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 S_{HI} 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 SCE-UA algorithm was first used to calibrate parameters related to incoming longwave radiation (using RMSE as the objective function). Subsequently, the 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.

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.

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.

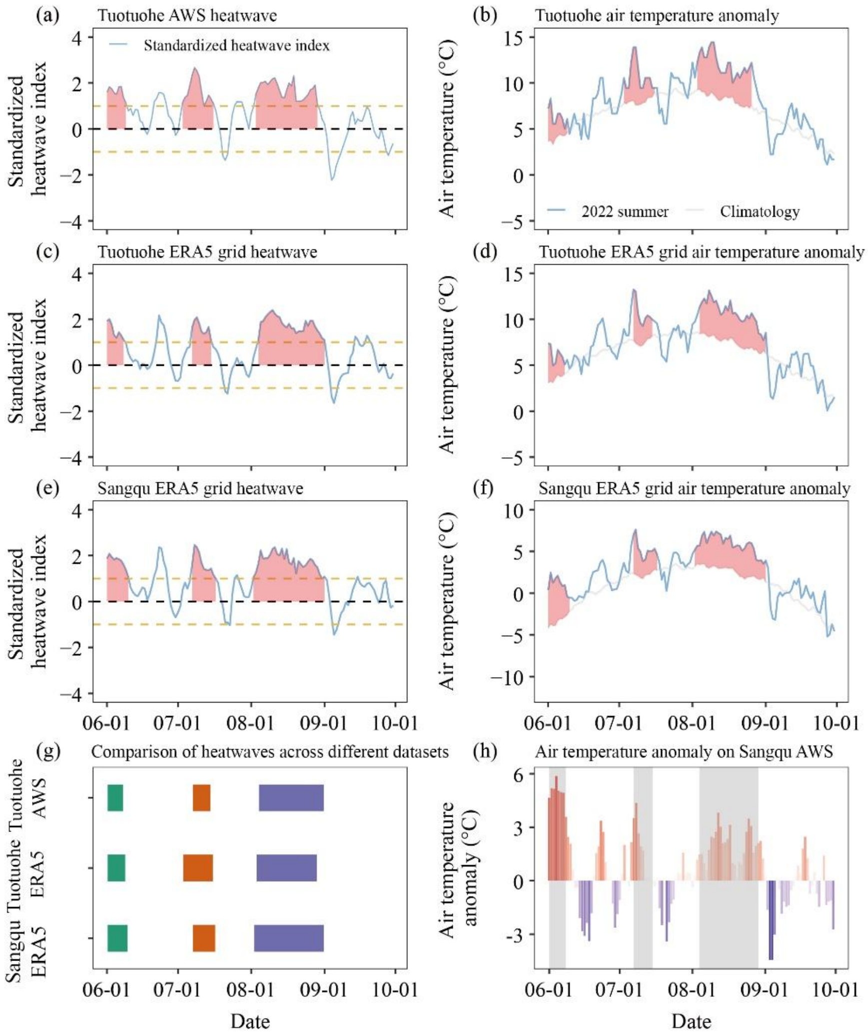

Figure 3. Identification of heatwave events from June to September 2022. Air temperature anomalies were calculated relative to the 1992–2021 climatology. (a) Heatwave events and (b) temperature anomalies at the Tuotuohe meteorological station; (c) and (d) same as (a) and (b) but for the ERA5 grid cell containing the Tuotuohe meteorological station; (e) and (f) same as (a) and (b) but for the ERA5 grid cell containing the Sangqu Glacier; (g) comparison of heatwave events identified from (a), (c), and (e); (h) summer (June–September) temperature anomalies at the 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–S6), 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–S6.

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.

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 (S_{HI}) 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.