseline. The three identified periods meet the criteria for duration and extremity, even if individual daily peaks occur outside these periods.

We have added a detailed description of this methodology in Section 4.3 and updated Figure 3 (formerly Figure 7) to show the overlapping periods clearly.

3. Unsupported Extrapolation from Point to Glacier-Wide Data:

The study relies on a **single Automatic Weather Station (AWS) at 5,700 m a.s.l.** (Fig. 1). Yet, the energy and mass balance results are presented as "glacier-wide" (e.g., Figs. 4, 8).

The method for spatially interpolating meteorological data across the entire glacier's elevation range (5,400–6,104 m a.s.l.) from this single point needs to be sufficiently justified. Using a constant lapse rate and assuming wind speed and relative humidity are uniform across the glacier is a major simplification that likely introduces significant errors, especially during extreme and stable atmospheric conditions. The manuscript should explicitly discuss the uncertainties introduced by this extrapolation and provide evidence (e.g., from other studies or model sensitivity tests) that the chosen method is valid for deriving glacier-wide energy balance.

Response: We agree that this is a critical limitation of single-AWS studies. We have substantially expanded the relevant discussions and analyses as follows:

1. **Justification of spatial extrapolation (Section 4.1)** – Following established practices (Hock & Holmgren, 2005; Van Pelt et al., 2012), we interpolate meteorological variables using a temperature lapse rate of 0.6 °C/1000 m derived from local observations, a precipitation gradient calibrated against glacier-wide mass balance, and the assumption of spatially uniform wind speed and relative humidity (supported by the relatively small spatial variability of these variables in the study region).
2. **Explicit discussion of uncertainties (Section 6.4)** – We now systematically discuss that simplified lapse rates and gradients cannot fully capture spatiotemporal heterogeneity, especially during extreme events, and that the uniform wind/humidity assumption may introduce uncertainties in turbulent flux calculations. We acknowledge these as inherent limitations.
3. **Sensitivity analysis (Section 4.2 and Tables S1–S6)** – We perturbed the temperature lapse rate (±0.1 °C/1000 m) and precipitation gradient (±10%). Results show that glacier-wide mass balance is highly sensitive to the lapse rate (−141.4 mm w.e. a^{-1}) but less sensitive to the precipitation gradient

(−2.2 mm w.e. a^{-1}). In addition, we quantified the uncertainty in the extreme melt increase during the heatwave period as 99.3 mm w.e. (Table S1), based on a comprehensive sensitivity analysis of key surface energy balance parameters. These revisions significantly strengthen the transparency and rigor of our study regarding spatial representativeness.

4. Unjustified Application of the Framework to Other Glaciers (Fig. 6):

The application of the dual-threshold method to other glaciers is presented without a clear rationale. If intended as **validation**, it is circular, as it applies the method to other short records without an independent benchmark. If intended to show **prevalence**, it dilutes the focus from the detailed process-based analysis of Sangqu Glacier. The purpose needs to be clearly stated and justified.

Response: We thank the reviewer for this important clarification. In the revised manuscript, we have completely rewritten **Section 4.4** to clearly articulate the purpose and rationale. The dual-threshold framework is **not** intended as a formal validation of the method (which would require an independent benchmark). Instead, the purpose is twofold:

1. **To demonstrate robustness and transferability** – By applying the framework to five glaciers with long-term records across different climatic regimes (central, northern, and western Tibetan Plateau), we show that the method yields consistent and interpretable results under varying conditions. This supports its credibility when applied to the short (3-year) record of Sangqu Glacier.
2. **To quantify uncertainty** – The interval between the relaxed and strict thresholds explicitly captures the uncertainty arising from short training periods. Applying the framework to multiple glaciers allows us to observe how this uncertainty varies across different glacier types, further informing its application to data-scarce sites.

We have revised the text in **Section 4.4** to state this rationale clearly, and **Figure 5** (formerly Figure 6) now presents the results as a demonstration of the framework's utility, not as a validation in the strict sense.

Summary of major revisions

- **Model validation:** Replaced scatter plots with time-series (Figure 2); added cumulative mass balance, seasonal validation, and cross-validation with Xiao Dongkemadi Glacier.
- **Heatwave identification:** Adopted multi-source overlapping periods based on SHI method; clarified in Section 4.3 and Figure 3.
- **Spatial extrapolation:** Added justification (Section 4.1), uncertainty discussion (Section 6.4), and sensitivity analysis (Section 4.2, Tables S1–S6).
- **Dual-threshold framework:** Rewrote Section 4.4 to clarify purpose (robustness and transferability, not circular validation).

We believe these revisions fully address the fundamental issues raised by the reviewer and have substantially improved the quality of the manuscript.

Major Revisions Required:

Comment 1: **Fundamental Validation:** Replace scatter plots in Figs. 2 and 3 with **time-series comparisons**. Crucially, include **time series of observed vs. modeled cumulative mass balance** to validate the ablation process.

Response: We fully agree that time-series comparisons, particularly for cumulative mass balance, provide a more intuitive and rigorous validation of the model's ability to capture the ablation process. To address this, we have made the following improvements.

1. Model parameter optimization

To reduce uncertainties from manual parameter calibration, we implemented a stepwise calibration scheme. The Shuffled Complex Evolution (SCE-UA) algorithm was first used to calibrate parameters related to incoming longwave radiation (using RMSE as the objective function). Subsequently, the Nondominated Sorting Genetic Algorithm II (NSGA-II) multi-objective optimization algorithm was applied to calibrate the EBFM model parameters using both R^2 and RMSE as dual objective functions, constrained by daily albedo measurements and glacier-wide mass balance observations.

2. Enhanced multi-scale validation

Following calibration, we rigorously validated the model using independent in-situ observations, including albedo, L_{in}, and L_{out} from the on-glacier AWS, point-scale mass balance at different elevations, seasonal mass balance for June–September 2021, glacier-wide mass balance for the periods 2019–2020, 2020–2021, and October 2021–July 2022, as well as cumulative mass balance time series to directly validate the ablation process over the full study period.

3. Revised figures

Based on these new simulations and validation results, we have regenerated Figures 2 and 3, replacing the original scatter plots with time-series comparisons. Specifically, Figure 2 now includes time-series comparisons for albedo, L_{in}, L_{out}, and—most importantly—observed versus modeled cumulative mass balance (see updated Figs. 2f and 2d), which directly validates the ablation process throughout the study period. The corresponding methodological details have been expanded in Section 4.2 to clearly describe the calibration and validation procedures.

We believe these revisions substantially strengthen the model validation and directly address the reviewer's concern. All updates have been incorporated into the revised manuscript.

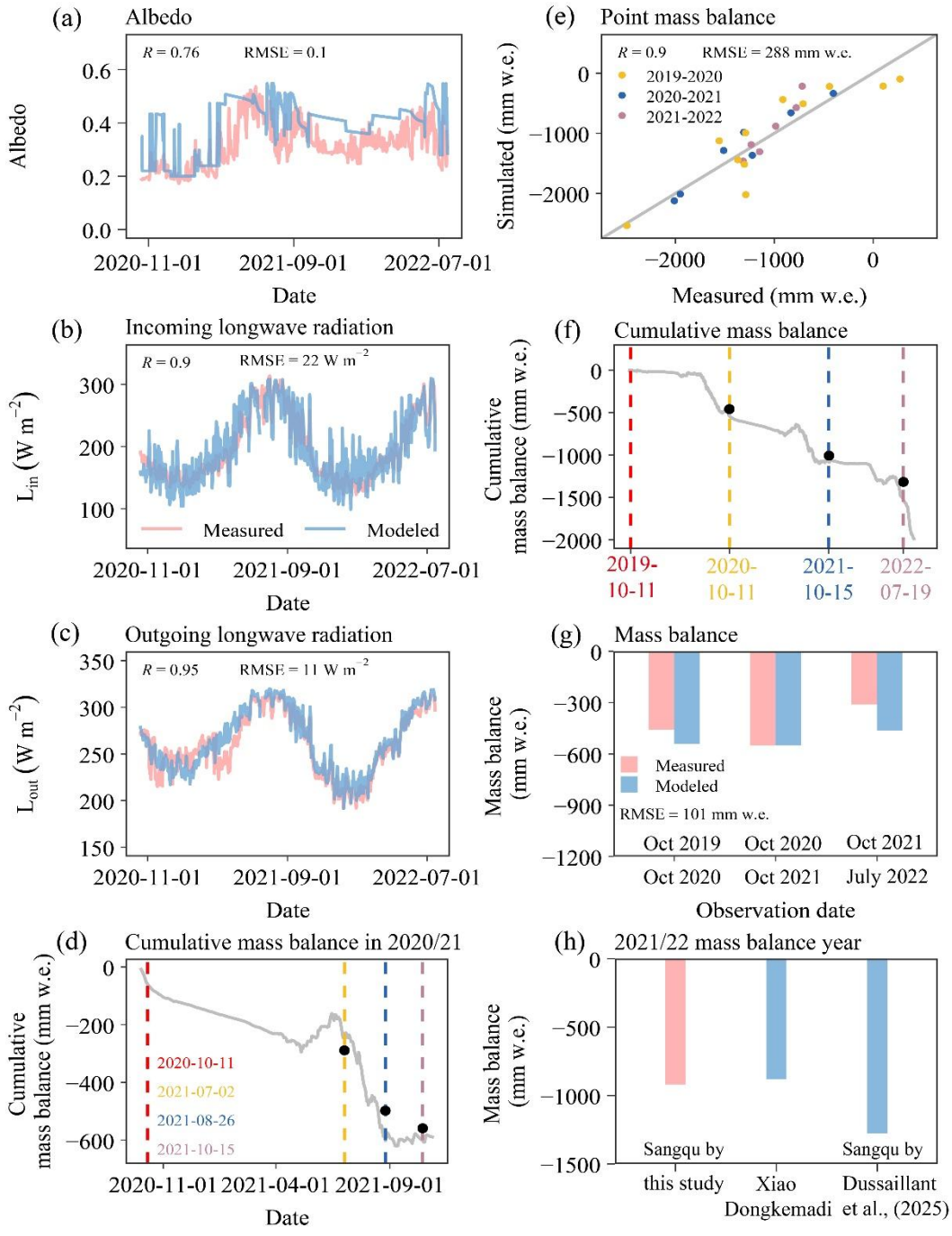


Figure 2. Comparison between observations and simulations of (a) albedo, (b) L_{in}, (c) L_{out}, (d) glacier mass balance from June to September 2021, (e) point mass balance at different elevations, (f) accumulation mass balance, and (g) glacier-wide mass balance for the periods 2019/20, 2020/21, and from October 2021 to July 2022. (h) Comparison of simulated glacier mass balance for Sangqu Glacier in 2021/22 with observed mass balance for Xiao Dongkemadi Glacier and with mass balance data for the same Sangqu Glacier derived from multi-source satellite remote sensing by Dussaillant et al. (2025).

Comment 2: **Justify Heatwave Classification:** Reconcile the apparent discrepancy in Fig. 7. Provide a compelling explanation for why the grey-shaded periods are classified as extreme heatwaves based on the in-situ AWS data, or re-evaluate the analysis periods.

Response: We thank the reviewer for raising this important point. To ensure the reliability and objectivity of heatwave identification, we adopted the standardized heatwave index method proposed by Shan et al. (2024). This method uses a dynamic 30-year moving window to calculate daily climatological expectations, thereby accounting for climate non-stationarity; it identifies warm spells as consecutive days with daily mean temperature exceeding the 90th percentile of the historical baseline; and it removes short-lived warm spells and merges adjacent events to ensure that the final heatwaves meet statistical criteria for extremity and independence.

Crucially, the in-situ AWS on Sangqu Glacier has a short observational record, which is insufficient to establish a reliable 30-year climatological baseline required by the standardized heatwave index. Therefore, we had to rely on long-term (≥30 years) temperature records from nearby meteorological stations and reanalysis data. To minimize uncertainties from individual data sources, we applied this method to three independent temperature time series derived from two data sources for the summer of 2022 (June–September). The Tuotuohe meteorological station (in-situ observations, ~150 km from Sangqu Glacier) identified three heatwaves: 1–9 June, 3–16 July, and 3–29 August. The ERA5 reanalysis at the Tuotuohe grid cell identified three heatwaves: 1–8 June, 7–15 July, and 4 August–1 September. The ERA5 reanalysis at the Sangqu Glacier grid cell identified three heatwaves: 1–10 June, 7–17 July, and 2 August–1 September.

To resolve discrepancies among the three datasets, we defined the final heatwave events as the overlapping periods consistently identified by all datasets. This yielded three periods: 1–8 June, 7–15 July, and 4–29 August 2022. These periods correspond exactly to the grey-shaded areas in Fig. 3, which are therefore classified as extreme heatwaves based on multi-source data consistency and an objective statistical

framework. The approach effectively reconciles apparent differences among datasets and ensures the credibility of the heatwave classification. We have added a detailed description of this methodology in Section 4.3 of the revised manuscript, and the identified heatwave events are now clearly indicated in Fig. 3.

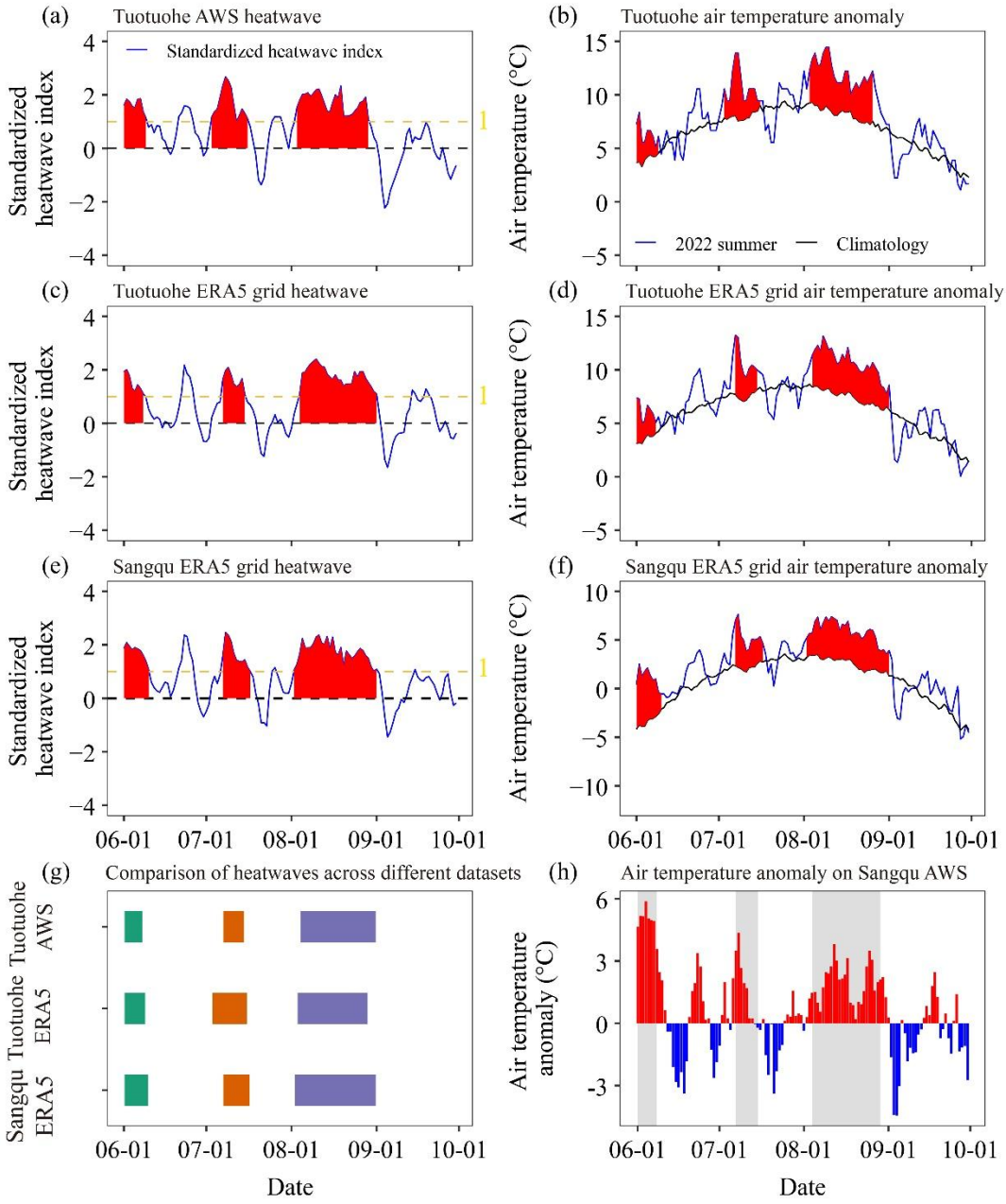


Figure 3. Identification of heatwave events from June to September 2022. Air temperature anomalies are calculated relative to the 1992–2021 climatology. (a, b) Heatwave events and corresponding temperature anomalies at the Tuotuohe meteorological station. (c, d) As in (a, b), but for the ERA5 grid cell containing the Tuotuohe meteorological station. (e, f) As in (a, b), but for the

ERA5 grid cell containing Sangqu Glacier. In panels (a), (c), and (e), the yellow line marks the standardized heatwave index threshold of 1; a heatwave is identified when the index exceeds this threshold for at least three consecutive days. (g) Comparison of heatwave events identified from the three datasets shown in (a), (c), and (e). (h) Summer (June–September) temperature anomalies at Sangqu Glacier in 2022 based on the identified heatwave events.

Comment 3: **Address Spatial Representation:** Justify the extrapolation from a single AWS to a glacier-wide scale. Discuss the associated uncertainties and limitations explicitly. A sensitivity analysis of the model to the spatial interpolation scheme would significantly strengthen the study.

Response: We thank the reviewer for this valuable suggestion. We fully agree that spatial representativeness is a critical limitation of single-AWS studies, and we have substantially expanded the relevant discussions and analyses in the revised manuscript.

1. Justification of spatial extrapolation

In the revised Section 4.1, we now explicitly state the assumptions underlying our spatial extrapolation approach. Following established practices in distributed glacier energy-balance modeling (Hock and Holmgren, 2005; Van Pelt et al., 2012), we interpolate meteorological variables from the single AWS using a temperature lapse rate of 0.6°C/1000 m derived from local observations, a precipitation gradient calibrated against glacier-wide mass balance, and the assumption of spatially uniform wind speed and relative humidity across the glacier—an assumption supported by the relatively small spatial variability of these variables in the study region.

2. Explicit discussion of uncertainties and limitations

In the revised Section 6.4 (Research Limitations), we have added a systematic discussion of uncertainties arising from spatial extrapolation. Specifically, we now highlight that the simplified lapse rate and gradient parameterizations cannot fully capture the spatiotemporal heterogeneity of meteorological variables under complex terrain conditions, particularly during extreme weather events, and that the assumption of spatially uniform wind speed and relative humidity may introduce uncertainties in turbulent heat flux calculations, especially in areas with complex topography. We

acknowledge that these limitations are inherent to single-AWS studies and represent a key direction for future research, which should aim to establish multi-elevation AWS networks.

3. Sensitivity analysis of spatial interpolation parameters

Following the reviewer's suggestion, we conducted a sensitivity analysis for key parameters that govern spatial interpolation. Specifically, we perturbed the temperature lapse rate (±0.1°C/1000 m) and the precipitation gradient (±10%) while keeping all other parameters at their calibrated values. The results show that the glacier-wide mass balance is highly sensitive to the temperature lapse rate, with a sensitivity of −141.4 mm w.e. a^{-1} (Tables S1–S4), indicating that uncertainties in lapse rate have a substantial impact on mass balance simulations. In contrast, the sensitivity to the precipitation gradient is relatively minor (−2.2 mm w.e. a^{-1}), suggesting that mass balance is less sensitive to precipitation distribution than to temperature distribution.

We believe these revisions substantially strengthen the transparency and rigor of our study regarding spatial representativeness. All corresponding updates have been incorporated into the revised manuscript, primarily in Sections 4.1, 4.2, 6.4, and Tables S1–S4.

Comment 4: **Streamline Narrative and Clarify Rationale:** Update Fig. 4 to include observed mass balance time series. Remove or modify redundant figures (e.g., Fig. 5). Provide a clear and convincing rationale for the analysis presented in Fig. 6.

Response: We thank the reviewer for these constructive suggestions. We have carefully addressed each point as follows.

1. Figure 4 and the inclusion of observed mass balance

We fully agree that including observational data enhances the credibility of the analysis. In the revised manuscript, we have reorganized the presentation of observational and simulated data to improve clarity. All validation comparisons between observations and simulations are now consolidated in Figure 2, which includes time-series comparisons for albedo, L_{in}, and L_{out}(Figs. 2a–c), seasonal mass balance for June–September 2021 (Fig. 2d), point-scale mass balance at different elevations

(Fig. 2e), glacier-wide mass balance for the three balance years (Figs. 2f–g), and cross-validation with the neighboring Xiao Dongkemadi Glacier (Fig. 2h). Figure 4 now focuses exclusively on the simulated energy and mass balance characteristics, presenting monthly variations of energy balance components (Fig. 4a), monthly variations of mass balance components (Fig. 4b), interannual variations of annual mass balance (Fig. 4c), and anomalies in accumulation mass balance (Fig. 4d). This reorganization streamlines the narrative by clearly separating model validation (Fig. 2) from model results (Fig. 4), and ensures that observational data are prominently featured in the validation figure. Regarding the incomplete 2022 record (observations only through July), we have maintained transparency by indicating the data gap in Fig. 2g and discussing it explicitly in Section 4.2.

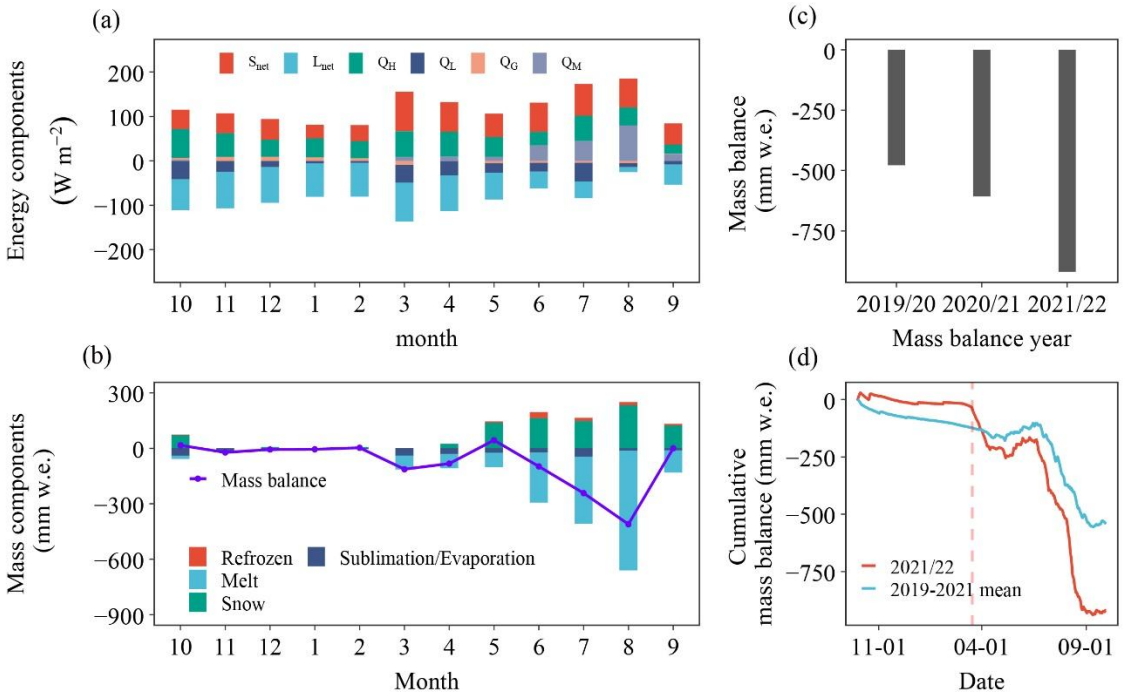


Figure 4. Monthly variations of (a) energy and (b) mass balance components during the 2021–2022 mass balance year. (c) Interannual variations of annual mass balance. (d) Differences in accumulation mass balance between 2021/22 and the 2019–2021 mean.

2. Removal of redundant figures

Following the reviewer's suggestion, we have removed the original Figure 5 and

its associated content from the manuscript. The key information previously presented in that figure has been either integrated into other figures or relocated to the supplementary materials where appropriate.

3. Clarified rationale for Figure 6

We have substantially revised the analysis presented in Figure 6, which is now renumbered as Figure 5 in the revised manuscript. The corresponding methodological description has been completely rewritten in Section 4.4, where we now provide a clear and convincing rationale for the dual-threshold framework used to identify extreme mass loss events from short-term records. The revised section begins by articulating the core idea—translating the conventional percentile-based definition into a parametric form ($\mu - k\sigma$) that can be calibrated from short training periods—and explicitly explains how the interval between the relaxed and strict thresholds quantifies the uncertainty arising from the short time series.

All corresponding updates have been incorporated into the revised manuscript, primarily in Sections 4.2, 4.4, Figures 2, 4, 5 (formerly Fig. 6), and the removal of the original Figure 5.

Conclusion

In my opinion, the manuscript requires major revisions. The current validation is insufficient to prove the model accurately simulates ice melt dynamics. The identification of the key driver, the heatwaves, is not convincingly supported by the presented in-situ temperature data, and the extrapolation of point measurements to the entire glacier is a major, unquantified source of uncertainty. Until these fundamental issues are addressed, the analysis of energy balance mechanisms remains speculative and is built on unverified premises.

We sincerely thank the reviewer for the thorough and critical assessment of our manuscript. We fully acknowledge that the original submission had fundamental weaknesses in model validation, heatwave identification, and spatial extrapolation. In the revised manuscript, we have systematically addressed each of these issues as follows.

On model validation: We have fundamentally revised our validation approach by replacing scatter plots with time-series comparisons (Fig. 2), adding observed versus modeled cumulative mass balance to directly validate the ablation process, and separately validating ablation and accumulation using seasonal observations (Fig. 2d). These revisions provide robust evidence that the model accurately simulates ice melt dynamics.

On heatwave identification: We have adopted the standardized heatwave index (SHI) method (Shan et al., 2024) and applied it to three independent temperature time series from two data sources. By defining the final heatwave events as the overlapping periods consistently identified by all three datasets (1–8 June, 7–15 July, and 4–29 August 2022), we have resolved the apparent discrepancy with the in-situ AWS data. These periods now correspond exactly to the grey-shaded areas in Fig. 7, providing a rigorous and objective justification for the classification.

On spatial extrapolation: We have explicitly justified our spatial interpolation assumptions in Section 4.1, added a systematic discussion of uncertainties in Section 6.4, and conducted sensitivity analyses for key spatial interpolation parameters (temperature lapse rate and precipitation gradient). The results show that while the lapse rate has a substantial impact (−141.4 mm w.e. a^{-1}), the precipitation gradient has a minor effect (−2.2 mm w.e. a^{-1}), and the overall model accuracy at the glacier-wide scale (RMSE of 101 mm w.e.) is sufficient for our extreme event analysis. Independent validation from the neighboring Xiao Dongkemadi Glacier further supports the reliability of our conclusions.

With these revisions, we believe the fundamental issues raised by the reviewer have been fully addressed. The model validation is now comprehensive, the heatwave identification is objectively justified, and the uncertainties associated with spatial extrapolation are explicitly quantified and discussed. We are confident that the analysis of energy balance mechanisms is now built on a solid and well-verified foundation. We thank the reviewer again for the rigorous and constructive feedback, which has substantially improved the quality of our manuscript.

Response to Reviewer 2

The manuscript 'Extreme Glacier Melt in the Central Tibetan Plateau during the Summer of 2022: Detection and Mechanisms' by Zhu and co-authors present a detailed analysis of the energy-balance of Sangqu Glacier using the EBFM model. The authors further apply a statistical approach to determine whether the summer 2022 was extreme in terms of glacier melt and discuss the underlying atmospheric mechanisms leading to the high recorded melt rates.

This is a good study presenting a detailed process-oriented approach to better understand the links between glacier and atmosphere. The model used was applied with care and extensively validated against in situ measurements. Despite this, I still had a number of points to raise that I feel would need to be addressed before this manuscript can be considered for publication.

Response: We are grateful for the reviewer's insightful suggestions and acknowledge the remaining points that need to be addressed. We fully agree with the reviewer that these points are essential for improving the clarity and quality of our manuscript. We carefully addressed all the comments raised by the reviewer and provide a point-by-point response below.

Thank you again for your time and expertise.

General comments:

Comment 1: Statistical analysis with very short time series: I found it hard to follow the description of the new framework used to identify these extreme events from time-series of just a few years. This methodological section goes straight into technicalities, without explaining the general principle, making it difficult to understand what the authors are actually after. Furthermore, given the extremely short focus period, I have serious doubts about the value of such an approach. Along these lines, I fail to

understand why the authors followed what seems to be a purely statistical approach without leveraging the energy-balance model that they validated at this site and could be expanded in time.

Response: We sincerely thank the reviewer for these insightful comments. They have helped us substantially improve the clarity and rigor of our manuscript.

Section 4.4 has been rewritten to clarify the logic of the dual-threshold framework, and a brief justification for using a statistical approach instead of model extension has been added at the end of that section.

1. Methodological description and conceptual framing

We fully agree that the original description of the dual-threshold framework in Section 4.4 was too abrupt and lacked a clear conceptual introduction. In the revised manuscript, we have completely rewritten this section. We now begin by stating the core idea explicitly: to translate the conventional percentile-based definition of an extreme event (which requires long records) into a parametric form ($\mu - k\sigma$) that can be calibrated from short training periods. By establishing a mathematical equivalence between the two definitions within the training period, we obtain a range of valid multipliers $[k_{\min}, k_{\max}]$, which yields both a relaxed threshold and a high-confidence threshold. This design serves two purposes: (i) it provides a parametric rule applicable to the test period, and (ii) the interval between the two thresholds explicitly quantifies the uncertainty arising from the shortness of the training series. We believe this revision clarifies the logical chain of our methodology and addresses the reviewer's concern about the intelligibility of the approach.

2. Value of the statistical approach with short time series

The reviewer raises a critical question about the value of such a method given the limited observation length. We fully acknowledge that, in an ideal setting, long observational records are preferable. However, over 90% of glaciers on the Tibetan Plateau have mass balance records shorter than ten years, making traditional methods (e.g., percentile thresholds, generalized extreme value distributions) inapplicable. Our dual-threshold framework is specifically designed for this common but challenging

scenario. It transforms a non-parametric definition into a parametric rule, and—importantly—the interval between the relaxed and strict thresholds provides a quantitative measure of the uncertainty introduced by the short record. This allows, for the first time, a statistically defensible identification of extreme events in short-term glacier mass balance series. We have now emphasized this point more clearly in the revised manuscript (Section 4.4).

3. Why not extend the energy-balance model instead?

We appreciate this thoughtful question. Indeed, our energy-balance model (EBFM) was carefully validated and could theoretically be extended in time to identify extreme melt events. However, we see two main reasons for developing a separate statistical framework. First, the EBFM requires high-quality, high-frequency meteorological forcing (radiation, humidity, wind, etc.). For years prior to 2020 in our study region, such observations are either unavailable or subject to large uncertainties, and using reanalysis-only forcing would introduce uncontrolled errors that could compromise the detection of extremes. Second, and more importantly, the two approaches serve complementary purposes. The EBFM provides a process-based understanding of *how* extreme melt occurs, while the statistical framework offers a simple, transferable tool that relies solely on annual mass balance observations. This makes it readily applicable to other data-scarce glaciers across the Tibetan Plateau and beyond, where high-quality meteorological data are often lacking. We have added a brief clarification of this rationale in the revised manuscript (Section 4.4).

Comment 2: Parallel with extreme glacier mass losses in other regions of the world: This study is very much focused on the Tibetan Plateau, and I find it a bit disappointing that it does not draw more parallels with recent glacier extreme events that have been reported in other regions of the world, such as Canada, Switzerland, the Andes or Svalbard. Some mechanisms described here are also described for these other regions and it would be interesting to insist on how these events relate or differ.

Response: We thank the reviewer for this constructive suggestion. We agree that

situating our findings within the broader context of recent global glacier extreme events significantly strengthens the discussion.

Section 6.2 has been substantially expanded to systematically compare extreme glacier melt events across different mountain ranges worldwide, highlighting common drivers and regional differences.

Following this suggestion, we have substantially expanded Section 6.2 (formerly titled "Comparison with extreme glacier mass loss events in other regions") to include a systematic comparison with glacier mass loss extremes reported in diverse mountain regions worldwide, including the Swiss Alps (Van Tiel et al., 2026), Austria (Voordendag et al., 2023), Svalbard (Bernard and Friedt, 2025), the Pyrenees (Izagirre et al., 2024), the Tien Shan (Xu et al., 2024), western Canada and the United States (Menounos et al., 2025), and New Zealand (Vargo et al., 2020).

In the revised discussion, we first synthesize common characteristics across these regions, such as record ablation rates, rapid glacier degradation, and the critical role of albedo feedback. We then examine shared driving mechanisms, including anomalous large-scale atmospheric circulation (e.g., westerly jet shifts), summer heatwaves superimposed on winter snow deficits, and energy balance anomalies characterized by enhanced incoming shortwave radiation and sensible heat flux. We also highlight regional differences, such as the role of early spring melt onset on the Tibetan Plateau versus the importance of liquid precipitation events in the Arctic, and the varying contributions of anthropogenic forcing across regions. This expanded comparison not only demonstrates that the extreme melt observed on Sangqu Glacier is part of a broader global phenomenon but also helps to contextualize the specific mechanisms operating in the central Tibetan Plateau.

We believe this revision substantially enhances the global relevance of our study and directly addresses the reviewer's suggestion. The added content is presented in Section 6.2 of the revised manuscript.

Comment 3: Differences in mass balance measurement periods: The mass balance records are reported to end in July in 2022, compared to October in the previous years,

which bodes the question of how comparable these measurements are, especially considering that 2022 is pointed out as the extreme year of the time series.

Response: We thank the reviewer for raising this important issue. We fully acknowledge that the 2022 mass balance record ends in July, whereas the previous years' records extend through October, and we agree that this discrepancy warrants careful consideration regarding the comparability of these measurements.

The shortened observational period in 2022 was unfortunately unavoidable due to the COVID-19 outbreak in Xizang during that summer, which prevented field access after July. To be explicit: our direct stake measurements at Sangqu Glacier in 2022 are only available through July. To determine whether the 2021–2022 balance year (October–September) as a whole qualifies as an extreme mass loss event, we therefore had to rely on a model-based reconstruction of the complete year, supported by multiple independent lines of validation.

We took the following four complementary approaches to ensure robustness:

1. Model parameter optimization

   Following the suggestion of Reviewer 1, we applied the NSGA-II multi-objective optimization algorithm to better constrain the key parameters of the energy-balance model (EBFM) using all available in situ observations. This improved the model's ability to simulate the full seasonal cycle, especially the ablation season.

2. Seasonal-scale validation

   We validated the model against observed seasonal mass balance from the summer of 2021 (June–September), which is the period most relevant to the 2022 extreme event. The comparison shows good agreement, with an RMSE of 329 mm w.e. (see updated Fig. 2d).

3. Cross-validation with a neighbouring glacier (Xiao Dongkemadi Glacier)

   We collected independent mass balance observations from Xiao Dongkemadi Glacier, located approximately 90 km from Sangqu Glacier, for the complete 2021–2022 balance year (October–September). Importantly, Zhu et al. (2026)

also identified that Xiao Dongkemadi Glacier experienced extreme mass loss in 2021–2022. The observed annual mass balance at Xiao Dongkemadi Glacier is –881 mm w.e., which closely agrees with our simulated value for Sangqu Glacier (–919 mm w.e.). This provides strong independent evidence that the 2021–2022 mass loss at Sangqu Glacier was indeed exceptional.

4. Comparison with independent satellite-derived mass balance

   We additionally compared our simulated 2021–2022 annual mass balance for Sangqu Glacier with the multi-source remote sensing product of Dussaillant et al. (2025). That product reports a mass balance of –1278 mm w.e. for Sangqu Glacier in 2021–2022, which is more negative than our modelled value of –919 mm w.e. We acknowledge this numerical difference, which may arise from several factors: differences in spatial representation (the remote sensing product may cover a larger or different elevation range of the glacier), methodological differences in mass balance definition (e.g., inclusion of different glacier zones), and uncertainties in model input data. Despite the difference in absolute magnitude, both independent methods consistently indicate that the 2021–2022 mass loss at Sangqu Glacier was far greater than in previous years (for example, –1278 is about 2.7 times the observed 2019/20 balance of –477 mm w.e.), and together they confirm that 2021–2022 was an extreme mass loss year. This comparison has been added to the updated Fig. 2h, and we discuss the possible reasons for the discrepancy in the revised manuscript.

To directly address the comparability issue:

The annual mass balance values reported for 2019/20 (–477 mm w.e.) and 2020/21 (–607 mm w.e.) are derived from full-year observations (October to October). For 2021/22, because our direct measurements stop in July, the reported annual mass balance (–919 mm w.e.) is a model-reconstructed value for the complete October–September year, calibrated and validated using all available observations. Thus, the comparison between 2021/22 and previous years is between a model-reconstructed full year and observation-based full years, not between an incomplete record and complete records.

We acknowledge that model performance during extreme heatwaves may carry additional uncertainty. However, the close agreement with the in-situ observations from Xiao Dongkemadi Glacier (complete year, also identified as extreme by Zhu et al., 2026) , together with the independent satellite-derived evidence from Dussaillant et al. (2025) – even though the absolute values differ – both point to the same conclusion: 2021–2022 was an extreme mass loss year at Sangqu Glacier. Taken together, these four lines of evidence give us confidence that our reconstructed 2021–2022 mass balance reliably captures the exceptional mass loss.

We have incorporated these improvements into the revised manuscript, with the updated model validation and multi-source comparisons now reflected in Fig. 2 (including the comparison with Dussaillant et al., 2025, in Fig. 2h, and the citation of Zhu et al., 2026).

Dussaillant, I., Hugonnet, R., Huss, M., Berthier, E., Bannwart, J., Paul, F., and Zemp, M.: Annual mass change of the world's glaciers from 1976 to 2024 by temporal downscaling of satellite data with in situ observations, Earth Syst. Sci. Data, 17, 1977–2006, https://doi.org/10.5194/essd-17-1977-2025, 2025.

Zhu, F., Yao, T.-D., Zhu, M.-L., An, L., Wang, P., Huang, G.-F., and Hu, Z.-F.: The role of mid-latitude zonal wind anomalies in extreme glacier mass loss on central Tibetan Plateau, Adv. Clim. Change Res., https://doi.org/10.1016/j.accre.2026.03.010, 2026.

Comment 4: Calibration and validation of the energy-balance model over the same time period: if I understand correctly the model was calibrated (for the precipitation gradient) and validated against the same data. This makes me confused as to why there seems to be a bias between the simulated and measured point mass balances in Fig. 3a? And if this statement is correct, this is a limitation that would need to be raised in the discussion.

Response: We thank the reviewer for raising this important methodological point. We fully understand the concern regarding the potential overlap between model calibration and validation data, as well as the observed point-scale deviations in Fig. 2e (formerly Fig. 3a). We address these issues through the following clarifications and improvements.

1. Clarification of calibration and validation data separation

We acknowledge that the original description was not sufficiently explicit. In the

revised manuscript (Section 4.2), we now clearly distinguish between calibration and validation datasets:

- **Calibration**: The NSGA-II multi-objective optimization algorithm was applied using **glacier-wide annual mass balance observations from 2019–2020 and 2020–2021** (full years, October–October) and **daily albedo measurements from the AWS** to constrain key parameters (snow/ice albedo, turbulent exchange coefficients, etc.). The precipitation gradient was also calibrated within this step.
- **Validation**: Model performance was evaluated using independent datasets **not used in calibration**, including:

  – Point-scale mass balance at different elevations (Fig. 2e);

  – Seasonal mass balance for June–September 2021 (Fig. 2d);

  – Cross-comparison with Xiao Dongkemadi Glacier observations for the 2021–2022 balance year;

  – Comparison with the satellite-derived product of Dussaillant et al. (2025).

2. Quantitative assessment of point-scale model performance

The reviewer correctly noted scatter between simulated and observed point mass balances. **We now provide a comprehensive set of statistical metrics to quantify this performance.** Based on the 24 paired point observations (2019, 2021, 2022), we obtained:

Metric	Value
ME (Mean Error)	37.25 mm w.e.
RMSE (Root Mean Square Error)	288 mm w.e.
MAE (Mean Absolute Error)	234.74 mm w.e.
R (Pearson correlation)	0.9
R^2 (Coefficient of determination)	0.78
NSE (Nash-Sutcliffe Efficiency)	0.78
PBIAS (Percent Bias)	-3.4%

These metrics demonstrate that at the point scale, the model captures the observed spatial variability well (R = 0.9) without systematic over- or under-estimation (ME = +37.25 mm w.e., PBIAS = –3.4%). The RMSE of 288 mm w.e. is comparable to or better than many distributed glacier mass balance models reported in the literature, especially given the challenging high-altitude environment and data scarcity on the Tibetan Plateau.

3. Explanation of point-scale scatter and scale dependency

The remaining scatter (RMSE = 288 mm w.e.) arises primarily from **scale mismatch**: the model operates at 100 m grid cells, while stake measurements represent discrete points. Local variability in snow depth, microtopography, wind redistribution, and ice surface roughness – which are not resolved by the model – contribute to point-scale deviations. Additionally, uncertainties in density assumptions (300 kg m^{-3} for snow, 900 kg m^{-3} for ice) and stake reading errors also play a role.

Critically, this point-scale scatter largely cancels out when aggregated to the glacier-wide scale. At the glacier-wide scale, the RMSE reduces to 101 mm w.e. (Figs. 2f, 2g), which is the primary scale of interest for our extreme event analysis.

4. Acknowledgment of limitation: partial overlap between calibration and validation data

We also wish to be transparent about a broader limitation. Due to the scarcity of in situ observations on the Tibetan Plateau, the same observational dataset (e.g., glacier-wide mass balance for 2019–2021) was used both for calibration and for constraining the model's overall performance, albeit at different aggregation levels (annual vs. seasonal/point). This is a common challenge in glacier modeling under data-scarce conditions. We now explicitly discuss this limitation in Section 6.4, noting that future studies with longer and more comprehensive observational records would allow a more rigorous separation of calibration and validation.

5. Additional validation to strengthen confidence

To further demonstrate model reliability for the 2022 extreme event, we added three independent lines of validation:

- **Seasonal-scale validation** using June–September 2021 mass balance (RMSE = 329 mm w.e.; Fig. 2d). This confirms the model's ability to capture ablation-season dynamics, which are most critical for the 2022 heatwave.
- **Cross-validation with a neighbouring glacier (Xiao Dongkemadi Glacier)** : Independent observations for the complete 2021–2022 balance year give –881 mm w.e., which closely agrees with our simulated value for Sangqu Glacier (–919 mm w.e.). Zhu et al. (2026) also identified that this neighbouring glacier experienced extreme mass loss in 2021–2022.
- **Comparison with the independent satellite-derived product of Dussaillant et al. (2025)** : Their multi-source remote sensing product reports a mass balance of **–1278 mm w.e.** for Sangqu Glacier in 2021–2022. While this absolute value is more negative than our modelled –919 mm w.e. (a difference we acknowledge and discuss in Section 6.4), **both independent methods unequivocally indicate that 2021–2022 mass loss was far greater than previous years**. Thus, despite the numerical difference, the satellite product independently corroborates our central conclusion that 2021–2022 was an extreme mass loss year. This comparison has been added to the updated Fig. 2h.

Together, these three independent validation strategies – seasonal, neighbouring glacier, and satellite-derived – provide strong confidence that our model reliably captures the extreme mass loss of 2021–2022 at Sangqu Glacier.

We have incorporated all these clarifications and improvements into the revised manuscript (Sections 4.2 and 6.4, and Fig. 2). We believe this addresses the reviewer's concerns.

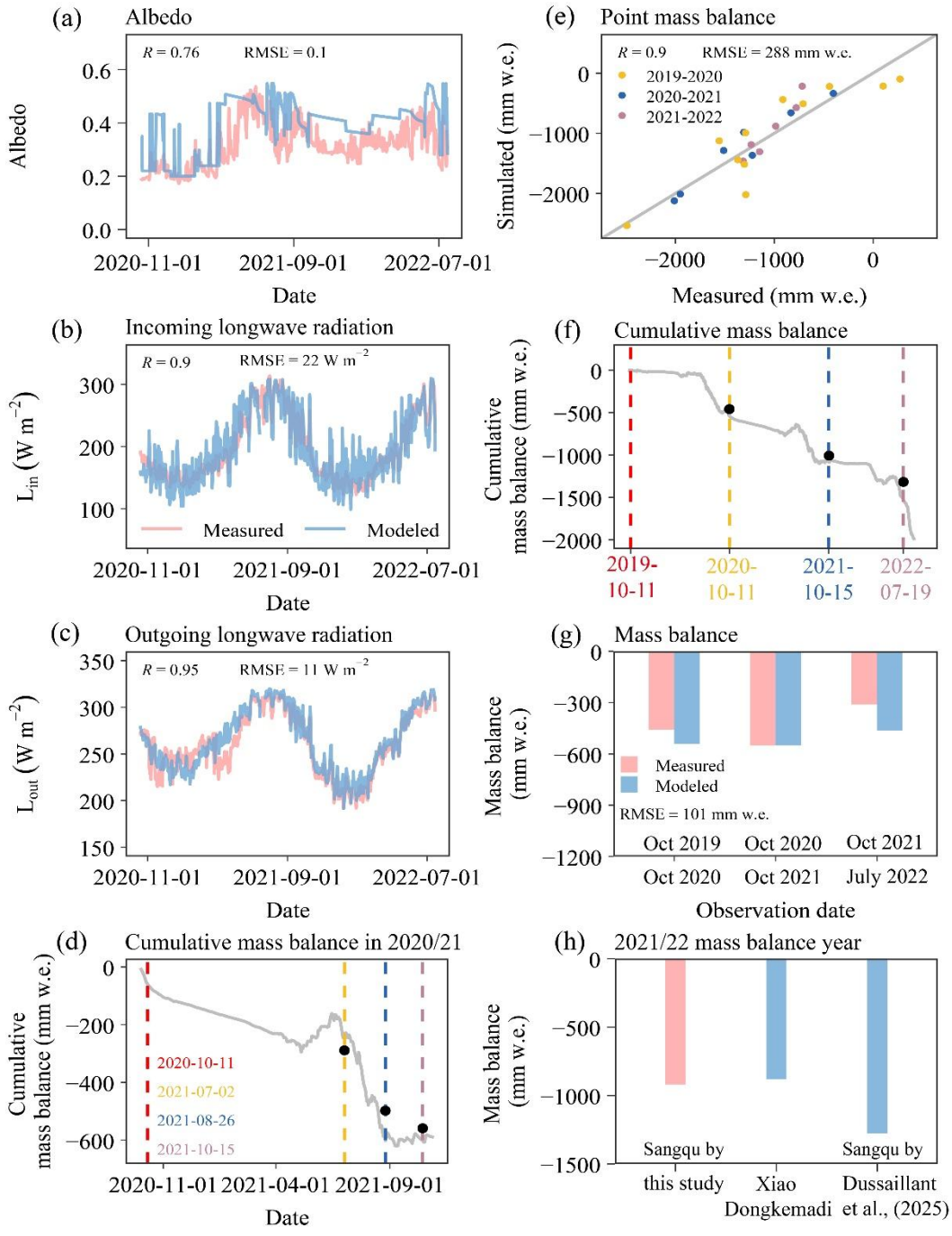


Figure 2. Comparison between observations and simulations of (a) albedo, (b) L_{in}, (c) L_{out}, (d) glacier mass balance from June to September 2021, (e) point mass balance at different elevations, (f) accumulation mass balance, and (g) glacier-wide mass balance for the periods 2019/20, 2020/21, and from October 2021 to July 2022. (h) Comparison of simulated glacier mass balance for Sangqu Glacier in 2021/22 with observed mass balance for Xiao Dongkemadi Glacier and with mass balance data for the same Sangqu Glacier derived from multi-source satellite remote sensing by Dussaillant et al. (2025).

Comment 5: Limited details on the heat transfer in the ice: my expectation is that this glacier is cold-ice, which would have potentially a strong influence on the energy budget. However there is very limited description of how the heat transfer into the ice was calculated. What did you take for the heat transfer coefficients? Was a sensitivity analysis performed? Was there any validation data available?

Response: We thank the reviewer for raising this important point regarding the subsurface heat flux calculation. We agree that the original manuscript lacked sufficient detail on this aspect, and we have substantially expanded the description of the model setup and validation in the revised manuscript.

The calculation of subsurface heat flux has been expanded in Section 4.1, validation against ice temperature measurements has been added in Section 4.2 (Fig. S3), and the glacier's thermal regime description has been supplemented in Section 2.

1. Calculation of subsurface heat flux

In our energy-balance model, the subsurface heat flux (Q_G) is computed by solving the one-dimensional heat conduction equation through the snow/firn/ice column, following the approach of Van Pelt et al. (2012). The effective thermal conductivity (κ) is parameterized as a function of density based on Östin and Andersson (1991) for snow and Sturm et al. (1997) for firn and ice. The heat capacity (c_p) is treated as a function of temperature following Yen (1981).

Effective thermal conductivity (κ) is parameterized as a function of density using the formulation of Sturm et al. (1997):

$$\kappa = 0.138 - 1.01 \times 10^{-3} \rho + 3.233 \times 10^{-6} \rho^2 \quad (1)$$

Specific heat capacity (c_p) follows Yen (1981):

$$c_p = p \cdot (152.2 + 7.122\, T) \quad (2)$$

The model resolves vertical profiles of temperature, density, and liquid water content, allowing for the simulation of heat conduction, meltwater percolation, and refreezing processes.

2. Sensitivity analysis

To assess the sensitivity of glacier-wide mass balance to subsurface heat transfer, we perturbed the effective thermal conductivity (κ) and specific heat capacity (c_p) by ±10% while keeping all other parameters at their calibrated values. The glacier-wide annual mass balance has a sensitivity of −9.9 mm w.e. a^{-1} to effective thermal conductivity and +10.5 mm w.e. a^{-1} to heat capacity. These values are substantially smaller than the sensitivities to surface energy balance parameters (e.g., snow albedo: 123.2 mm w.e. a^{-1}; temperature lapse rate: −141.4 mm w.e. a^{-1}), indicating that subsurface heat transfer has a relatively small impact on the overall mass balance. Detailed sensitivity results are provided in Tables S1–S6.

3. Validation against in-situ ice temperature measurements

To evaluate the model's ability to represent subsurface heat transfer, we compared simulated and observed ice temperatures at depths of 3 m and 5 m within the glacier. The simulations show strong agreement with observations, yielding correlation coefficients of 0.94 and 0.96, respectively. These results have been added to the revised manuscript and are now presented in Figure S3. This validation provides confidence that the model captures the thermal regime of the glacier adequately, even though the glacier is likely polythermal or cold-based in its upper reaches.

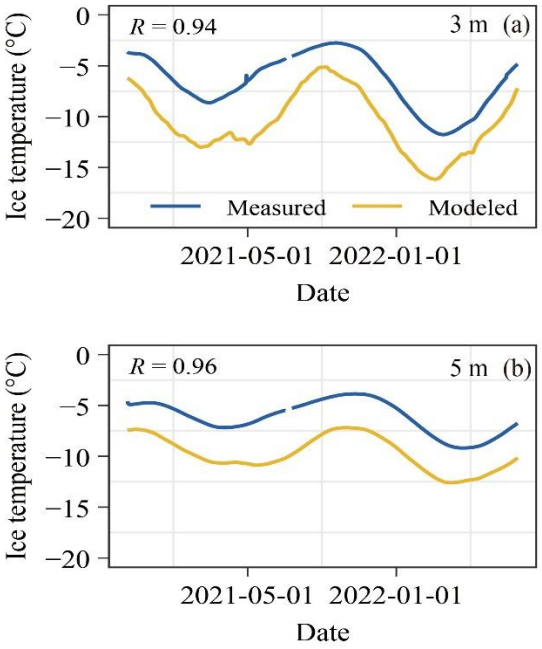


Figure S3. Comparison of observed and simulated ice temperatures at (a) 3 m and (b) 5 m.

4. Additional clarification on glacier thermal regime

Based on the measured ice temperatures and the observed presence of meltwater percolation and refreezing in the ablation zone, we consider Sangqu Glacier to be polythermal, with cold ice in the upper accumulation zone and temperate ice in the ablation zone. This thermal regime is consistent with observations on neighboring Xiao Dongkemadi Glacier, where cold ice temperatures and superimposed ice formation have been documented (Fujita et al., 1996).

Summary of revisions

In summary, we have: (1) completely rewritten Section 4.4 to clarify the dual-threshold framework; (2) substantially expanded Section 6.2 with global comparisons; (3) added multi-strategy validation (parameter optimization, seasonal-scale validation, and neighboring glacier cross-validation) to address the incomplete 2022 record; (4) systematically discussed model uncertainties in Section 6.4; and (5) provided detailed descriptions of subsurface heat transfer calculations with new sensitivity analyses and ice temperature validation. All corresponding updates have been incorporated into the revised manuscript.

Line-by-line comments:

Comment 6: L54: This feels a bit like an oversell. While the study itself is interesting, I didn't find anything absolutely groundbreaking there.

Response: Yes, we revised this sentence. In lines 46 to 49.

"This study proposes a framework for identifying extreme mass loss events in short-term glacier mass balance records and provides insights into glacier mass balance responses to extreme weather and climate events."

Comment 7: L61: I'm not convinced that wildfires fit in the category of 'weather and climate events'.

References would be welcome here.

Response: Yes, we revised this sentence, removed "wildfire," and added references.

Please see lines 55 to 57.

"Extreme glacier melt events directly reflect glacier mass balance responses to extreme weather and climate events, such as heatwaves and droughts (Zhu et al., 2024a; Xu et al., 2024; Zhu et al., 2026; Van Tiel et al., 2026)."

Comment 8: L62: Could you be a bit more specific here? Give some numbers?

Response: Thank you for the suggestion. We have revised the sentence to include specific temporal evidence from Xiao Dongkemadi Glacier, the longest-monitored glacier in the central Tibetan Plateau with observations since 1989. On this glacier, extreme mass loss events were concentrated in 2000–2022 compared to 1989–2000 (Yao et al., 2012; Zhu et al., 2026a). Please see lines 57 to 61.

The revised text is as follows: "Across the Tibetan Plateau and surrounding regions, evidence from Xiao Dongkemadi Glacier, which holds the longest mass balance record in the central Tibetan Plateau with observations since 1989, indicates that extreme mass loss events have been concentrated in 2000–2022, compared to 1989–2000 (Yao et al., 2012; Zhu et al., 2026a)."

Comment 9: L62, l95: Should be 'the Tibetan Plateau'

Response: Yes, we changed it to "the Tibetan Plateau" throughout the text.

Comment 10: L101-103: See general comment, this does not sound realistic.

Response: Thank you for this critical comment. We agree that the original statement may have overstated the feasibility. To address this, we have thoroughly revised Section 4.4 to present our dual-threshold framework with greater precision. The revised section now includes: (1) a clear rationale linking the percentile and mean–standard deviation

approaches; (2) a detailed description of the calibration and validation procedures using long-term glacier records; and (3) an explicit discussion of the framework's limitations when applied to short-term data. We believe this revision provides a more realistic and transparent account of our methodological contribution.

Correspondingly, we have also revised the introductory statement to reflect this more nuanced perspective, which now reads: "Consequently, developing and testing frameworks—such as the dual-threshold approach proposed in this study—for identifying extremes from short-term observations (under ten years) is essential for advancing glacier research in this region." Please see lines 94 to 97.

Comment 11: L110: remind here the period to which this anomaly applies.

Response: Yes, we revised this sentence to: "Hua et al. (2023) demonstrated that a 6.45 °C temperature anomaly, observed in central China over the 21-day period from 6 to 26 August 2022, broke historical records and established a new threshold for defining future record-breaking daily maximum temperatures in the region." Please see lines 103 to 106.

Comment 12: L153: It would be useful to indicate the mean value of annual precipitation at this site. I expect this to be a very dry site, and it would be useful to have these numbers in mind to put the energy budget into context.

Response: Yes, we have added the mean annual precipitation.

Please see lines 151 to 155.

The revised sentence reads: "Based on the combined in situ and ERA5-corrected precipitation data, the mean annual precipitation at the Sangqu Glacier site was 510.5 mm for the period October 2019–September 2022. Over 70% of this precipitation occurs during June–September, coinciding with the warmest season (Zhu et al., 2024a).

Comment 13: L157: Are there any available geodetic mass balance estimates for this glacier?

Response: Yes, based on the data provided by Dussaillant et al. (2025), we added the geodetic mass balance data for Sangqu Glacier from 1957 to 2024. Please see lines 153 to 155.

"Multi-temporal digital elevation model differencing and in situ stake measurements reveal continuous mass loss in this region since the 1990s (Yao et al., 2012; King et al., 2023). Geodetic mass balance estimates from Dussaillant et al. (2025) indicate that Sangqu Glacier lost ~515.4 mm w.e. yr^{-1} over the period 1957–2024, consistent with the regional mass loss pattern."

Comment 14: L161: Photo should be singular Figure 1: Panel d is hard to make sense of. It would make sense to show a DEM hillshade or satellite image of the glacier, or at the very least use a different colorscale.

Response: Based on your comments, we have thoroughly updated Figure 1 again: in Figure 1a, we added the geographical location of the glacier focused on in this study; Figure 1b now uses a DEM hillshade; Figure 1c uses a Sentinel image as the base map with added contour lines to highlight the real topographic features of the Sangqu Glacier. In addition, we have added photographs of the automatic weather station and the precipitation gauge (Figures 1d and 1e). These changes have been incorporated into the updated Figure 1. Thank you for your careful review.

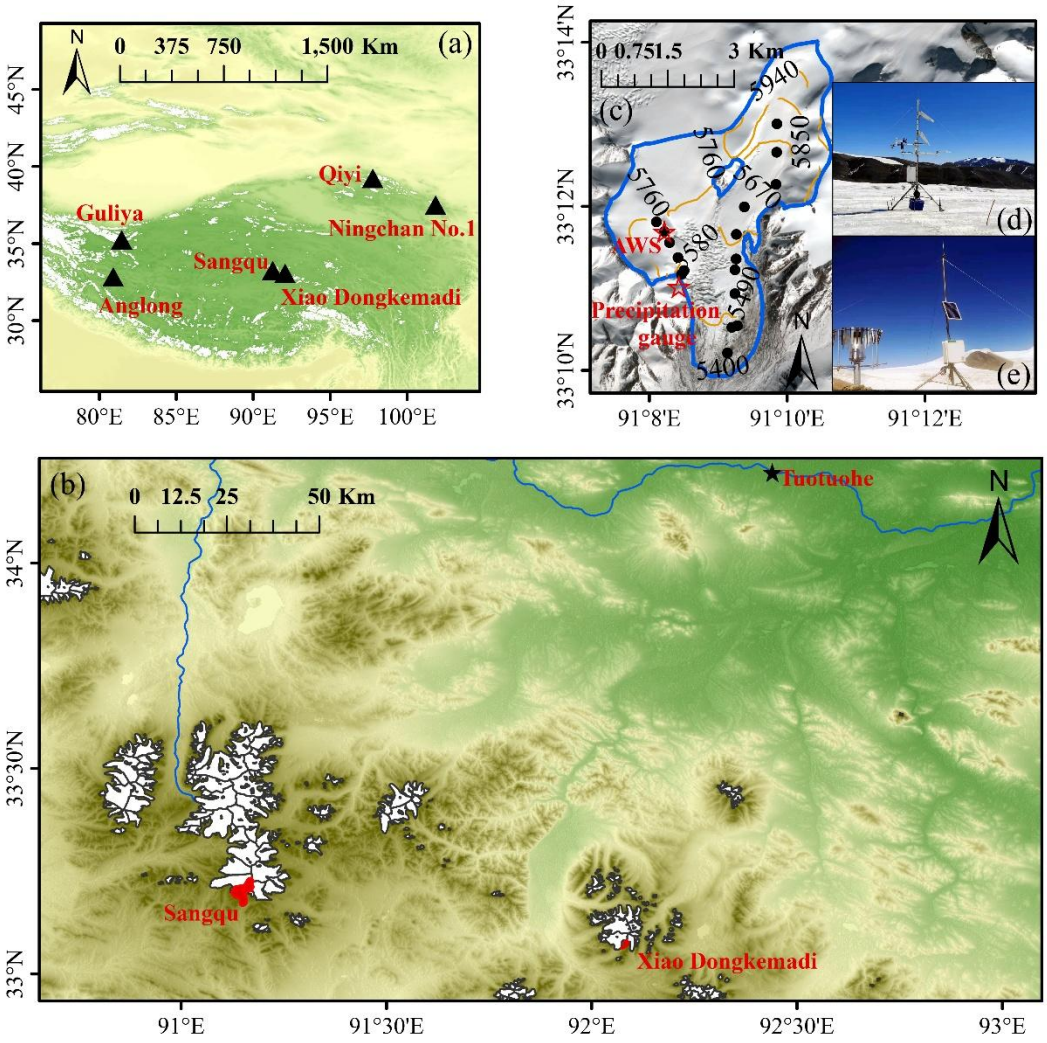


Figure 1. Overview of the study area. (a, b) Geographical location of the glacier on the Tibetan Plateau: (a) regional context; (b) enlarged local map. (c) Sentinel-2 image (2022-12-05) with contour lines showing the topography of Sangqu Glacier. (d, e) Photographs of the AWS and the precipitation gauge on Sangqu Glacier (Photo credit: Fei Zhu). The positions of the AWS, precipitation gauge, and stakes are indicated in panel (c).

Comment 15: L163: Please indicate the location of these precipitation gauges.

Response: Yes, we revised Figure 1 to make the position of the rain gauge more prominent.

Comment 16: L171: greater -> more

Response: Yes, we have revised it.

Comment 17: L174: More details would be welcome here, as well as a picture appended to Fig. 1. Was this precipitation gauge heated? Shielded?

Response: Thank you for this suggestion. We have updated Figure 1 to include a photo of the precipitation gauge, which is now shown in the revised panel (d). Regarding the gauge specifications, it is a Geonor T-200B precipitation gauge equipped with a double-alter shield to minimize wind-induced undercatch. These details have also been clarified in Section 3.1.

Comment 18: L210: 'where' does not fit here

Response: Yes, we deleted "Where" and made revisions elsewhere in the main text.

Comment 19: L241-244: I don't get this. Why do these time series need to be reconstructed if they already exist?

Response: Thank you for this question. The mass balance observations at Sangqu Glacier began in October 2019, whereas the automatic weather station (AWS) was not installed until October 2020. Consequently, to align the meteorological forcing with the full mass balance observation period (October 2019–September 2022), we reconstructed the missing meteorological variables—specifically air temperature, relative humidity, wind speed, and incoming shortwave radiation—using a random forest method that combines ERA5 reanalysis with AWS observations during the overlapping period. Please see lines 255 to 258.

Comment 20: L247: What is a 'three-fold cross validation'? And a 10-fold?

Response: We thank the reviewer for the valuable question. Five-fold cross-validation

is a widely used technique to assess the generalization performance of a model. In our study, we employed it to evaluate the predictive skill of the Random Forest model used for reconstructing near-surface meteorological variables (air temperature, relative humidity, wind speed, and incoming shortwave radiation) at the glacier site.

Specifically, the training dataset was randomly divided into five equal subsets. In each of the five rounds, four subsets were used for model training and the remaining one for validation. This process was repeated three times (three repetitions) with different random partitions to reduce the variability associated with a single data split. The final model performance was then averaged across all validation rounds, yielding robust estimates of predictive accuracy (e.g., R^2 values of 0.99, 0.96, 0.94, and 0.96 for the four variables at the hourly scale). This approach ensures that the reconstructed meteorological forcing used to drive the energy-mass balance model is reliable, especially for periods lacking in situ observations (e.g., August–September 2022).

We have clarified this in the revised manuscript (Section 4.1) to improve methodological transparency. Please see lines 268 to 271.

"To minimize overfitting and obtain robust estimates of model performance, we applied three repetitions of five-fold cross-validation, where the training data were repeatedly split into five subsets, each serving in turn as the validation set."

Comment 21: L251-252: Did you consider looking into seasonal bias corrections?

Response: We thank the reviewer for raising this important question regarding seasonal bias correction. Please see lines 272 to 276.

Our response is as follows:

The downscaling framework inherently captures seasonal signals.

In the current version, we adopted a multivariate Random Forest approach that uses not only ERA5 2 m temperature but also high-level atmospheric variables (temperature, humidity, wind components, and geopotential height at 300 hPa and 500 hPa) as predictors. These variables carry strong seasonal cycles, allowing the model to

implicitly learn and represent seasonal variations during training.

Validation for the ablation season shows no substantial seasonal bias.

To specifically assess model performance during the key study period (summer ablation season), we evaluated the agreement between downscaled and observed meteorological variables for June–September (Fig S8). The R^2 values for air temperature, relative humidity, wind speed, and incoming shortwave radiation during the ablation season are 0.96, 0.94, 0.90, and 0.95, respectively. These results indicate that the downscaled forcing does not exhibit systematic seasonal bias during the summer heatwave period.

Explicit seasonal bias correction was not applied, but we acknowledge its value.

Although we did not perform explicit seasonal bias correction (e.g., seasonally stratified downscaling or post-hoc adjustment), the evidence above supports that the current downscaled data are reliable for analyzing anomalies during extreme events. We agree with the reviewer that explicit correction represents a valuable improvement and have added this as a limitation in Section 6.4, highlighting it as a direction for future work.

We appreciate the reviewer's insightful comment, which has helped us strengthen the methodological rigor of our study.

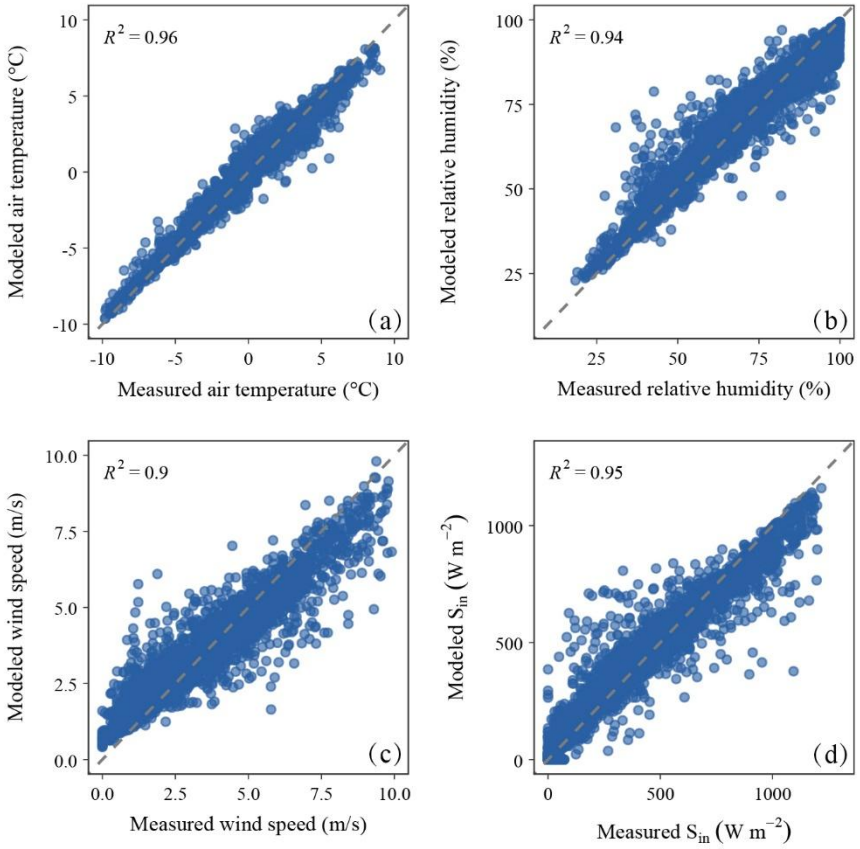


Figure S8. Comparison of observed and ERA5 downscaled (a) air temperature, (b) relative humidity, (c) wind speed, and (d) incoming shortwave radiation at the daily scale during the ablation season (June–September).

Comment 22: L270: How was this lapse-rate determines. It has been shown that the lapse rate on glaciers can strongly vary from the environmental lapse rate.

Response: We thank the reviewer for raising this important point. We fully agree that the lapse rate on glaciers often differs substantially from the environmental lapse rate and can exhibit strong spatiotemporal variability across different glaciers and seasons. Please see lines 287 to 295.

In this study, we did not assume a fixed or environmental lapse rate. Instead, the lapse rate was treated as a key parameter and determined through model calibration. Specifically, using in-situ observations from the automatic weather station (AWS) on the glacier surface and the energy-mass balance model (EBFM), we employed a multi-objective optimization algorithm (NSGA-II) to calibrate the lapse rate together

with other major parameters, ensuring that the simulated glacier mass balance matched the observations as closely as possible.

We also conducted sensitivity tests for the temperature lapse rate (see Sect. 4.2 and Supplementary Table S3). The results show that when the lapse rate varies within ±0.1 °C/1000 m of the calibrated value, the corresponding change in annual glacier-wide mass balance is approximately ±141.4 mm w.e. Although this sensitivity is relatively high, this uncertainty is within the overall mass balance uncertainty (264.3 mm w.e.) and does not alter the main conclusions regarding the extreme melt event and its driving mechanisms. We have presented these sensitivity results in the main text.

Comment 23: L270: 'the parameter' -> 'a calibrated parameter'

Response: Yes, we made revisions.

Comment 14: L271: Where does the albedo come from? Can the albedo at the AWS location be considered representative of the whole glacier.

Response: We thank the reviewer for raising these two important questions regarding the source and spatial representativeness of albedo. Our response is as follows:

(1) Albedo data source

At the automatic weather station located at 5,700 m a.s.l. on Sangqu Glacier, we installed radiation sensors to continuously measure incoming shortwave radiation (S_{in}) and outgoing shortwave radiation (S_{out}). Surface albedo was directly derived as the ratio S_{out} / S_{in} at the AWS site. This observed albedo time series was used for model calibration and validation. Please see lines 328 to 329.

(2) Spatial distribution of albedo in the model

In the energy-mass balance model (EBFM), albedo is not assigned using the point observation alone. Instead, we employed physically based parameterization schemes: snow albedo is parameterized following Oerlemans and Knap (1998) and Bougamont

et al. (2005) as a function of temperature and humidity, while ice albedo follows Mölg et al. (2008) based on dewpoint temperature. These schemes, combined with distributed inputs (topography, solar geometry), allow albedo to be calculated for each grid cell across the glacier, thus capturing spatial heterogeneity. Please see lines 296 to 307.

(3) Calibration and validation

Key parameters in the albedo schemes (e.g., snow albedo decay coefficients, reference ice albedo) were calibrated using the AWS-observed albedo together with multi-objective optimization (NSGA-II) to ensure the best agreement between simulated and observed glacier-wide mass balance. Fig. 2a in the main text compares the simulated daily albedo with observations at the AWS site. The good agreement (R = 0.76, RMSE = 0.1) confirms the reliability of the parameterization at the point scale.

(4) Limitations of point-scale albedo and model handling

We fully acknowledge that a single-point albedo measurement cannot fully represent the spatial albedo distribution across the entire glacier. This is precisely why we adopted a distributed modeling approach: the calibrated albedo parameters are applied to all grid cells, and the model accounts for topographic and radiative effects to simulate spatial variability. To quantify the uncertainty associated with albedo parameterization, we performed sensitivity tests (Sect. 4.2 and Table S2). The sensitivities of annual glacier-wide mass balance to snow albedo and ice albedo parameters are 123.2 mm w.e. and 126.3 mm w.e., respectively, and these uncertainties are included in the total mass balance uncertainty (264.3 mm w.e.). They do not affect our main conclusions regarding the identification and mechanisms of the extreme melt event.

Comment 25: L281: What emissivity value was used?

Response: We thank the reviewer for this detailed question. In the calculation of outgoing longwave radiation, we employed the Stefan-Boltzmann law: $L_{out} = \varepsilon\sigma T_s^4$, where ε is the surface emissivity, σ is the Stefan-Boltzmann constant, and T_s is the surface temperature. Following the parameterization scheme of van Pelt et al. (2012)

used in their Arctic glacier energy-mass balance model, we set ε = 1. Please see lines 296 to 306.

To validate this parameterization, we compared the simulated L_{out} with AWS observations (Fig. 2c in the main text). The good agreement (R = 0.95, RMSE = 11 W m^{-2}) confirms the applicability of this emissivity value in our study.

Comment 26: L281: What was used for the surface roughness and why?

Response: Regarding the parameterization of sensible and latent heat fluxes, we clarify as follows:

In this study, we did not employ the conventional aerodynamic method that relies on surface roughness for calculating sensible heat flux (Q_H) and latent heat flux (Q_L). Instead, we adopted the parameterization scheme of Klok and Oerlemans (2002), which uses bulk transfer coefficients to directly compute the turbulent fluxes, thereby avoiding explicit dependence on surface roughness length and stability correction functions. Please see lines 220 to 228.

We acknowledge that the bulk transfer coefficient method simplifies the representation of near-surface turbulent processes and may not fully capture stability effects. To assess the associated uncertainty, we conducted sensitivity tests for the exchange coefficients (Sect. 4.2 and Table S2), which show that the impact on annual mass balance is within acceptable limits. Moreover, the validation of energy balance components against AWS observations (Fig. 2) supports the adequacy of this parameterization for our study.

Comment 27: L286-288: I fail to see the relevance here, and in this entire paragraph – everything should be calculated relative to the AWS elevation?

Response: We deleted content that was not very relevant to this paper.

Comment 28: Figure 2: What does the color scale correspond to?

Response: We redrew Figure 2 and used a new color scheme.

Comment 29: L326: How were the glacier-wide values obtained? Since your model does not seem to account for any type of snow redistribution process, I wonder if a validation against a glacier-wide mass balance extrapolated from stake data is useful here.

Response: We thank the reviewer for these two interconnected and insightful questions. Our responses are as follows:

(1) How glacier-wide mass balance values were obtained

The glacier-wide mass balance in this study was simulated using the distributed energy-mass balance model (EBFM). The glacier was discretized into 100 m × 100 m grid cells, and energy and mass balances (including melt, snowfall, sublimation/evaporation, and refreezing) computed for each cell independently. The glacier-wide mass balance was then obtained by area-weighted averaging. Please see lines 185–193 and 231-233.

(2) Snow redistribution processes and the validity of validation against glacier-wide mass balance

We fully agree with the reviewer that snow redistribution processes (e.g., wind-driven snow transport, avalanches) can significantly influence the spatial distribution of mass balance on many glaciers, especially in complex terrain or regions with strong winds. We acknowledge that the current EBFM model does not explicitly account for snow redistribution, which is indeed an important simplification.

Nevertheless, we consider the validation against glacier-wide mass balance derived from stake measurements to be useful for the following reasons:

Glacier-wide mass balance is a core target for distributed model validation. Distributed models are designed to simulate spatial patterns, but validation against the

observed glacier-wide mass balance remains a critical test of overall model performance. In our study, we validated the model against both glacier-wide mass balance (Fig. 2f, 2g) and point-scale mass balance at different elevations (Fig. 2e). The point-scale comparison provides additional constraint on the elevation-dependent performance of the model.

We employed a multi-objective optimization algorithm (NSGA-II) to calibrate the model using both point-scale (stake) and glacier-scale (total mass balance) observations. Although snow redistribution is not explicitly represented, the calibration using point observations can indirectly capture its average effect on the spatial distribution.

The limitations of this omission have been added in lines 840–854.

Comment 30: L368-369: missing references here.

Response: Yes, we added references. Please see lines 599–600.

"This finding is further substantiated by two independent datasets (Hugonnet et al., 2021; Dussaillant et al., 2025)."

Comment 31: L388: How many are there?

Response: We rewrote this sentence. Please see lines 570–580.

"To evaluate the robustness of our proposed dual-threshold framework across different climatic regimes and record lengths, we applied it to five glaciers with long-term mass balance observations on the Tibetan Plateau (Fig 1a, Fig. 5, and Fig. S4)."

Comment 32: L432: I don't think I've seen any mention of the spatial resolution at which the model was run? Were snow redistribution processes (from avalanches and wind) accounted for in any way?

Response: We thank the reviewer for raising these two points. Regarding the spatial resolution, the distributed energy-mass balance model (EBFM) was run at a grid

spacing of 100 m. This information has been added to the revised manuscript in Sect. 4.1. Please see lines 252–255.

Regarding snow redistribution processes (e.g., wind-driven transport, avalanches), we confirm that the current EBFM model (van Pelt et al., 2012) does not explicitly account for these processes. We have added a clear statement acknowledging this simplification in Sect. 6.4 (Research Limitations) of the revised manuscript, along with a discussion of the associated uncertainties. We appreciate the reviewer's thorough reading, which has helped us better clarify the model's applicability and limitations.

Comment 33: Figure 4: It would be helpful to plot the net energy balance on top of this bar plot. Also, were there any altitudinal variations?

Response: We thank the reviewer for this constructive suggestion. We have revised Figure 4 by adding the net energy balance (the sum of net shortwave radiation, net longwave radiation, sensible heat flux, latent heat flux, and glacier heat flux) as an overlay on the bar plot, while retaining the individual energy balance components. The revised figure better illustrates the contribution of each component to the melt energy.

Regarding altitudinal variations, we agree with the reviewer that energy balance components exhibit clear elevation-dependent patterns. Accordingly, we have added a supplementary figure (Fig. S9) showing the altitudinal distribution of melt energy across different elevation bands. A brief summary of these results has been included in Sect. 5.1 of the revised manuscript.

We appreciate the reviewer's careful reading, which has helped us better present the spatial heterogeneity of the glacier surface energy balance.

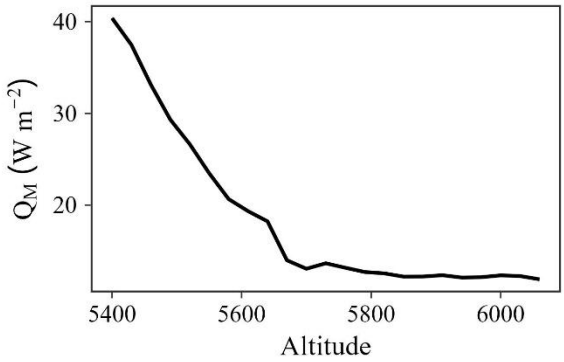


Figure S9. Melt energy versus elevation for the 2021–2022 mass balance year.

Comment 34: L522: why were these glaciers chosen specifically?

Response: We thank the reviewer for this question. The selection of these glaciers was based on the following considerations:

Climatic representativeness: The Tibetan Plateau and its surrounding regions are characterized by harsh environmental conditions, making long-term continuous glacier mass balance observations extremely challenging. The selected glaciers are located in the central (Xiao Dongkemadi Glacier) , northern (Qiyi Glacier, Ningchan No. 1 Glacier) , and western (Guliya Glacier, Xiao Anglong Glacier) Tibetan Plateau, covering the major climatic regimes of the region—from the relatively humid central plateau to the hyper-arid western plateau. This distribution allows us to capture the diversity of glacier responses to extreme events under different climatic backgrounds.

Spatial coverage and comparative analysis: By selecting glaciers distributed across different regions of the plateau, we aim to evaluate the transferability and robustness of our dual-threshold framework under varying climatic conditions and glacier types, thereby enhancing the generalizability of the methodological approach.

We have added a corresponding explanation in Sect. 5.2 of the revised manuscript to clarify the selection criteria and the climatic representativeness of these glaciers. Please see lines 570–580.We appreciate the reviewer's careful reading, which has helped us better articulate the rationale behind our study design.

Comment 35: L541-542: references missing.

Response: Yes, we have added the relevant references. Please see lines 598–605.

Comment 36: L738: this whole paragraph sounds like a repetition of what was stated above.

Response: We thank the reviewer for pointing this out. We agree that this paragraph largely repeated content presented earlier. Accordingly, we have removed this paragraph from the revised manuscript.

In addition, we have substantially rewritten Sect. 6.1 (Atmospheric circulation) to provide a more focused analysis of the circulation anomalies associated with the 2022 extreme melt event and their driving mechanisms, avoiding redundancy with the preceding sections. The revised content can be found in Sect. 6.1 of the updated manuscript.

We appreciate the reviewer's careful reading, which has helped us improve the conciseness and logical flow of the manuscript.

Comment 37: L755: I am missing a discussion of the limitations from the energy-balance model and its application.

Response: We thank the reviewer for raising this important point. We agree that a thorough discussion of the limitations of the energy-balance model and its application is essential for a transparent presentation of the study.

Accordingly, we have added a new Section 6.4 in the revised manuscript, where we systematically discuss the main limitations of the energy-mass balance model (EBFM) used in this study, including:

- Uncertainties associated with spatial extrapolation of meteorological variables from a single AWS.
- Simplified parameterization schemes for key parameters (e.g., albedo, turbulent

exchange coefficients).

- The absence of snow redistribution processes (wind-driven transport, avalanches).
- The assumption of fixed glacier geometry and its implications for long-term simulations.
- The impact of sparse observational data on model calibration and validation.

We also assess how these limitations affect the main conclusions of the study and outline directions for future improvements. Please refer to Sect. 6.4 of the revised manuscript for details.

We appreciate the reviewer's valuable suggestion, which has significantly improved the transparency and completeness of the discussion on model applicability.