COMMENTS FROM EDITORS AND REVIEWERS

This study by Xu and others evaluates the suitability of five different turbulent flux models on a glacier on the Tibetan Plateau. This is done by comparing modeled and observed turbulent fluxes over a five-month period in 2023. In a next step, the turbulent flux models are included in a mass balance model to test sensitivity of the mass balance to the choice of turbulent flux model. The manuscript is well written, presents new data, and contains useful new insights in performance of different turbulent flux models on a glacier on the Tibetan Plateau. This information will be valuable e.g. for future applications of mass balance modelling in the area. In my opinion the manuscript needs major revisions though before it can be published. One concern I have is that the description of the methods is incomplete. For example, it remains entirely unclear how the calibration of the mass balance model with the different turbulent flux models was done. Furthermore, I think it is a missed opportunity that no attempt has been made to calibrate e.g. roughness lengths to improve agreement between modeled and observed turbulent fluxes. This would give useful insight in what parameter settings of the turbulent flux models are most suitable to use on this glacier (and possibly other glaciers on the Tibetan Plateau). The calibrated turbulent flux models could then be used in the mass balance model to assess mass balance sensitivity to the turbulent flux formulations. Maybe some of this was in fact done; if so, it just needs to be described more clearly. Detailed comments are given below.

Answer: We sincerely thank the Editor and the Reviewers for providing us with the opportunity to revise our manuscript. We are also grateful to the Reviewers for their constructive and valuable comments, which have greatly helped us improve the quality of the manuscript. All comments have been carefully addressed in the revised version. A point-by-point response to the comments from the Reviewer (highlighted in blue) is attached to this submission, along with the corresponding revisions in the manuscript (marked in brown). We believe that these revisions and clarifications have significantly improved the manuscript and made it suitable for publication.

Specific comments:

Missing parameter values & errors in equations:

Section 3: For many model-specific and physical constants (e.g. C_{tub}, C_{tub2}, $L_{s/f}$, C_p, C_h, C_H etc) values are not specified in the manuscript (unless I missed it). Please add a table with all these parameters and their values, units and, where relevant, the source.

Answer: Thank you for this important suggestion. We agree that the original manuscript did not provide a sufficiently complete and centralized list of constants and scheme-specific parameters, which reduced reproducibility. In the revised manuscript, we added a dedicated parameter table that consolidates (i) physical constants, (ii) model-specific coefficients, (iii) units, and (iv) literature sources (or notes indicating calibrated vs. prescribed values). We also harmonized symbol usage across the table, equations, and the main text to avoid ambiguity (e.g., consistent notation for C_p, κ, and transfer coefficients). This revision ensures that readers can reproduce our turbulent-flux calculations and the downstream SEB/MB experiments without needing to infer missing values.

Following this revision, the parameter table has been added at **line 347** in the revised manuscript, and it includes the following contents:

Table 2: Summary of model-specific parameters and physical constants adopted in this study.

Parameter	Description	Value	Unit	Source
C_{turb}	dimensionless empirical constants	0.0001	–	this study
C_{turb2}	dimensionless empirical constants	0.007	–	this study
γ	potential temperature gradient	0.005	K m^{-1}	Oerlemans and Grisogono, 2002
P_r	Prandtl number	0.71	–	Standard value
$L_{s/f}$	latent heat of sublimation/evaporation	$L_f = 2.514\times10^6$ $L_s = 2.849\times10^6$	J kg^{-1}	Standard value
Cp	heat capacity of air	1005	J kg^{-1} K^{-1}	Standard constant
C_h	turbulent exchange coefficient (recalibrated in this study)	0.0005	–	this study
C_h	turbulent exchange coefficient (reference value)	0.00127	–	Oerlemans, 2000
κ	von Karman constant	0.4	–	Standard constant
P_0	mean atmospheric pressure at sea level	101325	Pa	Standard value

z_0	Aerodynamic roughness length (literature value)	0.003	m	Essery and Etchevers, 2004
z_{0m}	Aerodynamic roughness length (literature value)	0.01	m	Hock and Holmgren, 2005

Equation 1: Division by pressure is missing. Please check that it is not missing in your code. See for example Radic et al. (2017). I think the corresponding equation in Oerlemans and Grisogono (2002) may also miss this.

Answer: We appreciate the reviewer for identifying this issue. We re-checked both the manuscript and our implementation. While the computational implementation follows the physically consistent bulk-aerodynamic formulation, the equation as presented in the original manuscript could indeed be interpreted as missing a pressure-related term depending on the adopted variable form (mixing ratio/specific humidity/density-based). We therefore corrected Eq. (1) and revised the accompanying variable definitions to explicitly show the pressure (or equivalent density-based) dependence consistent with standard formulations (e.g., Radić et al., 2017). The specific modification is implemented in **Eq. (7) at line 271** of the revised manuscript:

$$LE = \frac{0.622}{P}\rho L_{s/f} C_{kat}(e_a - e_s) \quad (7)$$

Equation 2: The "0.622" should not be there and instead air density should be in the equation. Also here, see Radic et al. (2017), and please check that your code is correct.

Answer: Thank you for this detailed technical correction. We agree that the original manuscript presentation was confusing and could be interpreted incorrectly. In the revised manuscript, we rewrote Eq. (2) (now Eq. (8)) using the standard density-based formulation that explicitly includes air density ρ, ensuring dimensional consistency. We also clarified the role of $\epsilon = 0.622$(molecular weight ratio) and avoided placing it in the equation where it would be redundant or incorrect. Finally, we verified that the revised equation and definitions match the implemented code to ensure internal consistency. The revised formulation is provided as **Eq. (8)** at **line 272** of the revised manuscript, given by:

$$H = \rho C_P C_{kat}(T_a - T_s) \quad (8)$$

Discrepancies between modeled and observed turbulent fluxes:

L139-140: "the roughness lengths are computed dynamically". Does this mean that roughness lengths are time-dependent? It is important to add a bit more information about how the roughness lengths are determined and add the relevant equations from Andreas (1987). The information is relevant in order to be able to judge whether some of the discrepancies between modeled and observed fluxes (Fig. 4) could result from this.

Answer: We acknowledge that the phrase "computed dynamically" was insufficiently explained in the original manuscript. In the revised Methods section, we clarify that the aerodynamic roughness length (z_{0m}) was optimized within the range constrained by the eddy-covariance observations and was subsequently treated as a constant in the calculations. In contrast, the scalar roughness lengths are computed dynamically. Specifically, we now explicitly document the formulations used to estimate scalar roughness lengths, drawing on the parameterizations proposed by Andreas (1987) and Smeets and van den Broeke (2008). The resulting values of z_{0m}, $z_{0}q$, and $z_{0}t$ are consistently applied within the turbulent flux schemes to compute latent and sensible heat fluxes. These additions are essential for assessing whether the representation of roughness lengths contributes to the discrepancies between modeled and observed turbulent fluxes.

For details, please refer to **Section 3.1**, **lines 200–238** of the revised manuscript:

"**3 Methods**

3.1 The algorithm and quality control for roughness lengths

This study uses the roughness lengths of momentum derived from EC measurements to calculate turbulent fluxes. The EC data, with a 30-minute temporal resolution, provide three-dimensional wind speed components, which are then used to indirectly calculate friction velocity (u∗) and Obukhov length (L) as follows:

$$u_* = \left(\overline{u'w'}^2 + \overline{v'w'}^2\right), \quad (1)$$

$$L = -\frac{T_v u_*^3}{g\kappa \overline{w'T_v'}}, \quad (2)$$

where $u'$and v'represent the fluctuations in the horizontal wind components around their 30-minute mean values, w'denotes the fluctuation in the vertical wind component, T_v is the 30-min averaged virtual temperature (K), g is the acceleration due to gravity (9.8 m s^{-2}), and κ is the von Kármán

constant (0.4). The roughness length for momentum (z_{0m}) is then retrieved using friction velocity and L, according to the following equation:

$$z_{0m} = \exp\left[-\kappa\frac{u}{u_*} - \Psi_v\left(\frac{z}{L}\right)\right]z, \tag{3}$$

Here, z_{0m} is the momentum roughness length, and u represents the 30-min averaged wind speed. The term Ψ_v is the stability correction function for momentum. Under stable conditions ($\frac{z}{L} > 0$), it follows the formulation of Holtslag and De Bruin (1988), while under unstable conditions ($\frac{z}{L} < 0$), it follows the Businger-Dyer expression described by Paulson (1970).

Stable conditions:

$$-\Psi_v\left(\frac{z}{L}\right) = \frac{z}{L} + b\left(\frac{z}{L} - \frac{c}{d}\right)\exp\left(-\frac{dz}{L}\right) + \frac{bc}{d}, \tag{4}$$

Unstable conditions:

$$x = \left(1 - 16\frac{z}{L}\right)^{0.25}, \tag{5}$$

$$\Psi_v\left(\frac{z}{L}\right) = \log\left(\frac{1+x}{2}\cdot\left(\frac{1+x}{2}\right)^2\right) - 2\cdot\arctan(x) + \frac{\pi}{2}, \tag{6}$$

where the constants are a = 1, b = 2/3, c = 5, and d = 0.35; z is the measurement height. To provide a reference for future studies without reliance on EC data, we parameterized the scalar roughness lengths (z_{0t} and z_{0q}) as functions of the momentum roughness length z_{0m}, which was derived from EC data. This parameterization follows the methods proposed by Andreas (1987) and Smeets and van den Broeke (2008), which have been shown to provide accurate estimates of scalar roughness lengths over glaciers. Hereafter, these two parameterization schemes are referred to as A87 and SvdB, respectively. In this study, we performed quality control on the 30-minute z_{0m} values following the procedures outlined by Conway and Cullen (2013) and Li et al. (2016). Detailed processing steps can be found in Radić et al. (2017). After filtering, 183 valid z_{0m} data points remained, with inferred z_{0m} values ranging from 1.14 × 10^{-5} m to 2.08 × 10^{-2} m. We finally selected the median of the quality-controlled z_{0m} values (1.2 × 10^{-4} m) as the representative roughness length for evaluating the performance of the different turbulent flux parameterization schemes. For the measured sensible and latent heat fluxes, the same general quality control procedures as those applied to z_{0m} were adopted. However, with respect to the neutrality and wind speed criteria, flux calculations were allowed to include data satisfying $|z/L| < 2$ and $u > 1$ m s^{-1}. After screening, 592 turbulent flux data points were retained for subsequent model calculations and evaluation."

L308-309: "However, ... observed variability". It seems that the underestimation of variability of modelled turbulent fluxes applies over the whole period and for different turbulent flux models. Is this underestimation of temporal variability something that other studies also found (on other glaciers)? Could it be related to the calculation of the roughness lengths? How uncertain are the turbulent flux measurements themselves? Is it possible to assign a value to this uncertainty?

Answer: We thank the reviewer for this insightful and multi-faceted comment. We address the points regarding (i) the general underestimation of temporal variability, (ii) the role of roughness-length representation, and (iii) the uncertainty of EC turbulent-flux measurements as follows.

First, we agree that the underestimation of temporal variability in modeled turbulent fluxes is not limited to a specific scheme or period, but appears consistently across the full observation period and across different turbulent-flux formulations. Similar behavior has been reported in previous studies on other glacierized regions, although the number of long-term eddy-covariance (EC) datasets over glaciers remains very limited. For example, Radić et al. (2017) evaluated turbulent-flux parameterizations in the Cariboo Mountains, British Columbia, Canada, and found that H was generally underestimated. In contrast, because LE exhibits both positive and negative values, its modeling performance cannot be unambiguously assessed using statistical metrics or scatter plots alone.

Table 2. Comparison between modeled and OPEC-derived sensible (Q_H) and latent (Q_E) heat fluxes, expressed as root mean square error (RMSE), mean bias error (modeled minus observed; MBE), and Pearson correlation coefficient (r) for 2012 and 2010 (values in parentheses) observational periods, given for a set of six bulk methods and their variants (see text).

Method		Q_H			Q_E		
		RMSE $W\,m^{-2}$	MBE $W\,m^{-2}$	r	RMSE $W\,m^{-2}$	MBE $W\,m^{-2}$	r
(1)	C_{log}	29.6 (28.7)	15.4 (20.6)	0.80 (0.57)	14.3 (11.5)	7.3 (9.7)	0.89 (0.65)
	$C_{log}\ u_*$	15.2 (11.4)	−2.3 (−1.4)	0.87 (0.77)	6.7 (2.9)	1.5 (0.6)	0.94 (0.86)
	$C_{log}\ K_{Int}$	16.4 (13.0)	−4.4 (−1.7)	0.82 (0.66)	7.4 (4.4)	1.0 (1.3)	0.92 (0.73)
(2)	C_{Ri_b}	26.0 (23.7)	0.0 (3.1)	0.74 (0.46)	13.9 (8.7)	7.2 (4.5)	0.82 (0.43)
	$C_{Ri_b}\ u_*$	26.6 (21.5)	−21.9 (−17.0)	0.84 (0.65)	9.8 (5.0)	2.8 (−2.7)	0.88 (0.65)
(3)	C_{M-O}	26.9 (25.3)	7.3 (11.5)	0.77 (0.51)	14.0 (10.0)	7.5 (7.1)	0.85 (0.54)
	$C_{M-O}\ u_*$	20.3 (13.7)	−13.6 (−8.7)	0.84 (0.79)	6.5 (3.0)	1.2 (−0.6)	0.94 (0.85)
	$C_{M-O}\ \frac{z}{L}$	20.1 (18.3)	−5.3 (1.4)	0.78 (0.53)	9.9 (6.4)	4.3 (3.0)	0.88 (0.57)
	$C_{M-O}\ \frac{z}{L}\ u_*$	22.3 (15.1)	−15.0 (−9.7)	0.80 (0.74)	7.0 (3.3)	0.9 (−1.0)	0.93 (0.82)
	$C_{M-O}\ u_*\ Pr$	16.5 (12.5)	−8.0 (−6.2)	0.86 (0.78)	6.9 (3.3)	1.7 (0.4)	0.94 (0.81)
(4)	C_{SR}	23.8 (28.3)	5.3 (16.1)	0.78 (0.52)	12.1 (11.7)	6.6 (9.0)	0.86 (0.54)
	$C_{SR}\ u_*$	18.7 (10.6)	−11.7 (−2.3)	0.85 (0.80)	6.4 (3.3)	1.3 (1.6)	0.94 (0.86)
	$C_{SR}\ \frac{z}{L}$	18.7 (19.0)	−6.3 (6.2)	0.80 (0.54)	8.9 (7.4)	3.6 (4.7)	0.90 (0.61)
	$C_{SR}\ \frac{z}{L}\ u_*$	20.8 (12.2)	−13.1 (−2.8)	0.81 (0.73)	6.9 (3.5)	1.0 (1.2)	0.93 (0.82)
(5)	C_{kat}	19.5 (17.9)	−4.7 (1.4)	0.78 (0.44)	10.3 (5.2)	4.3 (3.2)	0.87 (0.76)
(6)	K_{Int}	39.8 (31.5)	−2.4 (6.8)	0.15 (0.20)	22.4 (13.6)	−6.2 (4.7)	0.74 (0.46)
	$K_{Int}\ Pr$	37.7 (30.6)	2.9 (8.7)	0.26 (0.23)	22.2 (13.6)	−6.2 (5.5)	0.77 (0.45)

Likewise, Haven et al. (2025) reported a systematic underestimation of sensible heat flux magnitude over the Greenland Ice Sheet, under a sign convention in which fluxes directed from the glacier to the atmosphere are defined as positive.

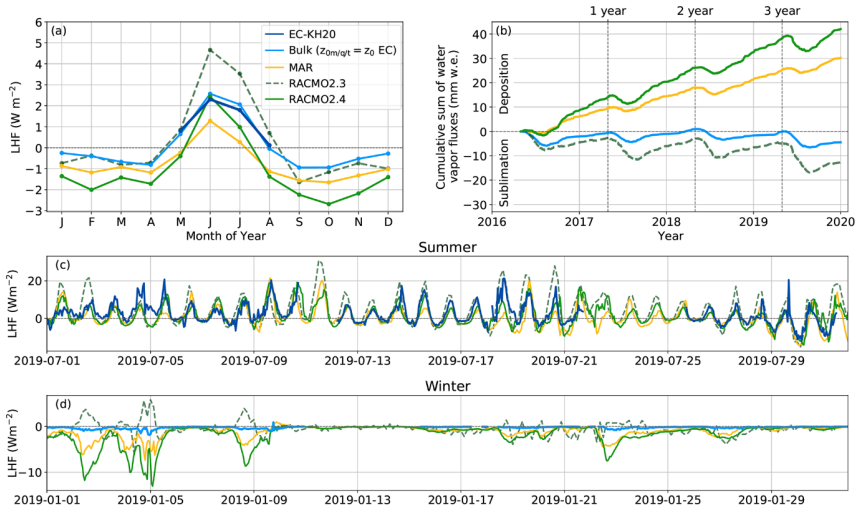


Figure 6. **(a)** Seasonal cycle of the LHF from May 2016 until the end of 2019. **(b)** Cumulative LHF expressed in mm water equivalent (mm w.e.) (note the fluxes follow the SMB sign convention in this panel). Panels **(c)** and **(d)** show examples of time series of the LHF during summer and winter, respectively. LHFs are shown based on observations from the EC-KH20 and from calculations using the bulk method using observations from the PROMICE AWS. Simulated LHFs are based on the RCM's MAR, RACMO2.3 and RACMO2.4. The observational time series from the EC-KH20 is used due to its longer available record.

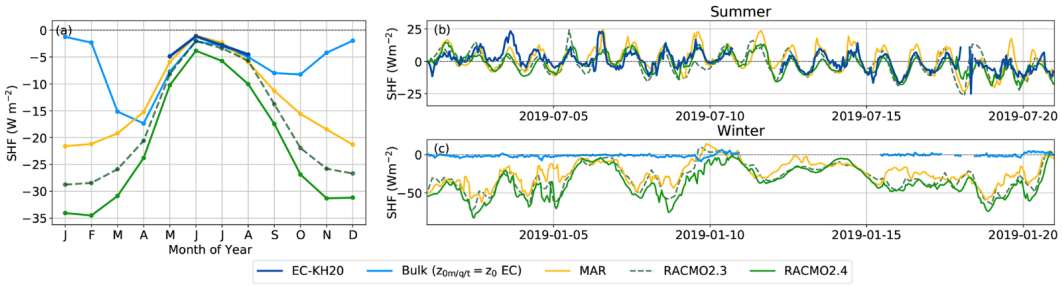


Figure 7. **(a)** Seasonal cycle of the SHF from May 2016 until the end of 2019. Panel **(b)** and **(c)** examples of SHF time series for a 21 d period in summer and winter, respectively. SHFs are shown based on observations from the EC-KH20 and from calculations using the bulk method using observations from the PROMICE AWS. Simulated SHFs are based on the RCM's MAR, RACMO2.3 and RACMO2.4. The observational time series from the EC-KH20 is used due to its longer available record.

In addition, Guo et al. (2011), who compared different roughness-length parameterizations, found a slight underestimation of H. Together, these studies suggest that muted temporal variability and systematic bias are common features in bulk turbulent-flux modeling over snow and ice surfaces, rather than artifacts unique to the present study.

Second, we agree that roughness-length representation is a likely contributor to the underestimated variability. In many previous studies, roughness lengths are prescribed as fixed values or derived from empirical relationships that may not fully capture temporal changes in surface conditions. Motivated by this concern, we substantially revised the manuscript by incorporating turbulent-flux calculations using aerodynamic roughness lengths derived from EC observations. Specifically, we (1) applied aerodynamic

roughness lengths inverted directly from EC observations, (2) tested multiple scalar roughness-length parameterizations. These additions allow us to better isolate the influence of roughness-length representation on both the magnitude and variability of modeled turbulent fluxes, and they are now explicitly discussed in the revised Results **(see Lines 200–238 of the revised manuscript)**, (1) applied aerodynamic roughness lengths inverted directly from EC observations:

"**3 Methods**

3.1 The algorithm and quality control for roughness lengths

This study uses the roughness lengths of momentum derived from EC measurements to calculate turbulent fluxes. The EC data, with a 30-minute temporal resolution, provide three-dimensional wind speed components, which are then used to indirectly calculate friction velocity (u∗) and Obukhov length (L) as follows:

$$u_* = \left(\overline{u'w'}^2 + \overline{v'w'}^2\right), \quad (1)$$

$$L = -\frac{T_v u_*^3}{g\kappa \overline{w'T_v'}}, \quad (2)$$

where $u'$and v'represent the fluctuations in the horizontal wind components around their 30-minute mean values, w'denotes the fluctuation in the vertical wind component, T_v is the 30-min averaged virtual temperature (K), g is the acceleration due to gravity (9.8 m s^{-2}), and κ is the von Kármán constant (0.4). The roughness length for momentum (z_{0m}) is then retrieved using friction velocity and L, according to the following equation:

$$z_{0m} = \exp\left[-\kappa \frac{u}{u_*} - \Psi_v\left(\frac{z}{L}\right)\right] z, \quad (3)$$

Here, z_{0m} is the momentum roughness length, and u represents the 30-min averaged wind speed. The term Ψ_v is the stability correction function for momentum. Under stable conditions ($\frac{z}{L} > 0$), it follows the formulation of Holtslag and De Bruin (1988), while under unstable conditions ($\frac{z}{L} < 0$), it follows the Businger-Dyer expression described by Paulson (1970).

Stable conditions:

$$-\Psi_v\left(\frac{z}{L}\right) = \frac{z}{L} + b\left(\frac{z}{L} - \frac{c}{d}\right)\exp\left(-\frac{dz}{L}\right) + \frac{bc}{d}, \quad (4)$$

Unstable conditions:

$$x = \left(1 - 16\frac{z}{L}\right)^{0.25}, \quad (5)$$

$$\Psi_v\left(\frac{z}{L}\right) = \log\left(\frac{1+x}{2}\cdot\left(\frac{1+x}{2}\right)^2\right) - 2\cdot\arctan(x) + \frac{\pi}{2}, \quad (6)$$

where the constants are a = 1, b = 2/3, c = 5, and d = 0.35; z is the measurement height. To provide a reference for future studies without reliance on EC data, we parameterized the scalar roughness lengths (z_{0t} and z_{0q}) as functions of the momentum roughness length z_{0m}, which was derived from EC data. This parameterization follows the methods proposed by Andreas (1987) and Smeets and van den Broeke (2008), which have been shown to provide accurate estimates of scalar roughness lengths over glaciers. Hereafter, these two parameterization schemes are referred to as A87 and SvdB, respectively. In this study, we performed quality control on the 30-minute z_{0m} values following the procedures outlined by Conway and Cullen (2013) and Li et al. (2016). Detailed processing steps can be found in Radić et al. (2017). After filtering, 183 valid z_{0m} data points remained, with inferred z_{0m} values ranging from 1.14 × 10^{-5} m to 2.08 × 10^{-2} m. We finally selected the median of the quality-controlled z_{0m} values (1.2 × 10^{-4} m) as the representative roughness length for evaluating the performance of the different turbulent flux parameterization schemes. For the measured sensible and latent heat fluxes, the same general quality control procedures as those applied to z_{0m} were adopted. However, with respect to the neutrality and wind speed criteria, flux calculations were allowed to include data satisfying $|z/L| < 2$ and $u > 1$ m s^{-1}. After screening, 592 turbulent flux data points were retained for subsequent model calculations and evaluation."

(2) tested multiple scalar roughness-length parameterizations, please refer to **Section 4.3.2 - 4.3.3**, **lines 468–562** of the revised manuscript:

4.3.2 Overall model performance during the entire study period

4.3.3 Evaluation of daily variations

These additions allow us to better isolate the influence of roughness-length representation on both the magnitude and variability of modeled turbulent fluxes.

Third, regarding the uncertainty of turbulent-flux measurements themselves, we acknowledge that EC-derived fluxes are subject to non-negligible uncertainty. However, assigning a single quantitative uncertainty value to the measured turbulent fluxes is challenging. The EC processing chain—including coordinate rotation, frequency response correction, density corrections, quality control, and gap filtering—can substantially modify the final flux estimates, making it difficult to propagate uncertainty in a straightforward manner. For this reason, and in line with the practice adopted in most glacier EC

studies, we do not prescribe an explicit uncertainty range for the measured turbulent fluxes. Notably, previous studies relying on EC observations over glaciers (including those cited above) similarly refrain from assigning fixed uncertainty values to the measured fluxes. We note, however, that further investigation of measurement uncertainty is an important topic, and we intend to address this aspect in future work.

L336-341: "All five methods ... turbulent exchange." Since all methods seem to give roughly the same errors, I am inclined to think that it could also be related to the way roughness lengths are determined (since all methods depend on them). It could be worth looking into this further and check e.g. whether an alternative description (or calibration; see next point) of roughness lengths would strongly change the model results.

Answer: We agree with the reviewer's interpretation that the broadly similar errors across schemes may reflect a shared driving factor rather than differences intrinsic to the schemes themselves. To directly examine this hypothesis, we introduced an experiment in which the aerodynamic roughness length (z_{0m}) was derived from the observed eddy-covariance turbulent fluxes and then uniformly applied across all turbulent flux schemes. This approach allows us to largely isolate and reduce the influence of differing roughness-length representations when comparing model performance.

In the revised manuscript, we now explicitly report the EC-derived z_{0m} and discuss the differences in simulated turbulent fluxes among the schemes when a unified roughness length is applied. In addition, it should be noted that, due to changes in the filtering criteria used for evaluating the turbulent flux data, the screened observational flux dataset became discontinuous. Accordingly, we revised the presentation of the comparison between observations and model simulations, replacing the original line plots with scatter-based time series plots (Figure 5).

For details, see **lines 468–511** of the manuscript:

4.3.2 Overall model performance during the entire study period

We compared the temporal evolution of modeled LE and H at the Dunde Glacier against in-situ observations (Fig. 5). The results from the C_{kat} and C_{log} methods were obtained using their respective optimal parameters, whereas the other schemes used EC-derived z_{0m} as an input parameter and applied different parameterizations of scalar roughness lengths. Over the observation period, nearly all models reasonably captured the overall temporal evolution of turbulent fluxes, but most simulations exhibited

varying degrees of underestimation. For LE, during the three periods of abrupt flux changes—early and mid-July, and mid-August— nearly all schemes underestimated the magnitude of the observed transitions. Outside these transition periods, particularly during intervals of relatively low LE, the simulations generally showed better agreement with the observations. Although the C_{kat} and C_{log} methods substantially improved agreement with observations through parameter optimization, they still showed a tendency to underestimate LE. For H, schemes that use the bulk exchange coefficient as a function of roughness lengths and atmospheric stability clearly outperformed the C_{kat} method, especially during periods of elevated sensible heat flux. However, during periods when H exhibited a downward trend—most notably in mid-August—none of the methods successfully reproduced this behavior.

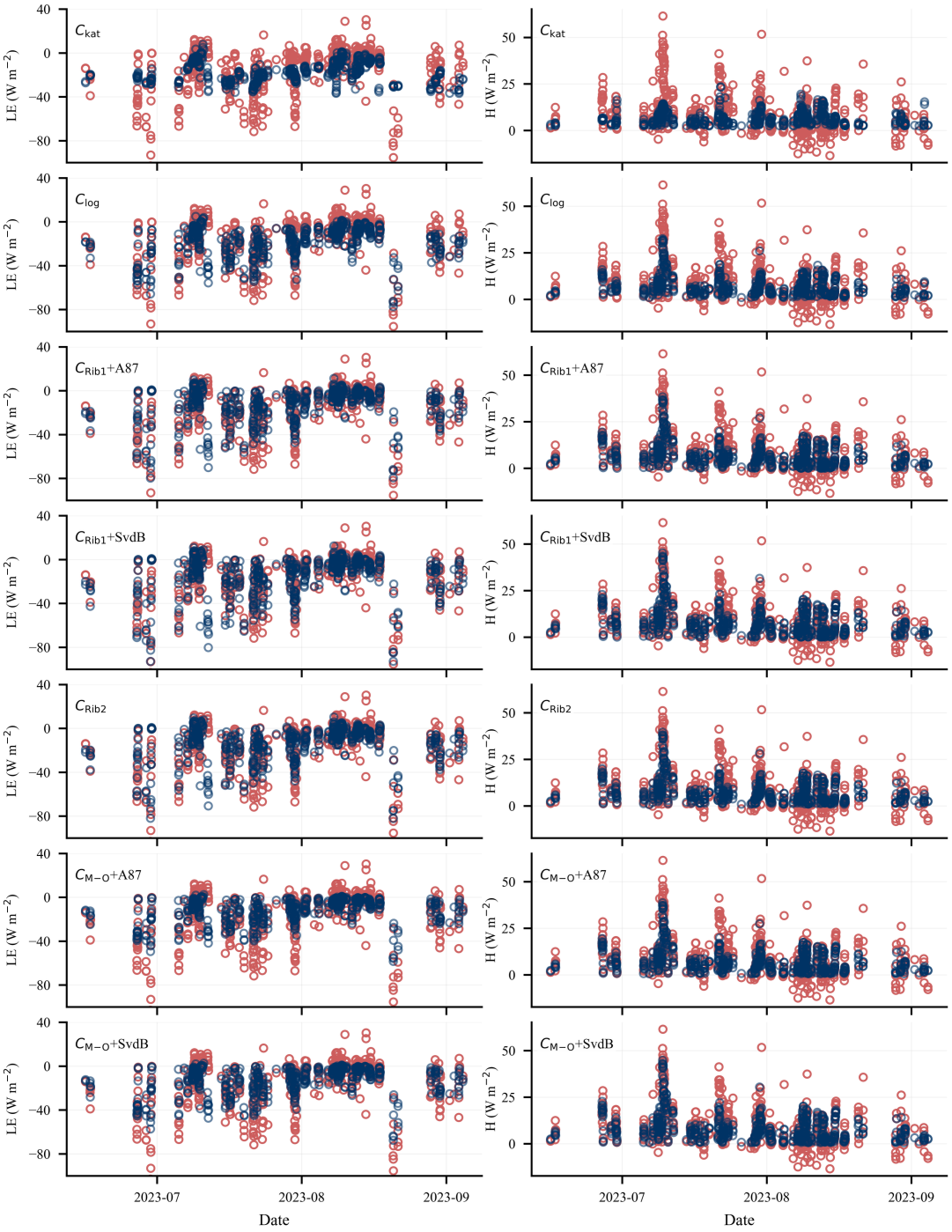


Figure 5: Time series of 30-min LE and H simulated by the five different schemes (C_{kat}, C_{log}, C_{Rib1}, C_{Rib2}, and C_{M-O}), compared with eddy covariance–derived values. The C_{Rib1}, C_{Rib2}, and C_{M-O} schemes were calculated by the observed aerodynamic roughness length (z_{0m} = 1.2 × 10^{-4} m), whereas C_{kat} and C_{log} were calculated using optimized parameters obtained by minimizing the differences between modeled and observed turbulent fluxes. Blue dots represent model simulations, and red dots denote observations.

Statistical comparisons between modeled and observed fluxes were conducted over the entire study period to quantitatively assess method performance (Table 4). For LE, the C_{kat}, C_{log}, and C_{Rib1} schemes using the SvdB scalar roughness length parameterization exhibited slight overall overestimation (with negative MBE indicating direction), whereas the remaining schemes—particularly C_{M-O}—tended to underestimate LE. Excluding C_{kat}, the other four schemes showed comparable performance, with RMSE values ranging from 11.2 to 13 W m^{-2}, indicating generally good agreement with observations. Although C_{kat} employed parameter optimization, its performance in terms of RMSE and MAD was inferior to that of the other schemes. Regarding the influence of scalar roughness length parameterizations, only minor differences were observed between the A87 and SvdB formulations for both the C_{Rib1} and C_{M-O} methods, suggesting that the choice of scalar roughness length parameterization had a limited impact on LE simulations over Dunde Glacier. For H, all schemes exhibited slight underestimation. However, overall model performance for H was better than that for LE. The observation-based schemes (C_{Rib1}, C_{Rib2}, and C_{M-O}), together with the parameter-optimized C_{log} method, performed slightly better than the katabatic-wind-based C_{kat} scheme. In contrast to the LE results, the choice of scalar roughness length parameterization exerted a noticeable influence on H simulations, with the SvdB formulation generally outperforming A87 for both C_{Rib1} and C_{M-O}.

Table 4: Comparative statistics of modeled LE and H over the entire study period using the observed aerodynamic roughness length, compared against eddy-covariance observations. RMSE = root mean square error; MAD = mean absolute deviation; MBE = mean bias error.

	LE (W m^{-2})			H (W m^{-2})		
Method	RMSE	MAD	MBE	RMSE	MAD	MBE
C_{kat}	15.8	12.5	−1.7	6.3	5.0	−1.7
C_{log}	11.2	7.6	−2.8	4.6	3.5	−1.6
C_{Rib1} + A87	12.7	7.4	1.2	4.1	2.9	−1.6
C_{Rib1} + SvdB	13.0	6.8	−1.1	3.5	2.4	−0.7
C_{Rib2}	12.4	7.0	0.3	3.9	2.9	−1.1
C_{M-O} + A87	12.4	8.4	4.0	4.0	2.9	−1.4

C_{M-O} + SvdB	11.5	7.6	2.5	3.6	2.5	−0.7

Figure 5: Given the weaker intra-day variability of in particular the latent heat flux, maybe the roughness length for the latent heat flux is off? I think it could have been an interesting strategy to try to find optimum values for (constant) roughness lengths by trying to minimize the differences between modelled and observed turbulent fluxes. Have you considered doing this?

Answer: We thank the reviewer for this valuable suggestion. In the revised manuscript, the aerodynamic roughness length (z_{0m}) used in the turbulent flux calculations was optimized within the range derived from EC observations to determine its optimal value.

In this study, we ultimately adopted the screened observationally derived aerodynamic roughness length ($z_{0}m$) for intercomparison among different parameterization schemes. Using a common observed $z_{0}m$ allowed us to evaluate the intrinsic differences among the schemes under identical roughness conditions, thereby ensuring a fair comparison. Moreover, the observationally constrained $z_{0}m$ demonstrated relatively good performance in reproducing turbulent flux magnitudes, and thus provides meaningful physical reference for continental glacier environments.

For schemes that do not explicitly prescribe roughness lengths (e.g., Ckat and Clog), we performed parameter optimization by minimizing the differences between modeled and observed turbulent fluxes, and identified optimal parameters. These optimized parameters are reported in the manuscript and may serve as useful references for future studies.

We agree that a systematic inversion of the optimal constant roughness length for each scheme represents a highly valuable extension of the present study. However, turbulent flux data obtained from the OPEC system over glaciers often contain spurious measurements. Parameters optimized using observations under specific conditions (e.g., near-neutral stability and relatively high wind speeds) may not be applicable to all periods, which can result in degraded performance when simulations are evaluated over the full-time span. On the other hand, directly optimizing parameters using turbulent flux observations from the entire period would not guarantee that the schemes are constrained under near-neutral conditions. Therefore, in this study, we ultimately chose to report the observed z_{0m} to provide a reference for future research. Nevertheless, we will consider implementing such a systematic inversion

approach in future work to further constrain roughness length parameterizations under varying atmospheric conditions. For details, please refer to **Section 3.1 (lines569−598)** in the newly revised draft:

3 Methods

3.1 The algorithm and quality control for roughness lengths

This study uses the roughness lengths of momentum derived from EC measurements to calculate turbulent fluxes. The EC data, with a 30-minute temporal resolution, provide three-dimensional wind speed components, which are then used to indirectly calculate friction velocity (u∗) and Obukhov length (L) as follows:

$$u_* = \left(\overline{u'w'}^2 + \overline{v'w'}^2\right), \quad (1)$$

$$L = -\frac{T_v u_*^3}{g\kappa \overline{w'T_v'}}, \quad (2)$$

where $u'$and v'represent the fluctuations in the horizontal wind components around their 30-minute mean values, w'denotes the fluctuation in the vertical wind component, T_v is the 30-min averaged virtual temperature (K), g is the acceleration due to gravity (9.8 m s^{-2}), and κ is the von Kármán constant (0.4). The roughness length for momentum (z_{0m}) is then retrieved using friction velocity and L, according to the following equation:

$$z_{0m} = \exp\left[-\kappa \frac{u}{u_*} - \Psi_v\left(\frac{z}{L}\right)\right] z, \quad (3)$$

Here, z_{0m} is the momentum roughness length, and u represents the 30-min averaged wind speed. The term Ψ_v is the stability correction function for momentum. Under stable conditions ($\frac{z}{L} > 0$), it follows the formulation of Holtslag and De Bruin (1988), while under unstable conditions ($\frac{z}{L} < 0$), it follows the Businger-Dyer expression described by Paulson (1970).

Stable conditions:

$$-\Psi_v\left(\frac{z}{L}\right) = \frac{z}{L} + b\left(\frac{z}{L} - \frac{c}{d}\right)\exp\left(-\frac{dz}{L}\right) + \frac{bc}{d}, \quad (4)$$

Unstable conditions:

$$x = \left(1 - 16\frac{z}{L}\right)^{0.25}, \quad (5)$$

$$\Psi_v\left(\frac{z}{L}\right) = \log\left(\frac{1+x}{2}\cdot\left(\frac{1+x}{2}\right)^2\right) - 2\cdot\arctan(x) + \frac{\pi}{2}, \quad (6)$$

where the constants are a = 1, b = 2/3, c = 5, and d = 0.35; z is the measurement height. To provide a reference for future studies without reliance on EC data, we parameterized the scalar roughness lengths

(z_{0t} and z_{0q}) as functions of the momentum roughness length z_{0m}, which was derived from EC data. This parameterization follows the methods proposed by Andreas (1987) and Smeets and van den Broeke (2008), which have been shown to provide accurate estimates of scalar roughness lengths over glaciers. Hereafter, these two parameterization schemes are referred to as A87 and SvdB, respectively. In this study, we performed quality control on the 30-minute z_{0m} values following the procedures outlined by Conway and Cullen (2013) and Li et al. (2016). Detailed processing steps can be found in Radić et al. (2017). After filtering, 183 valid z_{0m} data points remained, with inferred z_{0m} values ranging from 1.14 × 10^{-5} m to 2.08 × 10^{-2} m. We finally selected the median of the quality-controlled z_{0m} values (1.2 × 10^{-4} m) as the representative roughness length for evaluating the performance of the different turbulent flux parameterization schemes. For the measured sensible and latent heat fluxes, the same general quality control procedures as those applied to z_{0m} were adopted. However, with respect to the neutrality and wind speed criteria, flux calculations were allowed to include data satisfying $|z/L| < 2$ and $u > 1$ m s^{-1}. After screening, 592 turbulent flux data points were retained for subsequent model calculations and evaluation."

Mass balance modelling strategy:

Section 5.2: I can see why the authors want to run a mass balance model with the different turbulent flux schemes as it may give insight in sensitivity of the mass balance to the turbulent flux descriptions. However, in its current form I do not think the mass balance modelling experiment is of very much use. To me it would have made more sense to first calibrate the individual turbulent flux models against the observed turbulent flux data (by optimizing e.g. roughness lengths). In a next step these calibrated turbulent flux models could then be used in the mass balance model and the resulting mass balance from the runs with different turbulent flux schemes can be compared. There may be (like now) some biases between modeled and observed mass balance, but at least such an experiment will give robust insight in the sensitivity of the mass balance to using different turbulent flux models. If the above approach is what was already done by the authors, then it is just a matter of describing it more clearly.

Answer: We fully agree that the mass balance model provides more informative results when the turbulent flux schemes are first calibrated against observed eddy-covariance (EC) fluxes. Accordingly, we revised the workflow to ensure that each turbulent flux formulation is calibrated using observations

before being implemented in the surface energy balance (SEB) and mass balance (MB) models. For details, please refer to **Section 5.2 (lines610−649)** in the newly revised draft:

"**5.2 Performance improvement of the SEB model by optimizing turbulent flux methods**

To assess the importance of turbulent flux methods in simulating glacier mass balance, we included three representative turbulent flux methods (the C_{log}, C_{Rib2}, and C_{M-O} + SvdB methods) in a coupled energy balance–snow and firn model applied at the point scale. For each method, turbulent fluxes were computed using both the original parameter sets reported in the literature and the recalibrated parameters developed in this study, while all other model inputs, including incoming and outgoing longwave radiation, incoming and outgoing shortwave radiation, snow depth, surface albedo, and rainfall, were prescribed from in situ observations at AWS1 to ensure a controlled experimental setup.

Under this framework, the coupled model simulated glacier surface temperature, ice surface height, and cumulative mass balance (CMB). For each variable, model results obtained using the original and recalibrated turbulent flux parameter sets were evaluated against the corresponding observational data, allowing a direct assessment of the impact of turbulent flux parameterization and calibration on model performance (Fig. 7). In parallel, turbulent fluxes calculated with the original and recalibrated parameter sets were compared to quantify the effects of parameter recalibration on turbulent heat exchange itself. Model performance was assessed using the correlation coefficient (R), RMSE, MAD, and estimates of cumulative mass balance, thereby demonstrating the improvements achieved through turbulent flux optimization (Table 5).

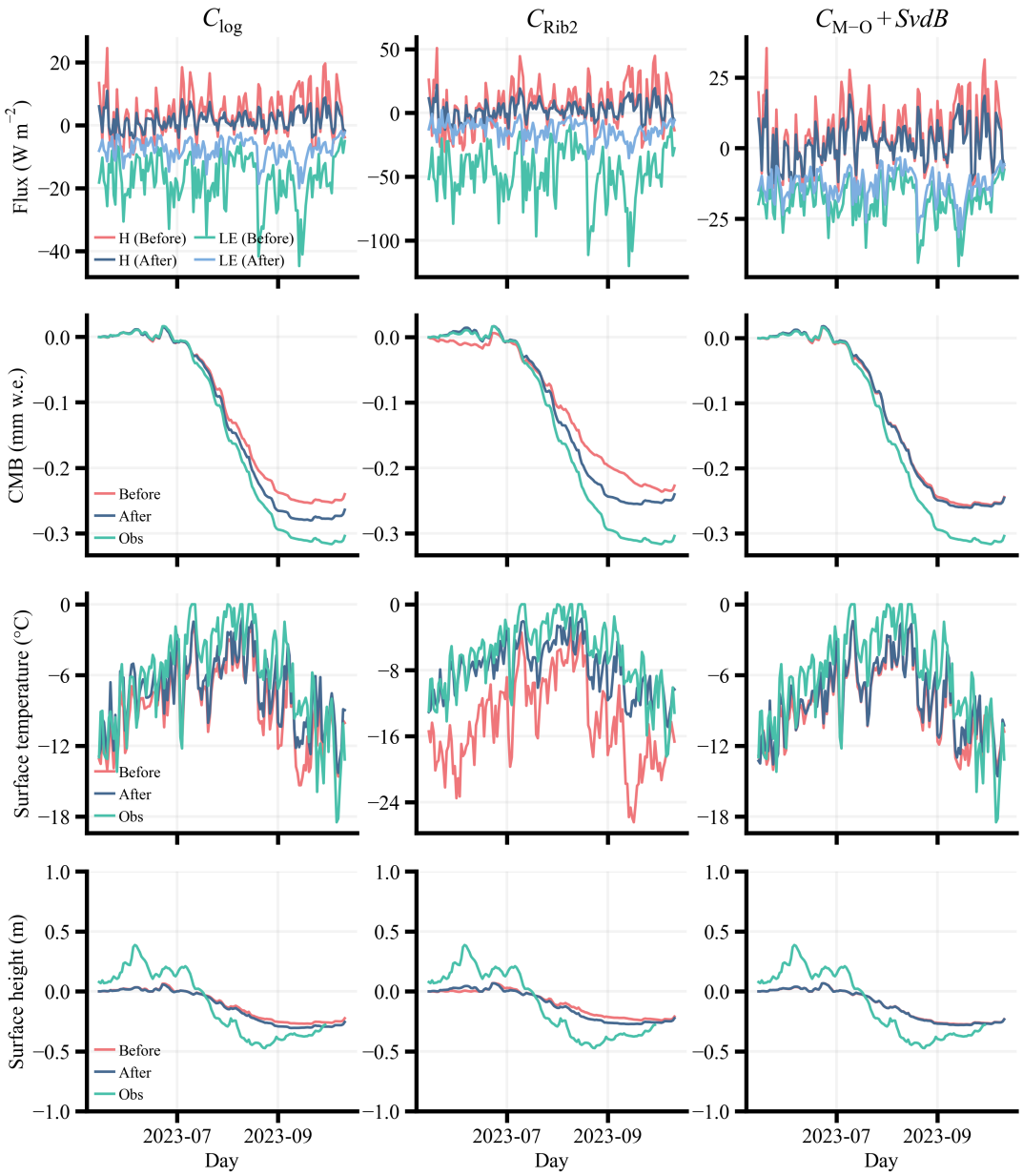


Figure 7: Time series of turbulent fluxes, cumulative mass balance (CMB), surface temperature, and surface height over the Dunde Glacier during the study period, based on original methods (before), recalibrated methods (after), and observations (obs).

Table 5: Comparative statistics between observations and surface energy balance model simulations of surface temperature (T_s), surface height, and cumulative mass balance at the Dunde Glacier, using both original (before) and recalibrated (after) parameters.

	Before	After

Method	T_s	Surface height	Mass balance	T_s	Surface height	Mass balance
C_{log}	R = 0.84 RMSE = 3.3°C	R = 0.93 RMSE = 0.16m MAD = 0.14m	−238.4mm	R = 0.85 RMSE = 2.5°C	R = 0.93 RMSE = 0.15m MAD = 0.14m	−260.3mm
C_{Rib2}	R = 0.78 RMSE = 8.8°C	R = 0.91 RMSE = 0.18m MAD = 0.16m	−225.8mm	R = 0.87 RMSE = 2.8°C	R = 0.94 RMSE = 0.15m MAD = 0.14m	−238.6mm
C_{M-O}	R = 0.86 RMSE = 3.2°C	R = 0.94 RMSE = 0.16m MAD = 0.14m	−242.4mm	R = 0.88 RMSE = 2.6°C	R = 0.94 RMSE = 0.14m MAD = 0.14m	−249.5mm

We first ran the SEB models under uncalibrated conditions with three turbulent flux methods (the C_{log}, C_{Rib2}, and C_{M-O} + SvdB methods). These gave RMSE values for T_s of 3.3, 8.8, and 3.2 °C, respectively. The R values between modeled and measured T_s exceeded 0.78 for all three schemes. The RMSE values for surface height change ranged from 0.16 to 0.18 m, while the CMB estimates were −238.4, −225.8, and −242.4 mm, respectively, all deviating markedly from the observed value of −298.6 mm. Following parameter calibration, all three schemes exhibited improvements in model performance. The C_{Rib2} method showed the greatest improvement, giving a CMB value of −238.6 mm, equating to an error reduction of 5.6% relative to the uncalibrated result. In parallel, the RMSE values for T_s and surface height decreased by 68.2% and 3.2%, respectively. The models based on the C_{M-O} and C_{Rib2} methods achieved relatively high overall accuracy after optimization, with R values of 0.87 and 0.88 for T_s, respectively, and 0.94 for surface height in both cases. Although a residual bias remained in the CMB simulation, it was substantially mitigated compared with the original configuration. Overall, these findings highlight that turbulent flux parameter optimization is an effective approach to improve glacier SEB model accuracy."

Selecting a best turbulent flux model:

L327: "indicating the best accuracy among the tested methods". I think it should rather be emphasized that the performance of the five methods is very similar. Given uncertainty in the observations, and these

small differences, it is not possible to argue (with statistical significance) that one method is better than the others based on these results.

Answer: We agree and have revised the wording to avoid over-interpretation. The revised manuscript emphasizes that overall performance differences among the five schemes are relatively small and generally comparable within typical error ranges. Accordingly, we removed or softened statements implying the existence of a single "best" method based solely on marginal differences in performance metrics. Instead, we focus on differences among models under a common roughness-length setting and on regime-dependent robustness (e.g., behavior during extreme events), while explicitly acknowledging that these differences may not be statistically significant. For details, please refer to **lines 491−511、lines 521−540、lines 546−562** in the newly revised draft:

Lines 491−511: Statistical comparisons between modeled and observed fluxes were conducted over the entire study period to quantitatively assess method performance (Table 4). For LE, the C_{kat}, C_{log}, and C_{Rib1} schemes using the SvdB scalar roughness length parameterization exhibited slight overall overestimation (with negative MBE indicating direction), whereas the remaining schemes—particularly C_{M-O}—tended to underestimate LE. Excluding C_{kat}, the other four schemes showed comparable performance, with RMSE values ranging from 11.2 to 13 W m^{-2}, indicating generally good agreement with observations. Although C_{kat} employed parameter optimization, its performance in terms of RMSE and MAD was inferior to that of the other schemes. Regarding the influence of scalar roughness length parameterizations, only minor differences were observed between the A87 and SvdB formulations for both the C_{Rib1} and C_{M-O} methods, suggesting that the choice of scalar roughness length parameterization had a limited impact on LE simulations over Dunde Glacier. For H, all schemes exhibited slight underestimation. However, overall model performance for H was better than that for LE. The observation-based schemes (C_{Rib1}, C_{Rib2}, and C_{M-O}), together with the parameter-optimized C_{log} method, performed slightly better than the katabatic-wind-based C_{kat} scheme. In contrast to the LE results, the choice of scalar roughness length parameterization exerted a noticeable influence on H simulations, with the SvdB formulation generally outperforming A87 for both C_{Rib1} and C_{M-O}.

Lines 521−540: At the daily scale, the simulated turbulent fluxes showed good agreement with the in-situ observations, with both LE and H exhibiting diurnal variations similar to the measured values (Fig. 6). For LE, the observed and simulated fluxes for all methods reached their maximum around 15:30;

however, the modeled values were generally lower than the observations (Fig. 6a). Meanwhile, differences among the parameterizations became apparent. Specifically, the two Richardson number-based stability correction schemes (C_{Rib1} and C_{Rib2}) reproduced LE more accurately, showing smaller deviations from the observations, particularly during periods of high LE, when they better captured the amplitude of flux variability. In contrast, the other methods—especially C_{M-O}—tended to underestimate LE during these high-flux periods. During relatively stable periods (after approximately 20:00), C_{kat} and C_{log} exhibited overestimation, with comparatively larger deviations. Regarding the scalar roughness length parameterizations, the two schemes performed similarly during low LE periods; however, at higher LE values, the SvdB parameterization outperformed A87. This pattern was evident in both the C_{Rib1} and C_{M-O} frameworks. For H, both observations and simulations showed a gradual diurnal increase, with minimum values occurring around 12:00 and maximum values around 23:00 (Fig. 6b). Differences among the parameterizations were relatively small. Compared with LE, all methods demonstrated generally better agreement for H, although a systematic underestimation persisted. The underestimation was generally more evident during periods when H values were relatively high (e.g., around 18:30 and 21:30). Concerning the scalar roughness length parameterizations, the SvdB scheme consistently performed slightly better than A87 across the evaluated time bins, although the differences between the two schemes remained modest.

Lines 546–562: In summary, among the five evaluated methods, both the C_{kat} and C_{log} schemes exhibited a tendency to underestimate the observed turbulent fluxes to some extent; however, the optimized parameters adopted in these schemes provide useful references for studies of continental glaciers. For the EC-based approaches derived from observations (i.e., the C_{Rib1}, C_{Rib2}, and C_{M-O} methods), the overall performance in simulating both LE and H was broadly comparable throughout the study period, with generally higher simulation accuracy for H than for LE. Nevertheless, differences among the schemes became more evident when examining the diurnal variations. In particular, the C_{Rib1} and C_{Rib2} methods showed closer agreement with observations in reproducing the diurnal evolution of LE, whereas the differences among schemes in simulating H were relatively small. In addition, considering both the overall period and the diurnal cycle, the SvdB scalar roughness length parameterization consistently outperformed the A87 scheme. Overall, although the differences among the five schemes were not

substantial, the methods employing Richardson number-based stability corrections combined with the SvdB scalar roughness parameterization demonstrated comparatively balanced performance in simulating both LE and H, and were able to more accurately reproduce the diurnal evolution of turbulent fluxes. Therefore, among the five evaluated schemes, this class of approaches can be regarded as a relatively favorable option for simulating turbulent fluxes over the Dunde Glacier. More broadly, it may also be applicable to turbulent flux simulations over continental glaciers on the TP."

L425: "the best performance". It is only marginally better than the other models. So statistically speaking the performance is not significantly different across the models.

Answer: We agree and have revised the manuscript to avoid stating or implying statistical superiority where it is not demonstrated. References to "best performance" have been replaced by more cautious language (e.g., "comparable performance," "marginal differences,"), and the discussion now highlights that multiple schemes achieve similar accuracy depending on the metric considered. For details, please refer to **lines 744−752** in the newly revised draft: "In addition, we systematically evaluated the performance of commonly used turbulent flux parameterizations. Among the five schemes assessed, considering both the overall period and the diurnal cycle, when the models were applied using the aerodynamic roughness length derived from observations, the combined approach based on Richardson-number stability correction and the SvdB scalar roughness length parameterization exhibited comparatively higher accuracy in simulating both LE and H, and more accurately reproduced the diurnal variability of turbulent fluxes. The RMSE for LE ranged between 12.4 and 13.0 W m^{-2}, while the RMSE for H ranged between 3.5 and 4.1 W m^{-2}. Moreover, statistical comparisons indicate that within both the C_{Rib1} and C_{M-O} frameworks, the SvdB scalar roughness length parameterization consistently outperformed the A87 scheme."

L549-552: "Among the five schemes ... variability.". Again, this is a rather bold statement based on the presented results. There are clear benefits for using the simpler (easier to implement and quicker to run) models too, especially if their performance is similar to the MO model.

Answer: We agree with this comment and have revised the Conclusions to provide a more balanced perspective. The revised text explicitly states that when all schemes are implemented with optimized

parameters, the differences in overall model performance are generally small and comparable. However, the relative performance of the models may vary across different temporal scales or target variables. Accordingly, we have reframed the results and provided a reference selection strategy. This clarification does not imply that the other schemes perform poorly, but rather highlights their respective strengths and applicable contexts. For details, please refer to **lines 744–752** in the newly revised draft: "In addition, we systematically evaluated the performance of commonly used turbulent flux parameterizations. Among the five schemes assessed, considering both the overall period and the diurnal cycle, when the models were applied using the aerodynamic roughness length derived from observations, the combined approach based on Richardson-number stability correction and the SvdB scalar roughness length parameterization exhibited comparatively higher accuracy in simulating both LE and H, and more accurately reproduced the diurnal variability of turbulent fluxes. The RMSE for LE ranged between 12.4 and 13.0 W m^{-2}, while the RMSE for H ranged between 3.5 and 4.1 W m^{-2}. Moreover, statistical comparisons indicate that within both the C_{Rib1} and C_{M-O} frameworks, the SvdB scalar roughness length parameterization consistently outperformed the A87 scheme."

Minor comments / technical corrections:

L30: "a negative total ... climate condition". Replace with "the commonly observed negative total turbulent heat flux".

Answer: We adopted the suggested wording to improve scientific precision and readability. For details, please refer to **line 32** in the newly revised draft "We also found that the Dunde Glacier experienced a sharp increase in H and reversal in LE during a humid heatwave event, shifting from the commonly observed negative total turbulent heat flux under the mean climate condition to positive values during the extreme event." which better reflects the general phenomenon without introducing unnecessary qualifiers."

L32: "extreme weather and climate event". Replace with "extreme event".

Answer: We revised "extreme weather and climate event" to "extreme weather event" or "extreme event" for concision and to avoid mixing timescale concepts. For example, in the revised manuscript we now state:

L32: We also found that the Dunde Glacier experienced a sharp increase in H and reversal in LE during a humid heatwave event, shifting from the commonly observed negative total turbulent heat flux under the mean climate condition to positive values during the extreme event.

L50: Under ongoing global warming, the frequency and intensity of extreme weather events are projected to increase, which may lead to severe future glacier mass loss due to multiple contributing factors, such as enhanced turbulent fluxes and increased incoming longwave radiation

L653: However, the ability of SEB models to simulate turbulent fluxes under extreme events remains uncertain and warrants further investigation. Between 8 and 11 July, 2023, the Dunde Glacier experienced a typical humid heatwave (Fig. 8a and b).

L47: "frequencies ... events". Replace with "frequency and intensity of extreme weather events".

Answer: We implemented the reviewer's suggestion and revised the wording to "frequency and intensity of extreme weather events," which is both grammatically correct and more precise. Please refer to **line 50** in the newly revised draft: "Under ongoing global warming, the frequency and intensity of extreme weather events are projected to increase, which may lead to severe future glacier mass loss due to multiple contributing factors, such as enhanced turbulent fluxes and increased incoming longwave radiation."

L48: Not only enhanced turbulent fluxes are responsible for increased glacier mass loss but also e.g. increased incoming LW radiation. Please reformulate.

Answer: We agree and revised the statement to acknowledge that enhanced turbulent fluxes are not the only driver of increased glacier mass loss. The revised text explicitly notes the role of radiative forcing (e.g., increased incoming longwave radiation). The revised sentence **(line 49)** now reads: "Under ongoing global warming, the frequency and intensity of extreme weather events are projected to increase, which may lead to severe future glacier mass loss due to multiple contributing factors, such as enhanced turbulent fluxes and increased incoming longwave radiation (Brun et al., 2018; Duan et al., 2012; Hugonnet et al., 2021; Yao et al., 2012)."

L63: "climatic mechanisms". Please note that climate refers to much longer timescales than considered here. Maybe use "weather patterns" or similar instead.

Answer: We revised the wording to avoid using “climatic” for processes discussed over a seasonal window. The revised manuscript uses “weather patterns” to match the timescale analyzed. The revised sentence **(line 66)** now reads: “This data gap has resulted in an inadequate understanding of the relative contributions of individual energy balance components to the overall surface energy budget of TP glaciers, which, in turn, is crucial for understanding the weather patterns controlling glacier variations across the TP.”

L64: "has become" --> "is".

Answer: We implemented the suggested grammatical correction and revised “has become” to “is” to improve consistency with the surrounding text. The revised sentence **(line 70)** now reads: “Owing to the scarcity of observational data, numerical modeling is the primary method…”

L65: "are" --> "have been".

Answer: We revised “are” to “have been” where appropriate to ensure consistent tense usage across the paragraph. The revised sentence **(line 70)** now reads: “...by which glacier turbulent fluxes have been estimated on the TP (Mölg et al., 2012; Yang et al., 2011; Zhu et al., 2018).”

L68: "intensive" --> "expensive".

Answer: We have revised the relevant statements to ensure the content is accurate and truthful, and correctly describes the characteristics of MOST. The revised sentence **(line 72)** now reads: “Current turbulent flux modeling approaches predominantly rely on Monin–Obukhov similarity theory (MOST) (Monin and Obukhov, 1954), which has been widely applied for parameterizing near-surface turbulent fluxes.”

L76-80: "Taking the ... 2006-2011". This comparison requires a bit more info to be useful. E.g. what methods were used (modelling or observations)? How large were uncertainties in the two studies? Could the different periods considered explain the differences?

Answer: We agree that such a comparison is only meaningful when additional contextual information is provided. We therefore expanded the text to briefly describe the methodologies adopted in the referenced studies. Specifically, Zhang et al. (2016) applied a highly simplified Monin–Obukhov–based bulk approach, in which constant exchange coefficients driven by near-surface wind speed were used (the C_{log} scheme). In contrast, Huintjes et al. (2015) employed a bulk method in which atmospheric stability is represented by the bulk Richardson number (Rib), corresponding to the C_{Rib2} scheme. In both studies, the relevant parameters were prescribed based on previous investigations conducted on glaciers with similar surface characteristics. In addition, because both studies reported multi-year mean turbulent fluxes, the influence of differences in study periods is reduced to some extent. Moreover, as both studies relied on model-based estimates and did not provide a dedicated uncertainty analysis, we do not further discuss uncertainty-related aspects in this context.

The revised sentence **(L87)** now reads: "Previous studies on the Zhadang Glacier clearly demonstrate this model-dependent variability, showing that the estimated multi-year mean turbulent heat fluxes for both winter and summer differ substantially when different modeling approaches are applied. Zhang et al. (2016) employed a highly simplified Monin–Obukhov–based bulk approach with constant exchange coefficients driven by near-surface wind speed and derived multi-year mean turbulent heat fluxes of 13.4 W m^{-2} in winter and 5.7 W m^{-2} in summer for the period 2011–2014. In contrast, Huintjes et al. (2015) adopted a bulk method in which atmospheric stability is represented by the bulk Richardson number, yielding multi-year mean turbulent heat fluxes of 8.0 W m^{-2} in winter and −28 W m^{-2} in summer during 2001–2011. Such differences may partly arise from variations in the study periods; however, a more important factor is the limited availability of observational data for calibrating turbulent-flux parameterizations, as well as differences among the turbulent-flux models themselves."

L88: "... on the TP.". Add "has previously been done".

Answer: We implemented the suggested revision and added clarifications indicating that related work has previously been conducted, including the evaluation of roughness-length parameterizations by Guo et al. (2011) and the comparison between EC measurements and alternative sublimation-monitoring methods by Liu et al. (2024). This clarification strengthens the study context and helps to more clearly delineate the remaining gaps addressed by the present work.

The revised sentence **(L98)** now reads: “Consequently, evaluations of the accuracy of different turbulent-flux parameterizations over the Tibetan Plateau remain limited, are largely confined to short time scales, and are constrained by the scarcity of EC observations. Moreover, comprehensive assessments of turbulent-flux schemes specifically for continental glaciers are still lacking. Guo et al. (2011) evaluated three turbulent flux parameterizations for the Parlung No.4 Glacier (maritime glacier) and found that the scheme proposed by Yang et al. (2002) produced lower errors in turbulent flux estimates during both individual melt phases and the entire ablation season. Compared with the other two schemes evaluated (Andreas, 1987; Smeets and van den Broeke, 2008), the mean absolute deviation (MAD) was reduced by approximately 12–29%. Liu et al. (2024) employed eddy-covariance measurements to validate a one-camera time-lapse structure-from-motion (O-T-SfM) photogrammetry approach for monitoring snow sublimation on the August-one ice cap. The results indicate that, under snow-free conditions, glacier surface sublimation rates estimated using the O-T-SfM method were in good agreement with EC measurements. Both methods successfully captured daily snow sublimation during the winter season.”

L90-92: "These uncertainties hinder ... between glaciers and climate.". This becomes a bit repetitive now. I suggest to remove it here.

Answer: We agree and removed/reduced the repetitive sentences to improve concision while retaining the essential message. The revised sentence **(L110)** now reads: “However, to date, no comprehensive analysis has been conducted specifically for continental glaciers. Previous assessments focusing on maritime glaciers on the TP cannot be directly applied to the widely distributed continental glaciers due to fundamental differences in glacier characteristics. Continental glaciers therefore require observation-based evidence that is more directly applicable to their climatic and surface conditions (Zhu et al., 2023).”

L93: "systematically analyze" --> "systematic analysis".

Answer: We revised the phrasing to “systematic analysis” as suggested to improve clarity and stylistic consistency. The revised sentence **(L116)** now reads: “Here we provide the first systematic analysis of meteorological and glacier mass balance observations and direct eddy-covariance-based turbulent flux measurements at the Dunde Glacier…”

L98: "the model robustness" --> "performance of the turbulent flux models".

Answer: We replaced "model robustness" with "performance of the turbulent flux models" to more accurately reflect what is being evaluated. The revised sentence **(L121)** now reads: "…to test the performance of the turbulent flux models."

Section 2: I suppose surface temperature is estimated from the observed outgoing longwave radiation. This should be described somewhere. Was an emissivity of 1 assumed?

Answer: We agree that surface temperature estimation must be explicitly documented. In the revised Methods, we clarify that T_s is derived from observed outgoing longwave radiation and we state the assumed surface emissivity used in the inversion. This improves transparency and enables reproducibility.

The revised sentence **(L153)** now reads: "The glacier surface temperature used in this study was derived from the observed outgoing longwave radiation, assuming a constant surface emissivity of 0.98. Details of the sensors used to measure each variable are provided in Table 1."

L119: "using" --> "using a".

Answer: We implemented the suggested grammatical correction by revising "using" to "using a." The revised sentence **(L150)** now reads: "…were primarily measured at the study site using a Campbell-ClimaVUE50 automatic weather station (AWS1) located at 5317 m a.s.l."

L190: "pressure Ps". Please note that air pressure was already defined as "P" in L165.

Answer: We standardized the pressure notation and removed duplicate definitions to avoid confusion (using a single symbol consistently throughout text and equations).

L201: Please note that the **ϵ** constant is the same as the "0.622" that is used in several equations (e.g. Eqs (1) and (3)).

Answer: We clarified that ϵ corresponds to the 0.622 constant (molecular weight ratio) and ensured consistent notation across equations and variable definitions. The revised sentence **(L331)** now reads: "…and ϵ is defined as the ratio of the molecular weights of water to dry air, with a value of 0.622."

Sect. 4.1: Please make sure that italic font is used for variables.

Answer: We revised the formatting to ensure variables are presented in italic font in accordance with standard mathematical typesetting.

L224-242: This is a lot of text about the temperature time-series, which are also shown in a figure and do not contain too many surprising features. I suggest to shorten this.

Answer: We agree and have condensed the descriptive text, retaining only the key features necessary for interpretation, including the range, mean values, and the months with the highest and lowest daily mean meteorological conditions, as well as the main characteristics of the diurnal variations. All other descriptive content has been removed to ensure conciseness and avoid redundancy. Please refer to **line 366** in the newly revised draft:"Figure 2 shows the temporal evolution of near-surface meteorological conditions during the observation period from 14 May to 12 October 2023. The daily mean T_a at the Dunde Glacier ranged from −14.1 to 3.3 °C, with an average of −4.0 °C (Fig. 2a). The warmest months were July and August, averaging −1 °C. October was the coldest month, with a mean daily T_a of −9.5 °C. The mean daily T_s ranged from −17.2 °C to 0 °C, with a mean of −5.4 °C (Fig. 2d) and exhibited seasonal variability. Over the study period, the mean surface temperature deficit ($T_a - T_s$) was 1.4 °C, reaching its highest values in July and September (1.9 °C) and its lowest value in June (0.7 °C). To isolate the diurnal variability from the dominant seasonal signal, Fig. 2c and f presents detrended daily cycles of T_a and T_s. After detrending, both variables display clear and consistent single-peak and single-trough diurnal patterns; The earlier T_s peak is interpreted to reflect the rapid radiative response of the glacier surface to incoming solar radiation; T_a responds more slowly owing to the thermal inertia of air and gradual development of convective mixing. At night, however, both the glacier surface and air cool synchronously under stable stratification and weak turbulence, resulting in the coincident early morning T_s and T_a minima. Generally, T_a was higher than T_s for most of a given day, indicative of a persistent near-surface temperature inversion layer over the glacier surface (Fig. 2t and u)."

L243-268: Also this can be shortened. Essentially, everything mentioned here is also visible in Figure 2 so just a short summary would work.

Answer: We shortened this subsection substantially and replaced lengthy narration with a concise summary that complements (rather than repeats) the figure.

Please refer to **line 381** in the newly revised draft: "Mean daily relative humidity at the Dunde Glacier was 56.3%, with notable seasonal variability (Fig. 2g). The highest seasonal average occurred in May-Jun (62.6%), whereas Sep-Oct was the driest season (50.7%). On a daily scale, relative humidity characterized by a distinct midday trough and nighttime peak (Fig. 2i).

Daily mean wind speeds ranged from 1.7 to 10.5 m s^{-1}, averaging 4.1 m s^{-1}, with marked seasonal variability (Fig. 2j). The strongest winds occurred during May-Jun (average of 5.51 m s^{-1}), while Jun-Aug recorded the weakest winds (average of 3.6 m s^{-1}). On a daily scale, wind speed exhibited s a pronounced diurnal cycle, with the mean daily cycle showing a peak-to-trough difference of approximately 1.5 m s^{-1} (Fig. 2l). Wind direction was presented using circular statistics. Nighttime wind speeds exhibited limited variability, consistent with stable atmospheric stratification. Prevailing wind directions were westerly and northwesterly (Fig. 2m and o), highlighting the dominant influence of the mid-latitude westerlies on regional meteorological conditions.

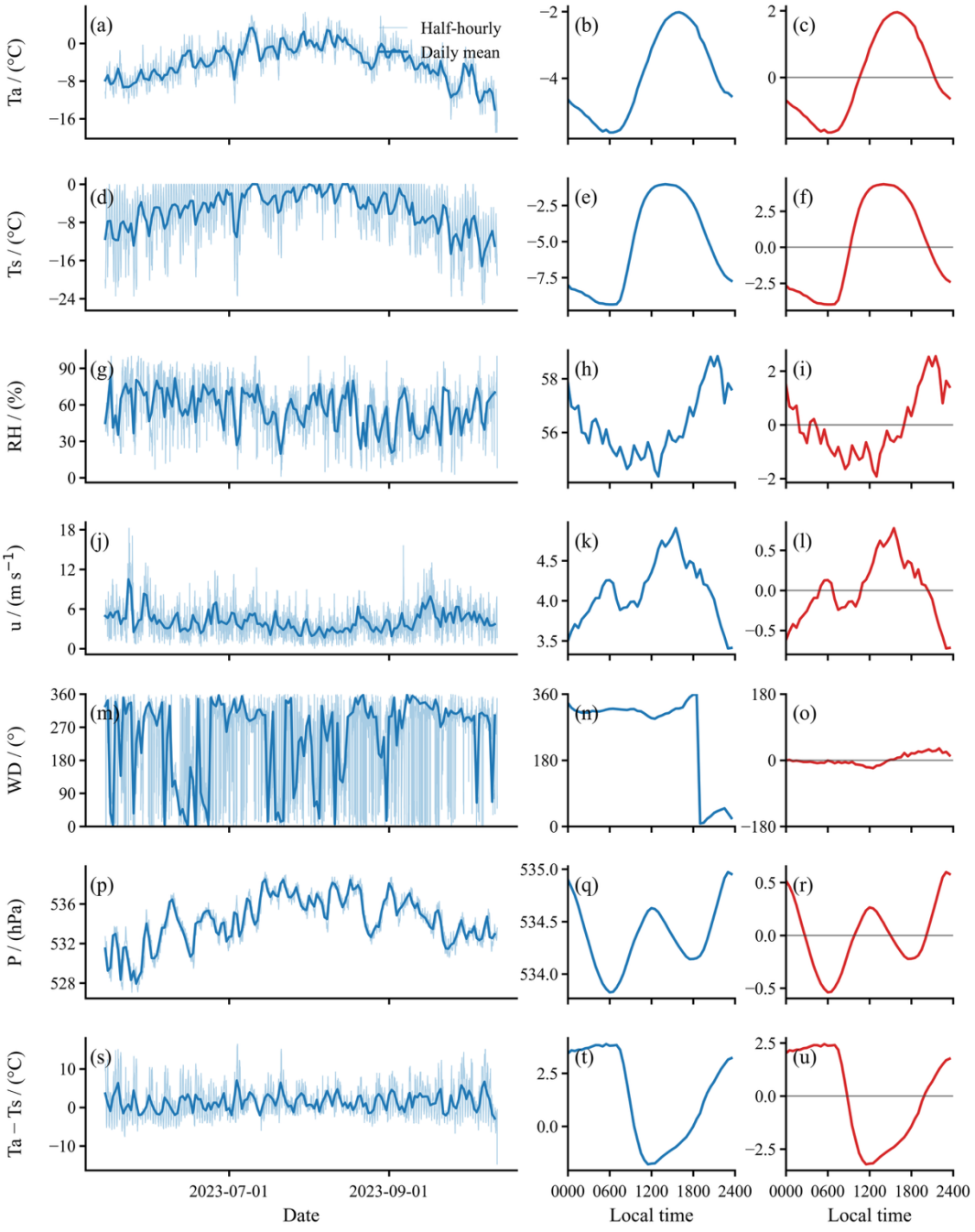
Daily mean atmospheric pressure at the Dunde Glacier during our study period ranged from 527.9 to 538.4 hPa, with an average of 534.4 hPa (Fig. 2p). The highest seasonal mean atmospheric pressure occurred in Jun-Aug (535.2 hPa) and the lowest in May-Jun (530.6 hPa). On a daily scale, atmospheric pressure exhibited a distinct double-peak, double-trough daily pattern (Fig. 2r). Overall, atmospheric pressure variations were relatively small over the study period and therefore likely played only a minor role in modulating turbulent fluxes.

Figure 2: Units on the y-axes should be in brackets. Additionally, please use the same variable names as in the main text.

Answer: We revised Figure 2 by adding unit annotations in parentheses on the vertical axes. In addition, we updated the variable labels to ensure consistency with those used in the main text.

Figure 2: You could consider adding a panel with Ta-Ts (the temperature deficit), which is an important parameter for the turbulent fluxes and their direction. Furthermore, albedo could be shown. It is discussed in connection to the extreme event in July 2023.

Answer: Thank you for this helpful suggestion. We agree that the temperature deficit ($T_a - T_s$) and surface albedo provide important physical context for interpreting the direction and magnitude of turbulent fluxes, particularly during extreme events. In the revised manuscript, we added an explicit panel showing the temperature deficit in Figure 2 of Section 4.1 and expanded the corresponding discussion. Because albedo is discussed in Section 5.3, we added the albedo panel to Figure 8 in the revised Section 5.2 to support the interpretation of the July 2023 event.



L286 (for example): With an observation period starting in May and finishing in October it maybe is better to not refer to "spring" and "autumn" but rather the months in question, e.g. "May-Jun" or "Sep-Oct".

Answer: We implemented the suggestion and replaced seasonal terms (e.g., “spring,” “autumn”) with explicit month ranges (e.g., May–Jun, Sep–Oct) consistent with the study period.

L287-288: "with June having the lowest monthly average". I suppose that coincides with the period with the lowest temperature deficit at the surface(?). If so, that could be mentioned here.

Answer: Thank you for this helpful suggestion. We expanded the discussion by explicitly linking sensible heat flux to the surface–air temperature deficit. We show that the monthly mean H reaches its minimum in June, which coincides with the period of the smallest surface–air temperature difference. This consistency supports the interpretation that reduced temperature gradients are associated with weakened turbulent exchange and helps explain the June minimum, as now clarified in the revised manuscript.

The revised sentence **(L417)** now reads: “The lowest values were recorded in Jun-Sep, which averaged 5.4 W m^{-2}, and June exhibited the lowest monthly average (3.1 W m^{-2}), coinciding with the period when the surface temperature deficit reached its minimum. During June, both T_a and glacier T_s increased; however, the glacier surface warmed more rapidly than the overlying air, reducing the surface temperature deficit to its lowest level. As a result, the monthly mean H reached its minimum in June.”

L291: "through turbulent heat exchange". So sublimation / evaporation are large. Related question, do you find penitentes on the glacier surface? If so, it could be worth highlighting in the introduction as it is a visible confirmation of the significance of sublimation and evaporation.

Answer: Thank you for this interesting question. We did not conduct a systematic field survey specifically aimed at documenting penitentes during this campaign, and therefore we cannot draw a definitive conclusion. Based on our field visits to the accumulation area of the Dunde Glacier, we did not note clear evidence of penitentes during the summer period (June–August 2023) or the winter period (December–January). We emphasize that this does not exclude their possible occurrence outside our observation periods or at other locations. We therefore consider penitentes formation an important

direction for future targeted surface observations, and plan to explicitly investigate this phenomenon in future field campaigns.

Figure 3: Please add the type of flux in the y-axis labels.

Answer: We revised Figure 3 axis labels to explicitly include the flux type (sensible vs. latent), improving interpretability and consistency with the text.

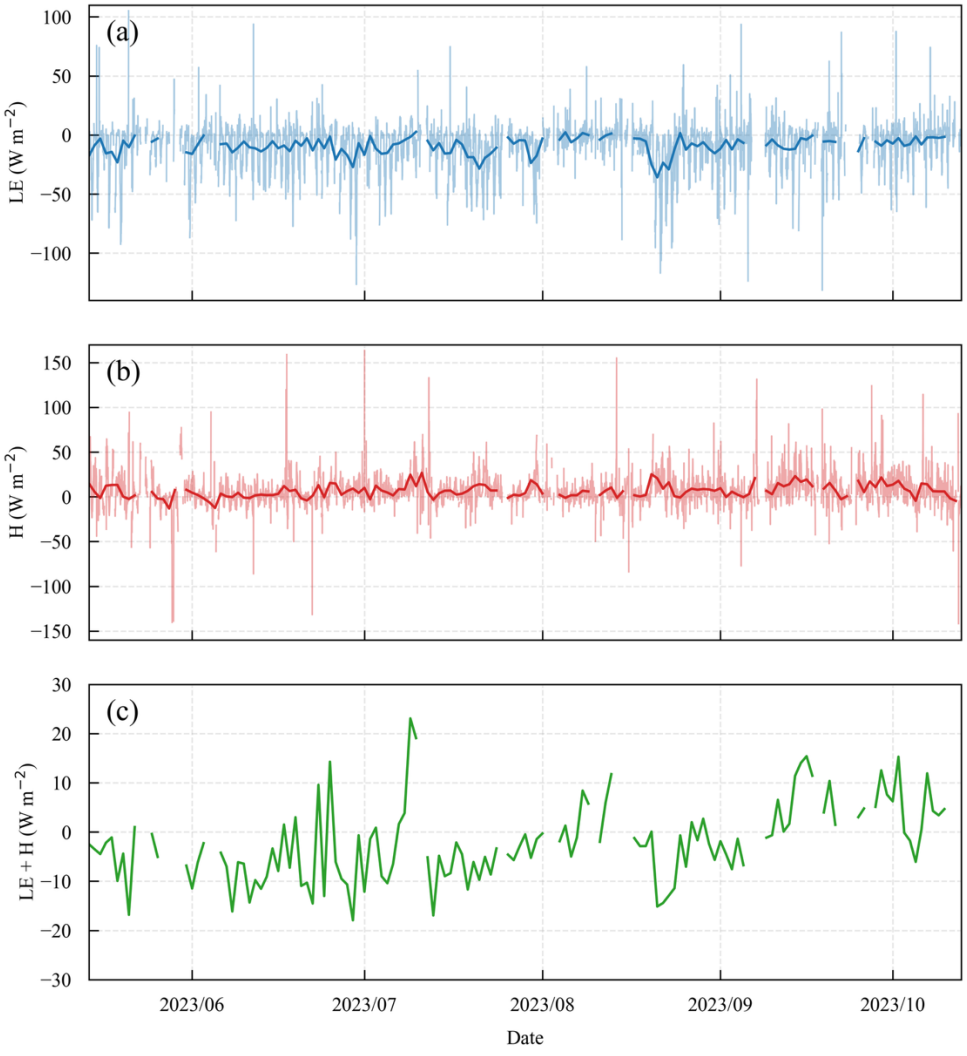


Figure 3: Since there are data gaps I suppose some filtering of the data was done. Unless I missed it, I do not think this is currently described. If so, please add this.

Answer: Thank you for this comment. The filtering and processing of EC data are fundamental to the entire analysis and therefore need to be explicitly documented. In the revised Methods section, we now clearly describe the full set of filtering and processing procedures applied to the EC data, including

but not limited to the removal of abnormal values, turbulent flux corrections, and other quality-control steps **(see Sect. 2.3, L170)**. The revised text is provided below:

"Turbulence data were processed using the EddyPro software. The processing workflow was organized based on a flux averaging interval of 30 min. Prior to flux calculation, raw CSAT3 data were screened using the instrument-provided quality flags to identify and remove invalid or unreliable measurements. This quality-flag-based filtering was applied at the high-frequency level before covariance calculation, and no interpolation was performed for the removed samples. Turbulent fluctuations were subsequently defined using the detrending options implemented in EddyPro. Time lags between wind components and scalar quantities were compensated using covariance maximization based on circular correlation techniques, which is recommended for open-path EC configurations (Moncrieff et al., 1997, 2004). Coordinate rotation was applied using the double-rotation method to correct for sensor tilt and streamline distortion (Kaimal and Finnigan, 1994). Flux correction was then applied to the calculated fluxes. Spectral corrections were first implemented to ensure that density corrections were based on environmentally representative fluxes. Density fluctuation effects were corrected using the Webb–Pearman–Leuning (WPL) formulation (Webb et al., 1980). The effects of humidity on sonic temperature and temperature-related covariances were corrected following the approach of Schotanus et al. (1983). Spectral attenuation resulting from both low-frequency and high-frequency losses was addressed using spectral correction schemes based on analytical transfer functions and reference cospectral formulations derived from the Kaimal spectrum (Kaimal et al., 1972) and its subsequent developments (Moore, 1986; Moncrieff et al., 1997, 2004; Massman, 2000, 2001). After the above processing steps, residual outliers in the 30 min flux data were further removed by excluding the following time intervals: periods with mean horizontal wind speed lower than 1 m s^{-1} or higher than 8 m s^{-1}; periods when the absolute value of vertical wind velocity exceeded 0.15 m s^{-1}; and periods when the wind direction corresponded to directions obstructed by the mounting structure. After this additional screening, 72% of the EC turbulent flux data were retained."

Figure 4: Since the modeled data are unique in every graph, I suggest to plot these lines on top of the lines with the measured values. Currently it is the other way around.

Answer: Because we filtered the observed turbulent flux data used for model evaluation (see Section 3.1 for details), the time series in the revised manuscript differ from those in the original version and are

no longer continuous. To better compare observations and model simulations, we replaced the original time-series line plots with 30-minute-resolution scatter time series. In addition, we adopted the suggestion to adjust the plotting order so that the modeled values are displayed on top, thereby improving visibility while maintaining consistent styling across panels. The revised **Figure 4 (now Figure 5, L483)** is as follows:

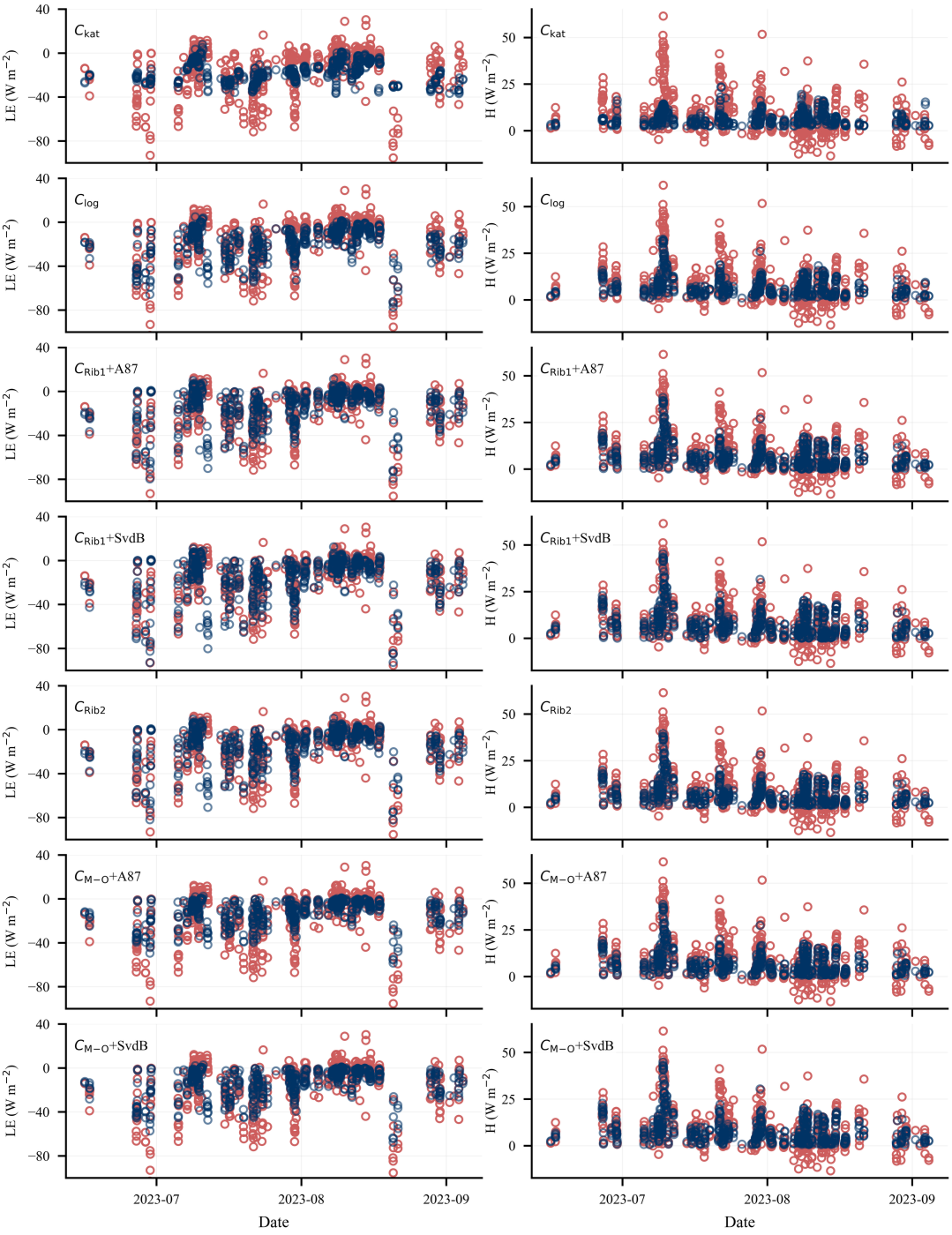


Figure 5: Time series of 30-min LE and H simulated by the five different schemes (C_{kat}, C_{log}, C_{Rib1}, C_{Rib2}, and C_{M-O}), compared with eddy covariance–derived values. The C_{Rib1}, C_{Rib2}, and C_{M-O} schemes were calculated by the observed aerodynamic roughness length (z_{0m} = 1.2 × 10^{-4} m), whereas C_{kat} and C_{log} were calculated using optimized parameters obtained by minimizing the differences between modeled and observed turbulent fluxes. Blue dots represent model simulations, and red dots denote observations.

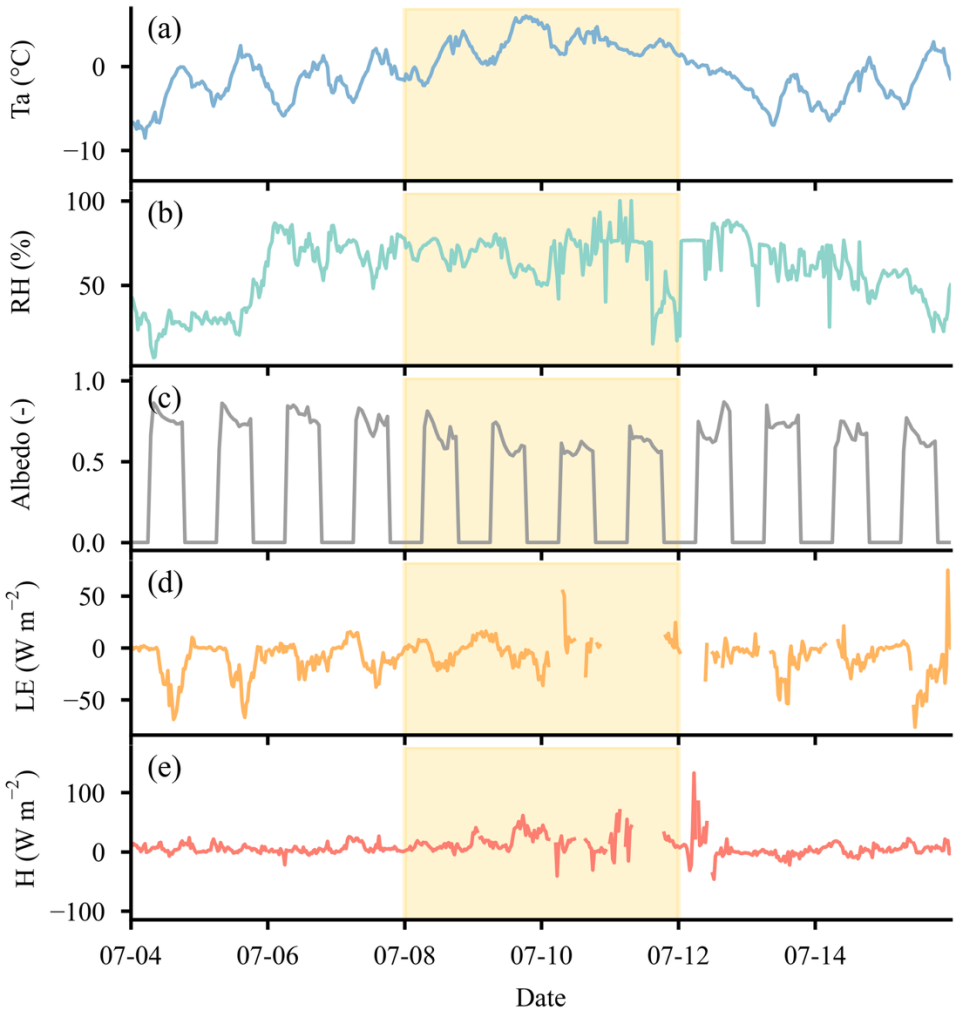
L462-463: "while CMB estimates ... -304.19 mm.". It is not surprising that there are biases as the mass balance model (EBFM), with standard values for model parameters, is not optimized for this region/glacier. Besides turbulent flux parameters, e.g. correct albedo parameters are at least equally important for the simulated mass balance. This should be acknowledged.

Answer: We agree with the reviewer that, in addition to turbulent flux parameterizations, other components such as albedo and longwave radiation are at least equally important for mass-balance simulations. To minimize the influence of these additional factors on the EBFM results, we clarified in **Sect. 5.1 (L610)** that, apart from the turbulent fluxes, all other model inputs (e.g., albedo, radiation, and meteorological forcing) are taken directly from AWS observations. This experimental design ensures that differences among simulations can be primarily attributed to the turbulent flux parameterizations, thereby allowing a meaningful and focused assessment of the sensitivity of the EBFM results to different turbulent-flux schemes. The revised text is provided below:

"To assess the importance of turbulent flux methods in simulating glacier mass balance, we included three representative turbulent flux methods (the C_{log}, C_{Rib2}, and C_{M-O} + A87 methods) in a coupled energy balance–snow and firn model applied at the point scale. For each method, turbulent fluxes were computed using both the original parameter sets reported in the literature and the recalibrated parameters developed in this study, while all other model inputs, including incoming and outgoing longwave radiation, incoming and outgoing shortwave radiation, snow depth, surface albedo, and rainfall, were prescribed from in situ observations at AWS1 to ensure a controlled experimental setup."

Figure 7: Why not zoom in only the heat wave period? The rest of the timeseries are also shown elsewhere, so no need to include them again here.

Answer: We agree that focusing the figure on the heatwave window improves clarity. However, because our discussion explicitly compares conditions during the heatwave period (8–11 July) with non-heatwave conditions in July, we retained the July context in the figure. In the revised manuscript, we therefore adjusted Figure 8 (formerly Figure 7) to emphasize the heatwave period by narrowing the time window from the full month of July to the heatwave interval together with four days before and after the event, allowing a clearer comparison with the surrounding conditions. In addition, albedo is now included in Figure 8 to support the subsequent discussion. The revised **Figure 8 (L663)** is as follows:



L506-507: "Such positive LE values ... maritime glaciers.". I agree, but also in the high Arctic where surface temperatures are typically much lower than air temperatures due to less incoming solar radiation.

Answer: We agree and revised the discussion to note that positive LE can also occur in the high Arctic, where radiative conditions and low surface temperatures can maintain strong humidity gradients and support condensation/deposition processes. The revised text **(L683)** is provided below:

"This shift implies a weakening of the upward latent heat flux and, at times, a reversal in flux direction, reflecting enhanced moisture condensation from the atmosphere onto the glacier surface and thus increased latent heat input (Zhang et al., 2017). Such weakly negative LE values, and occasional positive values, are more commonly observed over maritime glaciers and in high-latitude Arctic regions. In maritime settings, they are primarily associated with relatively high air temperatures and abundant precipitation, whereas in Arctic regions they are linked to weak solar radiation, which may occasionally cause glacier surface temperatures to fall below air temperature (Kuipers Munneke et al., 2009, 2018; Yang et al., 2011)."

L509: "18.53". Please be consistent in the use of significant digits (e.g. I recommend to use "18.5" here).

Answer: We standardized significant digits for consistency (and checked similar cases throughout the manuscript).

L512: "surface albedo decreased". I suppose the surface was already ice-covered at the start of the heat wave(?). If so, I do not see how the albedo dropped further during the heatwave. It would by the way be nice to see albedo time-series in the manuscript, e.g. in Fig. 2.

Answer: Thank you for raising this point. We carefully re-examined the albedo evolution and revised the text accordingly. For cases where a decrease in albedo is evident, we now provide a physically plausible interpretation grounded in site conditions and explicitly link it to the observed albedo record. During the humid heatwave period, elevated air temperatures (T_a) likely enhanced snowmelt at the glacier surface. The observation site is located in the accumulation area of the Dunde Glacier (5317 m a.s.l.), where a relatively thick snow cover generally persists throughout the ablation season (June–August), and widespread exposure of bare ice is unlikely. Under such warm and humid conditions, intensified melting can promote rapid snow metamorphism, characterized by grain growth and rounding, which is known to reduce surface albedo. As the humid heatwave persisted, sustained melting may have locally thinned the seasonal snow cover and altered surface conditions, potentially contributing to the observed decrease in surface albedo **(Fig. 9c)**. The revised text **(L693)** is provided below:

"This contrast underscores a distinct shift from turbulent heat loss to heat gain under humid heatwave conditions. Concurrently, rising T_a during the humid heatwave promoted snowmelt on the glacier surface, leading to rapid snow metamorphism characterized by grain growth and rounding. As the humid heatwave persisted, continued melting may have locally removed portions of the seasonal snow cover, potentially exposing underlying ice. This reduction in surface albedo enhanced melt energy, triggering intense ablation (Fig. 8c). Thus, although short-lived, such anomalous turbulent flux variations during the humid heatwave markedly altered the glacier's energy-balance processes by modifying albedo and melt energy, with potential implications for longer-term melt dynamics (Zhu et al., 2024b)."

L526: "warrant further investigation". I am not sure what is the cause either, but since the even the sign is off, I suppose it in some way must be related to the humidity gradient (which in turn depends on the temperature gradient).

Answer: We agree with the reviewer that the reversal in the sign of latent heat flux (LE) during the humid heatwave indicates a fundamental change in the near-surface humidity gradient, which is closely coupled with the temperature gradient. During this event, the observed sign reversal of LE implies that the direction of the vapor pressure gradient between the glacier surface and the overlying air temporarily reversed. However, none of the evaluated schemes reproduced this observed sign reversal, highlighting limitations in the representation of near-surface humidity gradients and surface wetness under extreme meteorological conditions.

The revised manuscript therefore emphasizes that the underestimated variability may partly arise from observational limitations. In particular, the meteorological variables used to drive the models are based on half-hourly observations. The averaging procedure may obscure short-lived temperature and humidity gradients near the surface, thereby limiting the models' ability to reproduce such transient phenomena. This issue clearly warrants further investigation. Future work will focus on targeted observations and higher-resolution process-based modeling to better resolve these dynamics.

The revised text **(L716)** is provided below:

"Overall, the methods, particularly the C_{M-O} + SvdB scheme, reproduced the anomalous increase in H during the humid heatwave with relatively high consistency and accuracy, and captured the decreasing tendency of LE during this period. However, the estimation errors for LE were non-negligible, and the models failed to reproduce the occasional reversal in LE direction (on July 11). Part of the

discrepancy between modeled and observed results may be attributed to observational limitations. The meteorological variables used to drive the models are based on half-hourly observations. Such temporal averaging may obscure short-lived near-surface temperature and humidity gradients, thereby limiting the models' ability to reproduce transient fluctuations. Nevertheless, the discrepancies in turbulent fluxes identified in this study suggest that current parameterization schemes may underestimate the mass loss driven by LE under extreme meteorological conditions, thereby underestimating its role in glacier melt processes."

L528-531: "All calculated values ... to glacier melting.". This applies to the uncalibrated turbulent flux models. It would be great to see if that still applies after calibrating the turbulent heat fluxes (e.g. by tuning the roughness lengths).

Answer: We agree with the reviewer. In the revised manuscript, our discussion and conclusions are now based on the turbulent-flux schemes implemented using the observed aerodynamic roughness length (z_{om}). The results indicate that, even when applying the observed z_{om}—particularly given that a substantial amount of quality-controlled turbulent flux data during the extreme event was retained for evaluation, suggesting that the observed z_{om} should be applicable to this period—all schemes still fail to fully reproduce the reduction in the absolute magnitude of LE and the occasional reversal in its direction during the extreme event. Nevertheless, they are able to capture the pronounced increase in H.

These findings indicate that, current parameterization schemes may underestimate the mass loss driven by LE under extreme meteorological conditions, thereby underestimating its role in glacier melt processes. Therefore, our qualitative conclusions remain unchanged after calibration, but are now more explicitly framed in the context of calibrated model performance.

The revised text **(L716)** is provided below:

"Overall, the methods, particularly the C_{M-O} + SvdB scheme, reproduced the anomalous increase in H during the humid heatwave with relatively high consistency and accuracy, and captured the decreasing tendency of LE during this period. However, the estimation errors for LE were non-negligible, and the models failed to reproduce the occasional reversal in LE direction (on July 11). Part of the discrepancy between modeled and observed results may be attributed to observational limitations. The meteorological variables used to drive the models are based on half-hourly observations. Such temporal averaging may obscure short-lived near-surface temperature and humidity gradients, thereby limiting the models' ability

to reproduce transient fluctuations. Nevertheless, the discrepancies in turbulent fluxes identified in this study suggest that current parameterization schemes may underestimate the mass loss driven by LE under extreme meteorological conditions, thereby underestimating its role in glacier melt processes.”

L535-536: "to increasingly ... increased H. ". It could be worth emphasizing that it is primarily that the surface temperature cannot exceed 0 ºC that causes positive LE and higher H during such extreme events.

Answer: We agree with the reviewer’s point. In the revised manuscript, we now emphasize this mechanism more explicitly in the Discussion. During extreme events, the glacier surface temperature is constrained by the melting point and cannot exceed 0 °C, whereas air temperature continues to rise. As a result, the surface–air temperature difference becomes anomalously large during heatwave conditions, which directly enhances the sensible heat flux (H).

Highlighting this constraint improves the physical interpretation of the observed increase in H. The revised text **(L729)** is provided below:

“Such conditions will cause continental glaciers, such as the Dunde Glacier, to increasingly exhibit turbulent flux characteristics that resemble those of maritime glaciers, including reduced LE magnitudes and even reversals in LE direction, as well as a pronounced enhancement of H. This phenomenon warrants further investigation and is likely related to the constraint that T_s cannot exceed 0 °C. Under humid heatwave conditions, T_a continues to rise while T_s is capped, leading to an increased surface temperature deficit and consequently a substantial enhancement of H. Meanwhile, because the humidity gradient is closely coupled to the temperature gradient, LE may also undergo anomalous changes. Consequently, continued warming may drive future transitions in glacier types.”