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.

General comments

1. As the results of this manuscript hinge on the measured fluxes, it is important to have reliable, reproducible flux measurements. However, there is no mention of how any of the turbulence data are processed. What coordinate system rotation is applied, if any? Is any filtering applied? Are any frequency corrections applied? Over what interval are the fluxes calculated? This last point is particularly important, as there is increasing evidence that 30 min is likely too long an averaging interval to capture turbulent fluxes on glaciers [Mott et al.,2020, Nicholson and Stiperski, 2020, Lord-May and Radi´c, 2024].

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."

Regarding the choice of the flux averaging interval, turbulent fluxes were calculated over a 30 min averaging period, which remains the most commonly adopted practice in glacier eddy-covariance studies and provides a well-established balance between statistical convergence, data availability, and comparability across sites and studies. We acknowledge recent work suggesting that shorter averaging intervals may better capture non-stationary turbulence and sub-mesoscale processes over glacier surfaces (e.g., Mott et al., 2020; Nicholson and Stiperski, 2020; Lord-May and Radić, 2024).

In the present study, several scientific and practical considerations motivated the use of a 30 min interval. First, the primary objective of this study is a multi-month, integrated evaluation of turbulent flux parameterizations and their implications for glacier energy and mass balance, for which temporal consistency and comparability with previous glacier studies are essential. Second, the long analysis period partially mitigates the influence of short-term non-stationarity on the overall statistical assessment of model performance.

We therefore retain the 30 min averaging interval in this study while explicitly acknowledging it as a methodological limitation. Future work will focus on exploring shorter averaging intervals, to better resolve intermittent turbulence over glacier surfaces and to further assess its impact on turbulent flux estimates.

2. The methods section lacks a clear flow and connection between the various bulk methods introduced. The connection is not made clear that many of these models differ only in their treatment of stability within the bulk exchange coefficient. I suggest a rewrite with a clear derivation, starting with the Clog method. All assumptions should be clearly presented throughout. You make this connection on L395, but this should be clarified much sooner.

Answer: We agree that the original Methods section did not sufficiently guide readers in distinguishing between the common elements shared by the schemes and their key differences. We therefore reorganized the Methods framework by first establishing a unified bulk-aerodynamic theoretical basis, and subsequently describing each scheme primarily in terms of its stability treatment and whether katabatic wind effects are considered. This restructuring improves readability and reduces the risk of misinterpreting minor structural differences as fundamental differences in the underlying physical assumptions. The revised sentence **(L239)** now reads:

3.2 Bulk methods

Variations in the performance of different turbulent flux parameterizations primarily arise from their respective approaches to computing turbulent exchange coefficients and applying atmospheric stability corrections. These methodological differences directly affect method accuracy under varying temperature gradients, humidity gradients, and wind speed conditions, leading to discrepancies in the simulated LE and H. In this study, we evaluate two categories of bulk aerodynamic methods for calculating turbulent fluxes over glacier surfaces. The first category comprises methods in which the bulk exchange coefficient (C) is formulated as a function of surface roughness lengths and atmospheric stability. The second category includes bulk methods that explicitly incorporate a simplified katabatic (glacier wind) parameterization.

Within this framework, we analyze five bulk methods that are commonly applied to glacier surfaces. These include: (i) a simplified bulk formulation without explicit stability corrections (C_{log}); (ii) and (iii) bulk methods that account for atmospheric stability through the bulk Richardson number (C_{Rib1} and C_{Rib2}); (iv) a method based on the full Monin–Obukhov similarity theory, employing universal stability functions with iterative closure (C_{M-O}); and (v) a bulk method derived from a simplified katabatic flow model that explicitly considers glacier wind effects (C_{kat}). The detailed formulations of methods (i) through (v) are described in the following subsections.

Model performance is evaluated by comparing 30-min turbulent fluxes (H and LE) simulated by each method with eddy covariance (EC) measurements, using standard statistical metrics including the root mean square error (RMSE), mean bias error (MBE), and mean absolute deviation (MAD). For consistency between modeled and observed fluxes, turbulent fluxes are defined as positive when directed from the atmosphere toward the glacier surface, and negative in the opposite direction. The naming of the turbulent flux methods follows the convention of Radić et al. (2017).

3.2.1. C_{kat} method

The C_{kat} method explicitly considers katabatic flows, introducing the katabatic bulk exchange coefficient (C_{kat}) to partially overcome the limitations of MOST during strong katabatic wind conditions (Oerlemans and Grisogono, 2002). This approach emphasizes intensified katabatic turbulence during nocturnal cooling or cold-air intrusions, enhancing turbulent exchange (Horst and Doran, 1988). However, the C_{kat} method lacks explicit stability corrections and dynamic roughness length adjustment, leading to a slow response under rapidly changing surface conditions (Denby, 2000). Hence, its application is more suited to small and medium-sized glaciers rather than large-scale domains. LE and H are expressed as:

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

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

where, ρ and C_P are the density (kg·m^{-3}) and heat capacity of air, respectively; P is atmospheric pressure (Pa), $L_{s/f}$ is the latent heat of sublimation/fusion, selected based on surface temperature; e_a and e_s are the atmospheric vapor pressure and saturated vapor pressure at the glacier surface (Pa), respectively; T_a and T_s denote the air temperature and glacier surface temperature, respectively; and C_{kat} is the katabatic bulk exchange coefficient, calculated using the following equation:

$$C_{\mathrm{kat}} = -C_{tub} C_{tub_2}^2 C \left(\frac{g}{T_0 \gamma P_r}\right)^{1/2}, \tag{9}$$

Here, C_{tub} and C_{tub2} are dimensionless empirical constants used to optimize the parameterization of turbulent fluxes; C is replaced by the air–surface temperature difference (T_a–T_s); T_0= 273.15 K; γ is the potential temperature gradient, which is prescribed as a constant value of 0.005 K m^{-1} (Oerlemans and Grisogono, 2002); and P_r is the Prandtl number (~ 0.71).

3.2.2. C_{log} method

The C_{log} method represents a highly simplified derivative of MOST. This method employs a constant exchange coefficient (C_h) driven by near-surface wind speed and the difference in air temperature (or humidity) and surface temperature (or humidity), providing a simplified, computationally efficient structure suitable for large-scale climate models (Oerlemans, 2000). It does not dynamically adjust atmospheric stability, and instead retains only a linear relationship for surface fluxes. Being a wind-speed-driven scheme, it performs well under less stable stratification conditions. However, the method also exhibits certain limitations. Owing to the absence of atmospheric stability parameters in its structure, it cannot identify or respond to surface inversions forming at night or in the early morning. LE and H are expressed as:

$$LE = 0.622\rho L_{s/f} C_h u(e_a - e_s)/P, \tag{10}$$

$$H = \rho C_P C_h u(T_a - T_s), \tag{11}$$

where, u is wind speed (m·s^{-1}), and C_h is the turbulent exchange coefficient. Following (Oerlemans, 2000), turbulent fluxes are regulated through the exchange coefficient C_h. Here, C_h is treated as a calibration parameter and optimized by minimizing the mismatch between modeled and observed turbulent fluxes.

3.2.3. C_{Rib1} method

The C_{Rib1} method retains the C_{log} formulation for turbulent flux calculation but introduces the bulk Richardson number (Rib) to allow for flux reduction under stable stratification (Suter et al., 2004). The bulk Richardson number can be related to the stability functions for momentum (ϕ_m), heat (ϕ_h), and water vapour (ϕ_q) in the following way ((Oke, 1987); see Eq. 14). In this formulation, the composite function ($f_h(Ri_b)$) is constructed from the stability functions for momentum (ϕ_m), heat (ϕ_h), and water vapor (ϕ_q) (Dyer, 1974; Holtslag and Bruin, 1988).

$$LE = \rho L_{s/f} \kappa^2 z_u z_q \left(\frac{\Delta\bar{u}\Delta\bar{q}}{z^2}\right)(\phi_m\phi_h)^{-1}, \tag{12}$$

$$H = \rho C_P \kappa^2 z_u z_t \left(\frac{\Delta\bar{u}\Delta\bar{T}}{z^2}\right)(\phi_m\phi_q)^{-1}, \tag{13}$$

$$\begin{cases} (\phi_m\phi_{h,q})^{-1} = (1 - 5Ri_b)^2 & (Ri_b > 0) \\ (\phi_m\phi_{h,q})^{-1} = (1 - 16Ri_b)^{0.75} & (Ri_b \le 0) \end{cases}, \tag{14}$$

$$Ri_b = \frac{g\frac{\Delta\bar{T}}{\Delta Z}}{\bar{T}\left(\frac{\Delta\bar{u}}{\Delta Z}\right)^2}, \tag{15}$$

where, $\bar{u}$, $\bar{q}$, and $\bar{T}$ represent the mean wind speed, specific humidity, and air temperature, respectively; where $z_{u,t,q}$ are the log mean heights defined as：

$$z_{u,t,q} = \frac{z - z_{0u,t,q}}{ln\left(\frac{z}{z_{0u,t,q}}\right)}, \quad (16)$$

and $z_{0u,q,t}$ denotes the surface roughness lengths for momentum, humidity, and temperature.

3.2.4. C_{Rib2} method

Similar to C_{Rib1}, the C_{Rib2} method is a non-iterative method for calculating turbulent fluxes (Essery and Etchevers, 2004). It employs Ri_b and $f_h(Ri_b)$, thereby reducing computational complexity while preserving the physical consistency of MOST. Turbulent fluxes in the C_{Rib2} method are calculated as:

$$LE = \rho L_{s/f} C_H u[q_{sat}(T_s, P) - q], \quad (17)$$

$$H = \rho C_P C_H u[T_s - T_a], \quad (18)$$

where, $q_{sat}(T_s, P)$ denotes the saturation specific humidity at surface temperature T_s and pressure P, and C_H is a surface exchange coefficient. Following Essery and Etchevers (2004) and Louis (1979), the exchange coefficient for surface sensible and latent heat flux is calculated as $C_H = C_{Hn} f_h$, where

$$C_{Hn} = \kappa^2 \left[\ln\left(\frac{z}{z_{0m}}\right)\right]^{-2}, \quad (19)$$

is the neutral exchange coefficient for roughness length z_{0m} and

$$f_h = \begin{cases} (1 + 10Ri_b)^{-1} & Ri_b \geq 0 \\ 1 - 10Ri_b\left(1 + 10Ri_b C_{Hn}\sqrt{-Ri_b}/f_z\right)^{-1} & Ri_b < 0 \end{cases}, \quad (20)$$

with

$$f_z = \frac{1}{4}\left(\frac{z_{0m}}{z}\right)^{1/2}, \quad (21)$$

Although atmospheric stability is also characterized by Ri_b, the calculation approach in C_{Rib2} method differs from that in C_{Rib1} method as follows:

$$Ri_b = \frac{g z_1}{u^2}\left\{\frac{T_a - T_s}{T_a} + \frac{q - q_{sat}(T_s, P)}{q + \epsilon/(1-\epsilon)}\right\}, \quad (22)$$

where, ϵ is defined as the ratio of the molecular weights of water to dry air, with a value of 0.622.

3.2.5. C_{M-O} method

This method replaces the composite stability function $f_h(Ri_b)$ used in the C_{Rib} with the universal stability functions (ψ) to introduce stability corrections. It adopts a complete MOST framework and explicitly calculates the friction velocity (u^*), ψ, and L (Hock and Holmgren, 2005). Separate roughness lengths are applied for momentum (z_{0m}), temperature (z_{0t}) and humidity (z_{0e}). Turbulent fluxes are obtained through an iterative solution procedure that ensures closure of the nonlinear stability equations.Under stable conditions, the nonlinear stability functions of Beljaars and Holtslag (1991) are

used, whereas under unstable conditions, the Businger–Dyer relationships are applied (Beljaars and Holtslag, 1991; Paulson, 1970). The variables $u*$ and L solved iteratively to achieve convergence.

$$LE = L_{s/f}\frac{0.622\rho}{P_0}\frac{\mathrm{k}^2}{\left[\ln\left(\frac{z}{z_{0m}}\right)-\psi_M\left(\frac{z}{L}\right)\right]\left[\ln\left(\frac{z}{z_{0e}}\right)-\psi_E\left(\frac{z}{L}\right)\right]}u(e_a - e_s), \quad (23)$$

$$H = \rho C_P\frac{\mathrm{k}^2}{\left[\ln\left(\frac{z}{z_{0m}}\right)-\psi_M\left(\frac{z}{L}\right)\right]\left[\ln\left(\frac{z}{z_{0t}}\right)-\psi_H\left(\frac{z}{L}\right)\right]}u(T_a - T_s)\,, \quad (24)$$

For clarity and reproducibility, all model-specific parameters and physical constants used in this study are summarized in Table 2, together with their numerical values, units, and data sources where applicable."

3. There are no tests of robustness or statistical significance in this manuscript. Considering how low the average fluxes are, this should be included to instill confidence in the results. At present, I am somewhat wary of the model results, as the models do not appear to accurately simulate the fluxes. For example, the average spring LE is−15.1 W/m2, but the best performing model had an RMSE of roughly 8.5 W/m2. This is a relative error of over 50%. For H, the relative error of the best-performing model exceeds 100%. Without a more clear breakdown of the flux processing methods, a section discussing modelled roughness lengths, and scatter plots to better present the measured vs. modelled heat fluxes, it's challenging to determine whether this performance is due to the models not applying, or whether there has been a procedural error. I do have several concerns about the application of the bulk methods, which I describe in my specific comments.

Answer: We thank the reviewer for raising this important issue. We agree that, in the original manuscript, the absence of explicit robustness tests and statistical significance analyses limited the confidence with which model performance could be interpreted. To address this concern, we substantially revised the manuscript by adding a dedicated statistical assessment of model biases and their significance, complemented by scatter plots and an expanded description of flux processing and roughness-length treatment. The revised sentence **(L440)** now reads:

4.3.1 Model–observation correspondence and statistical significance

The scatter plots reveal pronounced differences among the five evaluated schemes in estimating turbulent heat fluxes over the Dunde Glacier (Fig. 4). Overall, all methods exhibit relatively higher accuracy for both latent and sensible heat fluxes when the absolute flux magnitudes are small, with modeled values showing good agreement with the observations. However, as the absolute magnitude of

the fluxes increases, the simulation errors tend to increase, accompanied by a gradual decline in model performance. This behavior indicates that the stability of the parameterization schemes under higher flux conditions is limited, leading to reduced reliability of the modeled turbulent fluxes.

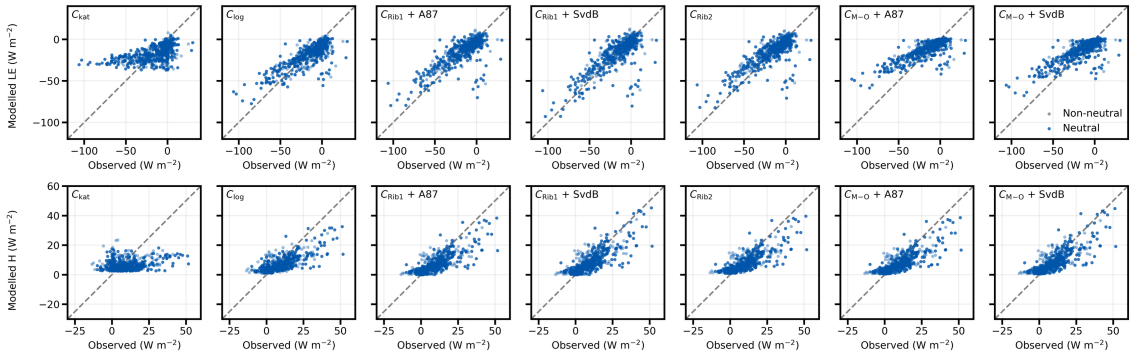


Figure 4: Comparison of 30-min observed (eddy-covariance-derived) and modeled H and LE according to the five turbulent flux methods.

Under near-neutral conditions, most parameterization schemes exhibit small mean biases in LE, with 95% confidence intervals encompassing zero. This suggests that, within the estimated uncertainty range, the majority of schemes do not show statistically significant systematic deviations from the eddy-covariance observations in terms of mean behavior. In contrast, all schemes display a consistent negative bias in H, with confidence intervals that do not include zero. This systematic underestimation may partly arise from uncertainties in the surface–air temperature gradient, such as potential errors in glacier surface temperature estimation, rather than from random variability alone. These findings highlight potential limitations of bulk turbulent flux parameterizations over glacier surfaces and underscore the need for careful interpretation of modeled H under near-neutral atmospheric conditions.

Table 3. Mean bias (Bias) and associated 95% confidence intervals of modeled latent heat flux (LE) and sensible heat flux (H) under near-neutral conditions for the five turbulent flux parameterization schemes. Positive (negative) bias values indicate overestimation (underestimation) relative to eddy-covariance (EC) observations.

	LE (W m^{-2})	H (W m^{-2})

Method	Bias (W m^{-2})	95% CI (W m^{-2})	Bias (W m^{-2})	95% CI (W m^{-2})
C_{kat}	1.2	[−3.0, 5.3]	−4.3	[−7.4, −1.6]
C_{log}	−2.2	[−5.3, 0.5]	−3.1	[−4.7, −1.7]
C_{Rib1} + A87	1.7	[−1.8, 4.6]	−3.0	[−4.2, −1.8]
C_{Rib1} + SvdB	−0.8	[−4.4, 2.0]	−1.7	[−2.7, −0.7]
C_{Rib2}	0.8	[−2.5, 3.8]	−2.4	[−3.7, −1.2]
$C_{M–O}$ + A87	4.4	[0.8, 8.1]	−2.7	[−4.0, −1.5]
$C_{M–O}$ + SvdB	2.8	[−0.5, 6.0]	−1.7	[−2.7, −0.7]

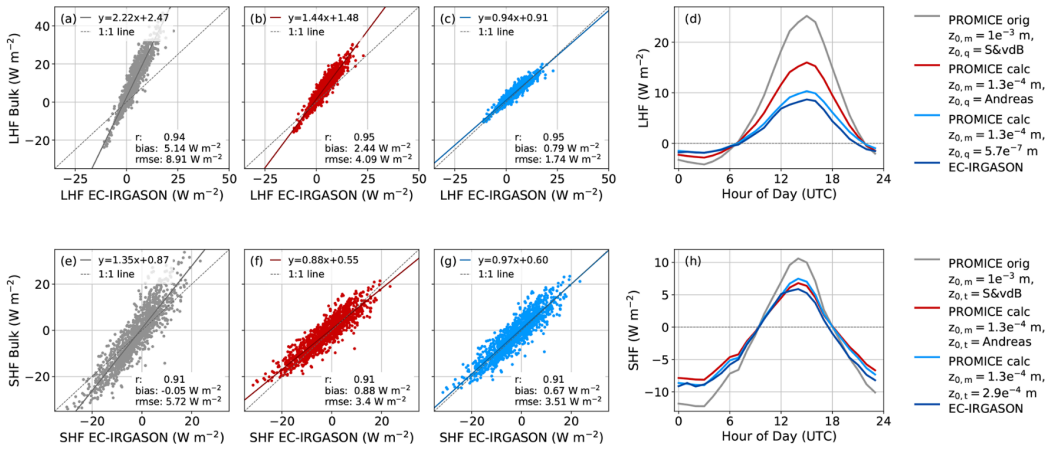
With respect to the reviewer's concern regarding the relatively large RMSE values compared to the seasonal mean fluxes, it should be noted that low mean turbulent flux magnitudes are a common characteristic of glacier environments. Under such conditions, previous studies have frequently reported RMSE values for both LE and H in the range of approximately 5–30 W m^{-2}, which does not necessarily indicate procedural errors in flux processing (Haven et al., 2025; Radić et al., 2017; Guo et al. 2011).

For example, Radić et al. evaluated turbulent-flux parameterizations in the Cariboo Mountains, British Columbia, Canada, and found that the RMSE between modeled and observed data ranged from 6.5 to 34.8 W m^{-2}:

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. reported that, using the bulk method combined with various roughness-length parameterization schemes, the RMSE between observed turbulent fluxes and modeled values over the Greenland Ice Sheet ranged from 1.7 to 8.9 W m^{-2}. In the present study, due to the relatively long observation period, the number of turbulent flux data points used for evaluation was larger and covered a longer time span. As a result, the RMSE between observations and simulations may be higher than in studies with shorter evaluation periods, but it remains within a reasonable range.



In addition, we expanded the Methods section to provide a detailed description of the EC flux processing chain and the derivation of surface roughness lengths. In this study, a temporally invariant aerodynamic roughness length z0m was applied throughout the entire analysis period. This choice ensured methodological consistency and allowed for a systematic evaluation of turbulent flux model performance under a unified roughness length and scalar roughness length parameterization framework.

Overall, the revised manuscript provides a more transparent and statistically robust basis for interpreting the model results.

The revised sentence (**L170-199、L201-238**) now reads:

2.3. Eddy-Covariance systems

Turbulent fluxes were measured with a CSAT3 three-dimensional sonic anemometer (Campbell IRGASON) and subsequently processed using the eddy covariance (EC) method during the period from 14 May to 12 October 2023; this is a widely adopted technique in micrometeorological research that enables real-time, accurate, and continuous monitoring of atmospheric turbulence. All turbulence raw data were collected at 10 Hz, including the three components of wind velocity, virtual temperature, and water vapor concentration.

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, 2005). Coordinate rotation was applied using the double-rotation method to correct for sensor tilt and streamline distortion (Kaimal and Kristensen, 1991). 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, 2005; 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.

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.

4. The language needs to be re-evaluated throughout to provide a precise description of the data at hand. For example, words including "marked", "notable", and "pronounced", often overexaggerate throughout this manuscript.

Answer: We thank the reviewer for this constructive comment and fully agree that precise and restrained language is essential for accurately describing the data and avoiding unintended overinterpretation. In response, we have carefully re-evaluated the language throughout the entire manuscript with particular attention to qualitative descriptors.

Specifically, we revised instances where terms such as "marked", "notable", and "pronounced" could overstate the magnitude or significance of the reported patterns. These expressions were replaced with more neutral, quantitative. This systematic language revision was applied consistently across the Results and Discussion sections.

5. The structure needs to be revised, as it lacks proper, deep discussion section that explores what these findings mean in a broader context. 5.1 Belongs in the methods, and most of 5.2 and 5.3 are results.

Answer: We thank the reviewer for this important comment regarding the manuscript structure. We agree that, in the original version, the separation between Methods, Results, and Discussion was not sufficiently clear, which may have obscured the distinct roles of each section.

Following this comment, we carefully reorganized the manuscript. The former Section 5.1 has now been fully relocated to the Methods section. This restructuring improves the logical flow of the Methods and ensures that all methodological descriptions are presented in a single, coherent framework.

In the revised manuscript, we introduced a new discussion on the characteristics of roughness length across different glacier types (Section 5.1). Together with the original Sections 5.2 and 5.3, these sections now collectively constitute the Discussion part of the manuscript. We would like to clarify that these

sections are deliberately framed as Discussion rather than Results. Although they are directly based on the quantitative findings presented in Section 4, their primary purpose is not to introduce new results, but to elaborate on the broader implications and applications of those findings. Specifically, these sections aim to provide guidance for roughness length selection in similar studies, and to clarify how different turbulent flux parameterization schemes influence glacier energy and mass balance simulations, as well as to evaluate their applicability under varying climatic conditions. Section 5.1 highlights the current scarcity of in situ roughness length measurements for continental glaciers on the Tibetan Plateau, and emphasizes that roughness lengths derived from other glacier types cannot be directly transferred to this setting. By comparing roughness length characteristics across different glacier types, we provide a reasonable range for roughness length selection for glaciers similar to that investigated in this study. Section 5.2 focuses on how improvements in turbulent flux representation translate into enhanced performance of SEB and mass-balance models, thereby placing the results within a modeling and process-oriented framework that extends beyond simple statistical comparison. Section 5.3 further broadens this perspective by examining the behavior and limitations of the turbulent flux schemes under extreme weather conditions, highlighting their broader implications for glacier response under ongoing climate warming. Together, these discussions emphasize the applied value of the results rather than presenting new quantitative findings.

The revised now reads:

3 Methods

3.1 The algorithm and quality control for roughness lengths

3.2 Bulk methods

3.3 Energy Balance Model

5 Discussion

5.1 Glacier type dependence of aerodynamic roughness length variability

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

5.3 Evaluation of turbulent flux methods under extreme weather conditions

6. It is somewhat unclear exactly what knowledge gap this study aims to fill, or how future studies should use their findings. What does it mean for one method to have better spring performance than another?

How could a future study leverage this information? This study would benefit from a more systematic exploration of why certain methods outperform others.

Answer: We sincerely thank the reviewer for this insightful comment. We fully agree that it is essential to clearly articulate the specific knowledge gap addressed by this study and the practical implications of its findings. Accordingly, we take this opportunity to clarify the particular knowledge gap filled by our work and to explain the broader scientific and practical significance of our results.

The primary knowledge gap addressed by this study can be summarized in four aspects.

First, this study represents the first systematic observation-based characterization of turbulent flux variability over a continental glacier on the TP. Focusing on the Dunde Glacier, eddy covariance measurements were conducted at the glacier summit within the high-altitude accumulation zone over an extended observational period spanning multiple seasons. This enables a comprehensive and quantitative assessment of turbulent flux behavior under diverse meteorological conditions, including overall variability, seasonal evolution, and diurnal patterns. To date, direct long-term turbulent flux measurements over continental glaciers on the TP have been lacking. Consequently, previous energy and mass balance simulations could not be independently validated in terms of turbulent flux representation, making it difficult to constrain the magnitude and numerical characteristics of turbulent exchange processes in these environments. Furthermore, due to the high logistical and maintenance costs of glacier instrumentation, most earlier measurements were conducted near glacier termini and over relatively short periods, often restricted to the ablation season. Long-term eddy covariance observations at glacier summits and high-elevation accumulation zones have therefore been largely absent. In this context, the present study fills a critical observational gap and provides a systematic reference for understanding turbulent flux variability over continental glaciers on the TP.

L116: 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 in the Qilian Mountains, the northeastern TP

Second, this study provides an observation-based reference for the selection of turbulent flux and roughness length parameterizations over continental glaciers on the TP. To our knowledge, this is the first comprehensive assessment of multiple widely used bulk turbulent flux formulations based on continuous eddy covariance measurements over a continental glacier in this region. Previous evaluations have primarily focused on polar ice sheets or maritime glaciers, with relatively few studies conducted on

glaciers within the TP. Within the Plateau itself, existing work has largely been limited to maritime glaciers, such as studies on Parlung No. 4 Glacier during the May–September 2009 period. However, due to fundamental differences in climatic regime, melt intensity, and surface characteristics between maritime and continental glaciers, parameterizations derived from maritime glacier observations cannot be directly generalized to the widely distributed continental glaciers on the TP. In this context, the present study provides a much-needed reference framework for selecting appropriate roughness lengths and turbulent flux models for glaciers under similar continental climatic conditions.

L111: 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).

Third, this study establishes a transferable framework for the selection, interpretation, and improvement of turbulent flux parameterizations within glacier energy- and mass-balance modeling. Due to practical constraints, many glacier change studies rely heavily on energy and mass balance models, yet the turbulent flux components within these models are often weakly constrained by observations, leading to considerable uncertainty in total energy budget simulations. By systematically incorporating different turbulent flux parameterizations into the energy balance framework, this study demonstrates that optimizing turbulent flux parameters can improve the overall performance of energy and mass balance simulations. This contributes to a more accurate quantification of surface energy partitioning and improves our understanding of the climatic controls governing glacier change on the TP.

L64: However, to date, there have been no reported observations of turbulent fluxes over continental glaciers on the TP characterized by low temperatures and precipitation, one of the most widely distributed glacier types on the TP. 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.

Finally, the observational period in this study spans multiple temporal scales and explicitly includes extreme weather conditions. The analysis of model behavior under such conditions is not intended as a purely descriptive ranking exercise, but rather provides diagnostic insight into the applicability and

limitations of different parameterization schemes. Importantly, the results suggest that systematic underestimation of turbulent fluxes during extreme events may lead to an incomplete assessment of glacier response under increasingly frequent climate extremes.

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 (Brun et al., 2018; Duan et al., 2012; Hugonnet et al., 2021; Yao et al., 2012). Therefore, a comprehensive understanding of the magnitude and variability of turbulent fluxes is essential to accurately quantify glacier SEB and project future glacier mass-balance changes.

Specific comments

L18: Here, and elsewhere: use daily instead of diurnal.

Answer: Thank you for the correction. We agree that "daily" is the appropriate term in the contexts flagged, and we revised the manuscript accordingly. We systematically replaced "diurnal" with "daily" where we refer to day-to-day statistics or daily aggregates, ensuring consistent terminology across the Introduction, Results, and figure captions.

L53: Estimation using numerical models is not the same as attaining turbulent flux data.

Answer: We agree and revised the wording to clearly distinguish between *measured* turbulent fluxes (derived from sonic-anemometer observations processed with the eddy covariance method) and modelled/estimated turbulent fluxes computed using bulk aerodynamic parameterizations. The revised text avoids implying that numerical modelling "attains" the observational turbulent flux data. The revised sentence **(L57)** now reads: "Currently, glacier turbulent fluxes are derived through two main approaches: direct measurements using the eddy covariance (EC) method and model-based estimates using numerical models."

L62: Is it the processes that aren't well understood? Or the relative contributions of these processes to the overall budget not understood?

Answer: Thank you for pointing out the ambiguity. We revised the sentence to clarify that the primary uncertainty lies in the relative contributions of different surface energy-balance components (and

their parameterized representations) rather than suggesting that the physical processes themselves are fundamentally unknown. This improves precision and aligns the stated motivation with the actual scope of our analysis. The revised sentence **(L64)** now reads: "However, to date, there have been no reported observations of turbulent fluxes over continental glaciers on the TP characterized by low temperatures and precipitation, one of the most widely distributed glacier types on the TP. 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."

L63: I'm not sure I understand how better flux measurements leads to a better understanding of glacier variations across the TP.

Answer: We thank the reviewer for pointing out that this causal relationship was not sufficiently articulated in the original text. Our intention was to emphasize the critical role of accurate turbulent flux measurements in constraining the evolution of glacier surface energy mass balance over the Tibetan Plateau.

Turbulent fluxes, including latent and sensible heat fluxes, constitute a substantial and highly variable component of the glacier surface energy balance on the Tibetan Plateau. Improved and well-constrained flux measurements enable a more reliable quantification of these energy exchange processes, which directly govern surface melt, sublimation, and refreezing. This, in turn, allows for more robust calibration and evaluation of surface energy- and mass-balance models that are widely used to interpret observed glacier changes and to project future glacier evolution across the Tibetan Plateau.

We have revised the corresponding statement in the manuscript to explicitly clarify this linkage, emphasizing that improved turbulent flux measurements deepen our understanding of glacier variations by reducing uncertainties in energy-balance partitioning and mass-balance attribution, rather than by providing descriptive information alone. The revised sentence **(L64)** now reads: "However, to date, there have been no reported observations of turbulent fluxes over continental glaciers on the TP characterized by low temperatures and precipitation, one of the most widely distributed glacier types on the TP. 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."

L67: MO theory is not computationally intensive. Its accuracy, especially on glaciers with strong stability and katabatic winds, has been questioned. The assumptions underpinning MO theory fundamentally break down when fluxes are not constant in height (often the case over glaciers) [Grisogono et al., 2007]. This is part of what the stability corrections aim to address.

Answer: We appreciate this important clarification. We revised the text to remove the statement implying that Monin–Obukhov similarity theory (MOST) is computationally intensive. We also acknowledge the known limitations of MOST over glacier surfaces under strongly stable stratification and katabatic-flow regimes, particularly in situations where the constant-flux-layer assumption breaks down and turbulent fluxes vary with height (e.g., Grisogono et al., 2007). These limitations motivate the use of stability correction functions within bulk formulations, while also highlighting the intrinsic challenges of applying MOST over complex glacier boundary layers. The revised sentence **(L72)** 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."

L71: Oerlemans 2000 does not do this.

Answer: Thank you for catching this. We re-checked Oerlemans (2000) and agree that the manuscript statement was inaccurate. We corrected the text to reflect what is actually implemented/discussed in that study, and removed any misleading implication. The revised sentence **(L79)** now reads: "To improve computational efficiency, Oerlemans (2000) applied a highly simplified bulk formulation to estimate near-surface turbulent fluxes, assuming logarithmic vertical profiles of wind speed, temperature, and humidity to be valid under prevailing stable conditions, without explicitly accounting for stability-dependent flux modifications."

L72: Stability functions have existed for many decades to modify turbulent fluxes predicted from bulk methods to account for suppressed fluxes due to stable stratification. Importantly, these functions do not themselves simulate turbulent fluxes. I would suggest Louis 1979 and other fundamental papers here.

Answer: We agree. We revised the text to clarify that stability functions do not themselves "simulate" turbulent fluxes; rather, they modify bulk transfer coefficients to account for suppressed exchange under

stable stratification. We also added Louis (1979) and other foundational references to appropriately ground the discussion. The revised sentence **(L74)** now reads: "However, its performance over glaciers remains uncertain, particularly under strongly stable stratification and katabatic flow conditions, where the fundamental assumption of vertically constant turbulent fluxes is frequently violated (Grisogono et al., 2007)."

L76: I do not understand this transition — how do poorly performing flux models relate to winter and summertime flux measurements?

Answer: We thank the reviewer for pointing out that this transition was not sufficiently clear in the original text. Our intention was to emphasize that differences among turbulent flux parameterizations not only lead to divergent estimates of turbulent fluxes at the same site, but can also result in fundamentally different interpretations of glacier surface energy and mass balance processes.

Specifically, under similar meteorological forcing, different bulk aerodynamic formulations may produce substantially different turbulent flux estimates for the same glacier, leading to contrasting multi-year seasonal mean values. As demonstrated by previous studies on the Zhadang Glacier, the application of different turbulent flux schemes can result in large discrepancies in both the magnitude and even the sign of multi-year mean turbulent heat fluxes during winter and summer. Such differences directly affect the interpretation of the relative role of turbulent heat exchange in seasonal energy and mass balance.

This highlights the necessity of observation-based evaluation and calibration of turbulent flux parameterizations, particularly in data-sparse high-mountain environments.

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."

L80: I do not agree. Would winter and summertime fluxes not be different because the seasons are different?

Answer: We thank the reviewer for this comment and agree that winter and summer turbulent fluxes are expected to differ because they correspond to distinct seasonal conditions. However, this was not the comparison we intended to make, and we acknowledge that the original wording was ambiguous.

Our intention was to highlight that, for the same glacier and for the same season (either winter or summer), different turbulent flux parameterizations can produce substantially different multi-year mean flux estimates. In other words, the contrast discussed refers to model-dependent differences within a given season, rather than to the expected physical differences between winter and summer conditions.

We have revised the corresponding text to clarify this point and to explicitly state that the large discrepancies arise from differences among turbulent flux models applied to the same site and season. The revised sentence **(L86)** now reads : "As a result, turbulent flux estimates for the same glacier can differ substantially depending on the choice of turbulent flux parameterization (Radić et al., 2017). 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."

L104: capitalize Northeastern

Answer: We implemented the capitalization correction ("Northeastern") for consistency with standard geographic naming. The revised sentence **(L128)** now reads: "The Dunde Glacier is located in the western part of the Qilian Mountains, on the Northeastern TP."

L106: Is it gaining mass? I don't understand how one mass balance estimate can come from two separate studies.

Answer: Thank you for pointing out this confusion. We have revised the manuscript's descriptive text to clearly indicate the source of the cited data. The revised wording provides appropriate contextual information, enabling readers to correctly interpret the referenced values. The revised sentence **(L130)** now reads :"The ice cap covers an area of approximately 60 km², with an average annual elevation change rate of −0.501 ± 0.08 m yr^{-1} (Hugonnet et al., 2021)."

L111: It's more correct to say that fluxes were measured with a sonic anemometer and processed using the eddy covariance method.

Answer: We agree and revised the description to the more technically accurate phrasing **(L171)**: "Turbulent fluxes were measured with a CSAT3 three-dimensional sonic anemometer (Campbell IRGASON) and subsequently processed using the eddy covariance (EC) method during the period from 14 May to 12 October 2023;"

L113: EC is accurate, not particularly precise (see Figure 3).

Answer: We agree with this distinction and revised the text accordingly. The revised manuscript avoids implying high precision. The revised sentence **(L177)** now reads: "Turbulent fluxes were measured with a CSAT3 three-dimensional sonic anemometer (Campbell IRGASON) and subsequently processed using the eddy covariance (EC) method during the period from 14 May to 12 October 2023; this is a widely adopted technique in micrometeorological research that enables real-time, accurate, and continuous monitoring of atmospheric turbulence."

L120: Are you assuming the precipitation to be the same between on- and off-glacier sites separated by 350 m vertically? You will need to substantiate this. Further, ERA5 is has been shown to simulate

precipitation quite poorly in glaciated environments [e.g., supplemental figure S2 in Draeger et al., 2024]. How did you correct for this? How much data was missing?

Answer: Thank you—this is an important forcing issue. We expanded the Methods to explicitly describe. The revised sentence **(L156)** now reads:

"Precipitation was treated separately from other meteorological variables due to the lack of continuous in situ measurements at AWS1. Precipitation at 4970 m a.s.l. was reconstructed using ERA5 reanalysis data following a two-step procedure. First, precipitation amounts were corrected for gauge under-catch as a function of air temperature and wind speed during precipitation events, following Zhao et al. (2014). Second, a constant scaling factor of 1.16 was applied to the daily precipitation totals. To suppress unrealistically small precipitation signals, daily precipitation was set to 0 when ERA5 precipitation in the corresponding grid cell was lower than 0.45 mm, consistent with the criteria adopted in previous studies (Yang et al., 2013; Zhu et al., 2021). After reconstructing the precipitation at 4970 m a.s.l., precipitation at 5317 m a.s.l. was estimated by extrapolating the reconstructed precipitation using three constant precipitation gradients (ΔP1, ΔP2, and ΔP3), following Zhu et al. (2022)."

L125/Figure 1: Is the sonic anemometer oriented in the direction of prevailing winds?

Answer: Thank you. We added clarification in the figure description regarding the sonic anemometer orientation relative to prevailing winds, and we note how this consideration is addressed in quality control. The revised sentence **(L139)** now reads: "the three-dimensional sonic anemometer was oriented toward the downwind direction of the prevailing westerlies (i.e., facing east), such that the sonic measurement volume was exposed to undisturbed incoming flow and flow distortion from the instrument and mounting structure was minimized (Photo credit: Meilin Zhu)."

L130/Table 1: Incorrectly identified as table 2. Where are the specifications of the sonic anemometer? Temperature accuracy is incorrect. That is for the operation range above 20◦C. Pascal has units of Pa. Precipitation, what is FS ? Why do the heights have different numbers of decimals?

Answer: Thank you for the detailed table-quality checks. We corrected the table numbering error and revised Table 1 to include the sonic anemometer specifications. We also corrected the temperature accuracy statement to match the appropriate operational range, standardized pressure units to Pa, defined

"FS" (full scale) where used, and harmonized the number of decimal places reported for sensor heights for consistency and professional presentation.

Table 1: Characteristics of the sensors installed in the eddy covariance system to measure turbulent fluxes and meteorological variables in this study.

Variable	Symbol (unit)	Sensor	Accuracy	Range	Height
Air temperature	T_a (°C)	Vaisala HMP155A	20 to 60 °C:± (0.055 + 0.0057*T_a) °C −80 to 20 °C:±(0.226 − 0.0028*T_a) °C	−80 to 60 °C	2.17 m
Relative humidity	RH (%)	Vaisala HMP155A	± 2%	0%–100%	2.17 m
Wind speed	u (m s^{-1})	LICOR LI-7500DS	± 1.5%	0–100 m s^{-1}	2.05 m
Wind direction	WD (°)	LICOR LI-7500DS	± 2º	360 º	2.05 m
Air pressure	P (Pa)	Vaisala PTB210	± 0.5 hPa	50–1100 hPa	2.05 m
Incoming and outgoing longwave radiation	LWI, LWO (W m^{-2})	Campbell CNR4	± 1%	−250 to 250 W m^{-2}	1.60 m
Incoming and outgoing shortwave radiation	SWI, SWO (W m^{-2})	Campbell CNR4	± 1%	0–2000 W m^{-2}	1.60 m
Turbulent wind components	u′, v′, w′ (m s^{-1})	3-D sonic anemometer (Campbell IRGASON / CSAT3)	±1 mm s^{-1} (u′, v′) ±0.5 mm s^{-1} (w′)	0−65 m s^{-1}	2.05 m
Precipitation	(mm)	Geonor T-200B	± 0.1% Full Scale	0–600 mm	1.70 m

L133: No overview or context of the methods are provided, instead jumping into naming conventions

Answer: We agree. We added an overview paragraph at the start of the turbulent-flux Methods section to provide context before introducing naming conventions. The overview now summarizes the shared bulk-aerodynamic framework and clearly states the primary differences among the five schemes

(especially their stability treatment and how roughness lengths enter). The revised sentence **(L240)** now reads：“Variations in the performance of different turbulent flux parameterizations primarily arise from their respective approaches to computing turbulent exchange coefficients and applying atmospheric stability corrections. These methodological differences directly affect method accuracy under varying temperature gradients, humidity gradients, and wind speed conditions, leading to discrepancies in the simulated LE and H. In this study, we evaluate two categories of bulk aerodynamic methods for calculating turbulent fluxes over glacier surfaces. The first category comprises methods in which the bulk exchange coefficient (C) is formulated as a function of surface roughness lengths and atmospheric stability. The second category includes bulk methods that explicitly incorporate a simplified katabatic (glacier wind) parameterization.

Within this framework, we analyze five bulk methods that are commonly applied to glacier surfaces. These include: (i) a simplified bulk formulation without explicit stability corrections (C_{log}); (ii) and (iii) bulk methods that account for atmospheric stability through the bulk Richardson number (C_{Rib1} and C_{Rib2}); (iv) a method based on the full Monin–Obukhov similarity theory, employing universal stability functions with iterative closure (C_{M-O}); and (v) a bulk method derived from a simplified katabatic flow model that explicitly considers glacier wind effects (C_{kat}). The detailed formulations of methods (i) through (v) are described in the following subsections.”

L156: Constants established how?

Answer: We agree that this must be explicit. We revised the manuscript to distinguish among (i) universal physical constants, (ii) literature-prescribed parameters, and (iii) calibrated/tuned parameters. This is now documented in the new parameter table (values, units, sources) and supported by brief methodological justification in the text **(L347)**.

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

What are your modelled roughness lengths? How do they compare with other studies?

Answer: We thank the reviewer for raising this question. In the revised manuscript, we now explicitly report the modeled aerodynamic roughness length and clarify how it was determined. In this study, the aerodynamic roughness length z_{om} was derived using an inversion approach based on eddy-covariance observations, rather than being prescribed a priori. The resulting value is $z_{om} = 1.2 \times 10^{-4}$ m.

Although the inferred z0m lies toward the lower end of the roughness length range reported for glacier surfaces, it remains consistent with values documented in previous studies. For example, Guo et al. (2011) reported z0m values ranging from 10−4 to 10−2 m over Parlung No. 4 Glacier. Haven et al. (2025) observed roughness lengths spanning approximately 10−6 to 10−2 m over the Greenland Ice Sheet. Similarly, Radić et al. (2017) reported z0m values on the order of 10−5 to 10−2 m in the Cariboo Mountains, British Columbia, Canada. These ranges indicate that the optimized value obtained in this study falls within the variability reported for glacier surfaces under different climatic and surface conditions. Such small roughness lengths are physically plausible and consistent with the smooth ice or

snow surfaces typical of continental glaciers, particularly under cold and dry conditions. To ensure transparency and reproducibility, we added a dedicated Section 3.1 that provides a detailed description of the methodology used to derive z0m. In addition, we introduced Section 5.1 to compare roughness length characteristics across different studies and glacier types. (Guo et al., 2011; Haven et al., 2025; Radić et al., 2017)

The revised sentence **(L201-238、L564-608)** now reads：

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.

5.1 Glacier type dependence of aerodynamic roughness length variability

Aerodynamic roughness length (z_{0m}) over glacier surfaces spans several orders of magnitude, and the reported ranges vary substantially among different glacier types and climatic settings. Compared with previous studies, the range inferred in this study (10^{-5} – 10^{-2} m) f is consistent with values documented in the literature. For example, calculated values over August-one Glacier on the TP were approximately 10^{-5} to 10^{-2} m (Guo et al., 2018), while observations from the Greenland Ice Sheet (Haven et al., 2025) documented ranges of roughly 10^{-6} to 10^{-2} m. These comparisons indicate that the values derived here are physically reasonable and consistent with established glacier surface characteristics. Previous studies of continental glaciers on the TP have predominantly relied on SEB models to investigate glacier change and its underlying mechanisms. In most of these studies, roughness lengths, together with other meteorological parameters, were adjusted to improve the simulated cumulative mass balance produced

by the surface energy mass balance framework. In such cases, parameter selection primarily aimed to optimize overall SEB model performance rather than to explicitly constrain turbulent flux representation. Moreover, the lack of independent turbulent flux observations has limited the constraint on simulated sensible and latent heat fluxes, introducing substantial uncertainty into SEB based flux estimates. This limitation, in turn, hampers our understanding of continental glacier change and its underlying mechanisms. By contrast, observations of roughness length and turbulent fluxes provide a constrained optimization range for parameter calibration in SEB models. This allows the simulated turbulent fluxes to better reflect actual surface conditions, thereby improving our understanding of glacier change. Given this importance, it is therefore essential to characterize the observed variability and magnitude of roughness lengths themselves, particularly across different glacier types and climatic regimes.

Glacier type appears to systematically influence the characteristic magnitude and range of z_{0m}. Reported values from studies over the Greenland Ice Sheet collectively span approximately 10^{-6} – 10^{-2} m, although the specific intervals vary among studies depending on observational period and methodology (Meesters et al., 1997; Lenaerts et al., 2014; Van Tiggelen et al., 2020). For maritime glaciers on the TP, the limited available observations—primarily from Parlung No. 4 Glacier—indicate values between 10^{-4} and 10^{-2} m (Guo et al., 2011). In contrast, continental glaciers on the Tibetan Plateau often show lower characteristic magnitudes but broader reported ranges, spanning approximately 10^{-5} – 10^{-1} m, although most of these estimates are derived from SEB model calibration rather than independent turbulent flux observations (Giesen et al., 2008; Guo et al., 2018; Sun et al., 2018). Taken together, these comparisons indicate that continental glaciers tend to exhibit lower but more widely distributed roughness lengths than maritime glaciers, which are generally characterized by relatively higher and narrower ranges. The broad z_{0m} range derived from the observational data in this study further suggests that surface roughness conditions over continental glaciers can vary considerably over time. A plausible reason why such variability is more clearly captured in the derived values is the comparatively slow evolution of continental glacier surfaces, which allows changes in roughness length to be tracked more consistently in the observational record.

It should also be emphasized that most previous measurements were confined to short ablation season campaigns, whereas the present study spans a longer observation period from May to October. The wider range identified here may therefore partly result from the extended temporal coverage. Furthermore, the observation site is located in the accumulation zone, where melt intensity is relatively

weak and surface conditions are comparatively stable; such conditions may allow a broader range of roughness length variations to be captured in the observations. Nevertheless, given the scarcity of long-term eddy-covariance measurements over continental glaciers, additional observations are still required to evaluate the reliability and broader applicability of these findings.

L157: How do you establish γ? Is it ever 0?

Answer: We agree that this must be clearly defined. We have revised the Methods section to explicitly state the parameter values, their units, and their data sources. This revision eliminates ambiguity and ensures readers can replicate the parameter processing procedure. For details, please refer to **Table 2** (γ=0.005 K m^{-1}).

L168: How do you calculate Rib?

Answer: We agree and added an explicit definition of the bulk Richardson number Ri_b, including the variables used and the height(s) over which it is computed. We also ensured consistent notation across the Methods, equations, and variable definitions **(L309)**.

$$Ri_b = \frac{g\frac{\Delta\bar{T}}{\Delta Z}}{\bar{T}\left(\frac{\Delta\bar{u}}{\Delta Z}\right)^2}, \quad (15)$$

where, $\bar{u}$, $\bar{q}$, and $\bar{T}$ represent the mean wind speed, specific humidity, and air temperature, respectively; z is the measurement height, where $z_{u,t,q}$ are the log mean heights defined as：

$$z_{u,t,q} = \frac{z - z_{0u,t,q}}{ln\left(\frac{z}{z_{0u,t,q}}\right)}, \quad (16)$$

and $z_{0u,q,t}$ denotes the surface roughness lengths for momentum, humidity, and temperature.

L170: The stability functions are typically denoted by ϕ. Ø is the empty set.

Answer: We corrected the notation to use the conventional stability-function symbol ϕ, avoiding the empty-set symbol Ø **(L300)**.

The C_{Rib1} method retains the C_{log} formulation for turbulent flux calculation but introduces the bulk Richardson number (Rib) to allow for flux reduction under stable stratification (Suter et al., 2004). The bulk Richardson number can be related to the stability functions for momentum (ϕ_m), heat (ϕ_h), and water vapour (ϕ_q) in the following way ((Oke, 1987); see Eq. 14). In this formulation, the composite

function ($f_h(Ri_b)$) is constructed from the stability functions for momentum (ϕ_m), heat (ϕ_h), and water vapor (ϕ_q) (Dyer, 1974; Holtslag and Bruin, 1988).

$$LE = \rho L_{s/f} \kappa^2 z_u z_q \left(\frac{\Delta\bar{u}\Delta\bar{q}}{z^2}\right)(\phi_m\phi_h)^{-1}, \quad (12)$$

$$H = \rho C_P \kappa^2 z_u z_t \left(\frac{\Delta\bar{u}\Delta\bar{T}}{z^2}\right)(\phi_m\phi_q)^{-1}, \quad (13)$$

$$\begin{cases} (\phi_m\phi_{h,q})^{-1} = (1-5Ri_b)^2 \quad (Ri_b > 0) \\ (\phi_m\phi_{h,q})^{-1} = (1-16Ri_b)^{0.75} (Ri_b \le 0) \end{cases}, \quad (14)$$

L177: κ is the typical notation for the von K´arm´an constant.

Answer: We revised the manuscript to consistently use $\kappa$for the von Kármán constant throughout equations and text.

L181: The Andreas surface renewal model does not assume these are equal.

Answer: We agree and have corrected the description. The revised text **(L312)** now explicitly presents the formulation used to derive scalar roughness lengths from the momentum roughness, consistent with the referenced framework.

$$z_{u,t,q} = \frac{z - z_{0u,t,q}}{ln\left(\frac{z}{z_{0u,t,q}}\right)}, \quad (16)$$

and $z_{0u,q,t}$ denotes the surface roughness lengths for momentum, humidity, and temperature.

L193: 0.16 does not come out of nowhere here — it is κ^2

Answer: Thank you for this helpful clarification. We revised the text to explicitly state that 0.16 corresponds to $\kappa^2$with $\kappa = 0.4$, thereby avoiding the impression that this coefficient appears without justification. The revised sentence **(L321)** now reads：

Following Essery and Etchevers (2004) and Louis (1979), the exchange coefficient for surface sensible and latent heat flux is calculated as $C_H = C_{Hn} f_h$, where

$$C_{Hn} = \kappa^2 \left[\ln\left(\frac{z_1}{z_0}\right)\right]^{-2}, \quad (19)$$

is the neutral exchange coefficient for roughness length z_0

L195: This equation is incorrect. The c∼C∗ in the Louis model is an empirical constant – not CHn.

Answer: We thank the reviewer for this comment. We agree that in the original formulation of the Louis model (Louis, 1979), the parameter often denoted as c or C∗ is an empirical constant. However, the formulation used in our manuscript follows the implementation described by Essery and Etchevers (2004), rather than the original Louis (1979) form.

In the Essery and Etchevers (2004) framework, the surface exchange coefficient is written as $C_H = C_{Hn} f_h$, where C_{Hn} is the neutral exchange coefficient and the stability function f_h is expressed as a function of the bulk Richardson number. In this formulation, particularly under unstable conditions, f_h explicitly depends on C_{Hn}, with the empirical constant fixed to 10. Therefore, the appearance of C_{Hn} in the expression for f_h is consistent with the referenced implementation and does not imply that the empirical constant itself is replaced by C_{Hn} .

To avoid ambiguity, we revised the text to explicitly state that the C_{Rib2} method is implemented following Essery and Etchevers (2004) and Louis (1979), and we clarified the definitions of C_{Hn} and f_h accordingly.

The revised sentence **(L314)** now reads:

"3.2.4. C_{Rib2} method

Similar to C_{Rib1}, the C_{Rib2} method is a non-iterative method for calculating turbulent fluxes (Essery and Etchevers, 2004). It employs Ri_b and $f_h(Ri_b)$, thereby reducing computational complexity while preserving the physical consistency of MOST. Turbulent fluxes in the C_{Rib2} method are calculated as:

$$LE = \rho L_{s/f} C_H u[q_{sat}(T_s, P) - q], \tag{17}$$

$$H = \rho C_P C_H u[T_s - T_a], \tag{18}$$

where, $q_{sat}(T_s, P)$ denotes the saturation specific humidity at surface temperature T_s and pressure P, and C_H is a surface exchange coefficient. Following Essery and Etchevers (2004) and Louis (1979), the exchange coefficient for surface sensible and latent heat flux is calculated as $C_H = C_{Hn} f_h$, where

$$C_{Hn} = \kappa^2 \left[\ln\left(\frac{z}{z_{0m}}\right)\right]^{-2}, \tag{19}$$

is the neutral exchange coefficient for roughness length z_{0m} and

$$f_h = \begin{cases} (1 + 10Ri_b)^{-1} & Ri_b \geq 0 \\ 1 - 10Ri_b\left(1 + 10Ri_b C_{Hn}\sqrt{-Ri_b}/f_z\right)^{-1} & Ri_b < 0 \end{cases}, \tag{20}$$

with

$$f_z = \frac{1}{4}\left(\frac{z_{0m}}{z}\right)^{1/2}, \tag{21}$$

Although atmospheric stability is also characterized by Ri_b, the calculation approach in C_{Rib2} method differs from that in C_{Rib1} method as follows:

$$Ri_b = \frac{gz_1}{u^2}\left\{\frac{T_a - T_s}{T_a} + \frac{q - q_{sat}(T_s,P)}{q+\epsilon/(1-\epsilon)}\right\}, \quad (22)$$

where, ϵ is defined as the ratio of the molecular weights of water to dry air, with a value of 0.622."

L200: The ratio of molecular weights is expressed by ϵ. $\in$denotes set inclusion.

Answer: We corrected the notation by expressing the molecular-weight ratio as $\epsilon$and removing the inappropriate set-inclusion symbol $\in$. The revised equations and definitions now use standard thermodynamic notation. The revised sentence **(L328)** now reads:

Although atmospheric stability is also characterized by Ri_b, the calculation approach in C_{Rib2} method differs from that in C_{Rib1} method as follows:

$$Ri_b = \frac{gz_1}{u^2}\left\{\frac{T_a - T_s}{T_a} + \frac{q - q_{sat}(T_s,P_s)}{q+\epsilon/(1-\epsilon)}\right\}, \quad (22)$$

where, ϵ is defined as the ratio of the molecular weights of water to dry air, with a value of 0.622.

L211: RMSE, MAD, and MBE are standard measures and don't need to be defined.

Answer: We agree and streamlined this part of the manuscript. Definitions of standard error metrics (RMSE, MAD, MBE) were removed or condensed to a minimal statement, improving concision and allowing more space for methodologically important details.

L222: This information was already provided.

Answer: We agree and removed the redundant information to improve the narrative flow and avoid repetition.

L224: In table 1, you say that pressure is P, not Pres.

Answer: We standardized pressure notation as Pconsistently across Table 1, the Methods text, and equations to eliminate inconsistency.

L246: This is fine as a sanity check, but is otherwise a given.

Answer: We agree and revised the text to treat this point as a brief sanity check rather than a substantive result. The revised wording is concise and keeps the focus on the main findings. The revised sentence **(L381)** now reads: "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)."

L255: How do you define the range? 1.5 between min and max? That isn't much, if so.

Answer: We thank the reviewer for pointing out this ambiguity. The value of approximately 1.5 m s^{-1} refers to the peak-to-trough amplitude of the mean diurnal wind speed cycle, rather than to the full range between minimum and maximum wind speeds. We agree that using the term "range" in this context may have been misleading.

We have revised the text to explicitly state that this value represents the diurnal amplitude derived from the mean daily cycle (Fig. 2l), while the overall wind speed range is defined separately by the observed minimum and maximum daily mean values (1.7–10.5 m s^{-1}). This clarification ensures that the different measures of variability are clearly distinguished and avoids confusion in interpretation. The revised sentence **(L387)** now reads: "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)."

L257: What do you mean by stable here?

Answer: We thank the reviewer for this question. In this context, the term "stable" is used in a qualitative sense to describe nighttime atmospheric stratification, rather than to indicate a specific stability regime defined by a quantitative metric such as the bulk Richardson number or Monin–Obukhov length.

Our intention was to convey that the reduced nighttime wind-speed variability is consistent with typically stable stratification over glacier surfaces at night, when radiative cooling suppresses turbulent mixing. To avoid ambiguity, we revised the text to explicitly refer to "stable atmospheric stratification" and to frame this statement as a qualitative interpretation, rather than a formal stability classification.

The revised sentence **(L389)** now reads： "Nighttime wind speeds exhibited limited variability, consistent with stable atmospheric stratification."

L259: How often was the wind aligned with the direction of the glacier slope? Or was the wind more often aligned with the westerlies? Were there katabatic winds? Channelization of flows from above?

Answer: We thank the reviewer for this question. The observation site is located on a relatively elevated and gently sloping section of the glacier surface, with surrounding terrain that is generally lower and does not form a clear upslope catchment. As a result, the site is not situated along a well-defined downslope flow pathway, and conditions are therefore less favorable for the development of persistent katabatic winds or strongly channelized flows originating from higher elevations.

Wind-direction statistics indicate that winds were more frequently aligned with the prevailing westerly and northwesterly directions, consistent with the influence of the mid-latitude westerlies, rather than being controlled by the local glacier slope orientation. Although katabatic winds may occur intermittently, they do not appear to represent a dominant or persistent feature of the wind record at this site.

We have clarified this interpretation in the revised manuscript, emphasizing that the wind regime at the observation site is primarily influenced by regional-scale circulation rather than by local downslope or channelized flow processes. The revised sentence **(L390)** now reads：: "Prevailing wind directions were westerly and northwesterly (Fig. 2m and o), highlighting the dominant influence of the mid-latitude westerlies on regional meteorological conditions."

L266: I don't think I would describe variations on this scale as a marked influence. Back-of-the-napkin math suggests density variations of less than 5%.

Answer: We agree and revised the language to avoid overstatement. The revised manuscript now describes density variability as a secondary influence. The revised sentence **(L396)** now reads: "Overall, atmospheric pressure variations were relatively small over the study period and therefore likely played only a minor role in modulating turbulent fluxes."

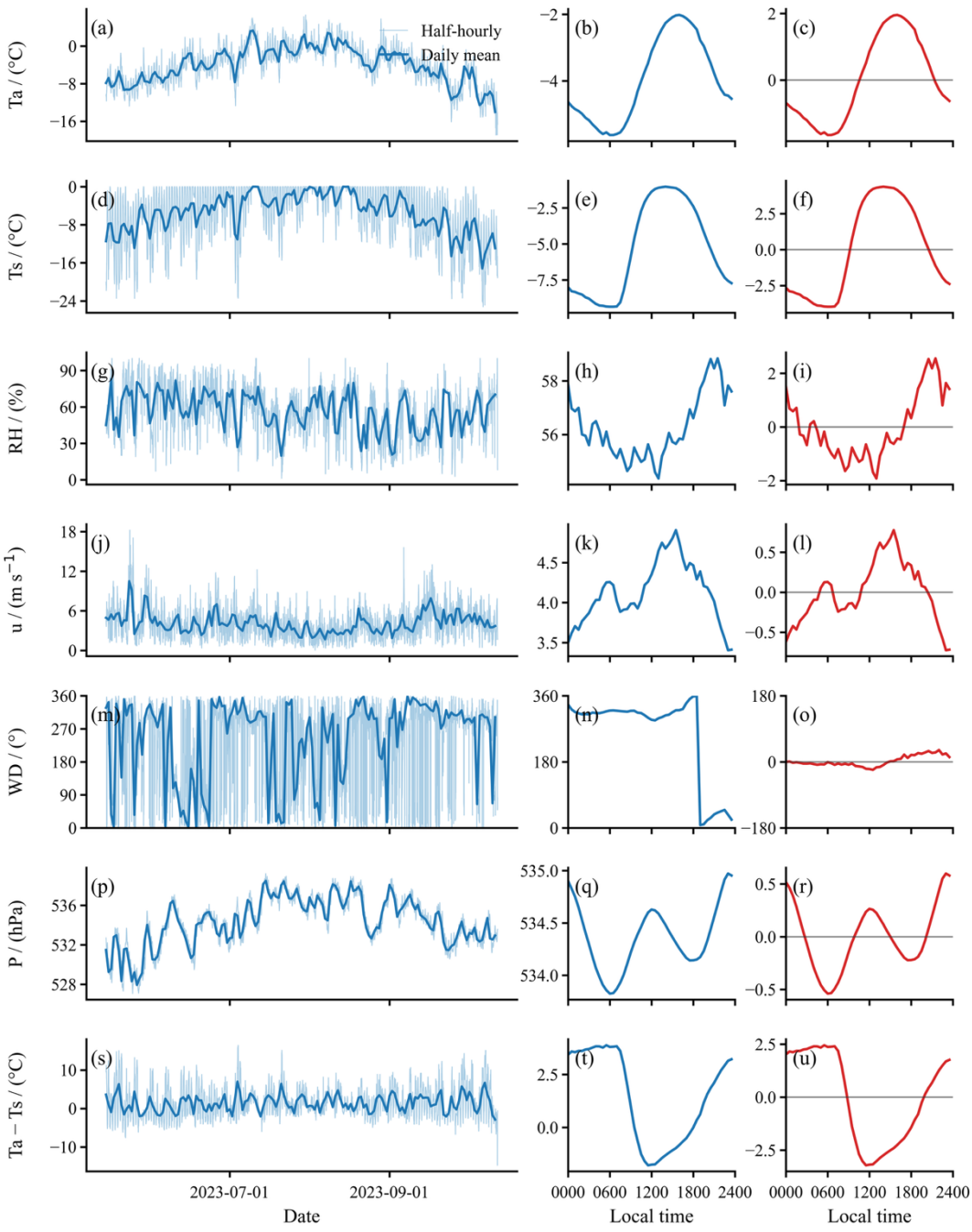
L271/Figure 2: How are you calculating the daily statistics? It appears, especially in (k), that the daily averages lead the hourly signals? By visual inspection, it appears that the average WD is calculated directly. WD is a circular variable so averages of cos(WD) and sin(WD) need to be calculated separately. I wonder if demeaned daily cycles would be even more informative?

Answer: We thank the reviewer for this detailed and constructive comment. The apparent lead of the daily averages relative to the half-hourly signals (e.g., Fig. 2k) is a plotting artifact rather than a calculation issue. Specifically, daily mean values were plotted at the start of each day (00:00 local time), which visually shifts the daily curve slightly earlier than the corresponding half-hourly time series.

Regarding wind direction, we agree that it is a circular variable and that direct arithmetic averaging is inappropriate. Following the reviewer's suggestion, we revised the calculation of daily mean wind direction using circular statistics, by separately averaging the sine and cosine components of wind direction before recomputing the mean angle. The revised approach is now consistently applied and clearly documented.

In addition, motivated by the reviewer's comment, we further analyzed demeaned diurnal cycles for all meteorological variables to better isolate intraday variability from longer-term changes. These demeaned daily cycles are now presented and discussed in Fig. 2 and Section 4.1, where they provide clearer insight into the characteristic diurnal patterns and their physical interpretation.

Overall, these revisions improve the statistical rigor of the daily analyses and enhance the interpretability of the diurnal signals shown in the manuscript.



L277: This should be explained in your methods.

Answer: We agree and moved the relevant methodological details to the Methods section. The revised manuscript now explicitly documents the procedure that was previously only implied, ensuring reproducibility.

L298/Figure 3: How much do you trust these measurements? Is the sensible heat flux really going from −150 Wm−2 to−20 Wm−2 in the span of an hour? See general comment on flux processing. Figure caption spills onto next page.

Answer: Thank you—this is a key concern. We strengthened the EC processing description (rotation, time lag, WPL correction, frequency-response considerations, and QC filtering) to support the reliability and reproducibility of the measured fluxes. The figure caption formatting issue was corrected to prevent spillover.

L316/Figure 4: Although time series can be nice to see general trends, they're not informative when making comparisons. I cannot tell which of these models performed best through this figure. A scatter plot would be more helpful. Why are the so few missing measurements here compared to Figure 3?

Answer: We thank the reviewer for this constructive comment. We agree that time series are primarily useful for illustrating general temporal evolution, but are limited for model comparison. In response, we have added scatter plots comparing modeled and observed turbulent fluxes and provided a performance analysis in Section 4.3.1 of the revised manuscript. These scatter plots allow a clearer assessment of relative model performance and complement the time-series presentation.

Regarding the number of missing measurements, the difference between Fig. 3 and Fig. 4 in the original version resulted from inconsistent data-screening criteria. In Fig. 3, daily values were excluded when more than half of the corresponding half-hourly measurements within a day were missing. This criterion was not consistently applied in Fig. 4, which led to an apparent discrepancy in data availability between the two figures. In the revised manuscript, 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).

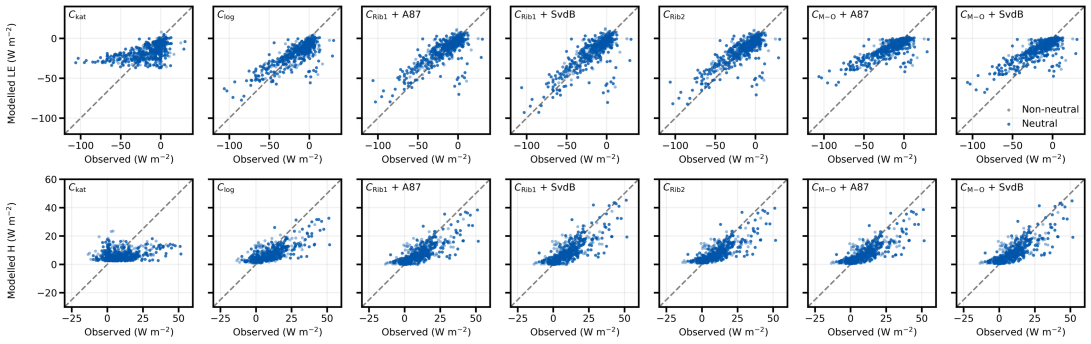


Figure 4: Comparison of 30-min observed (eddy-covariance-derived) and modeled H and LE according to the five turbulent flux methods.

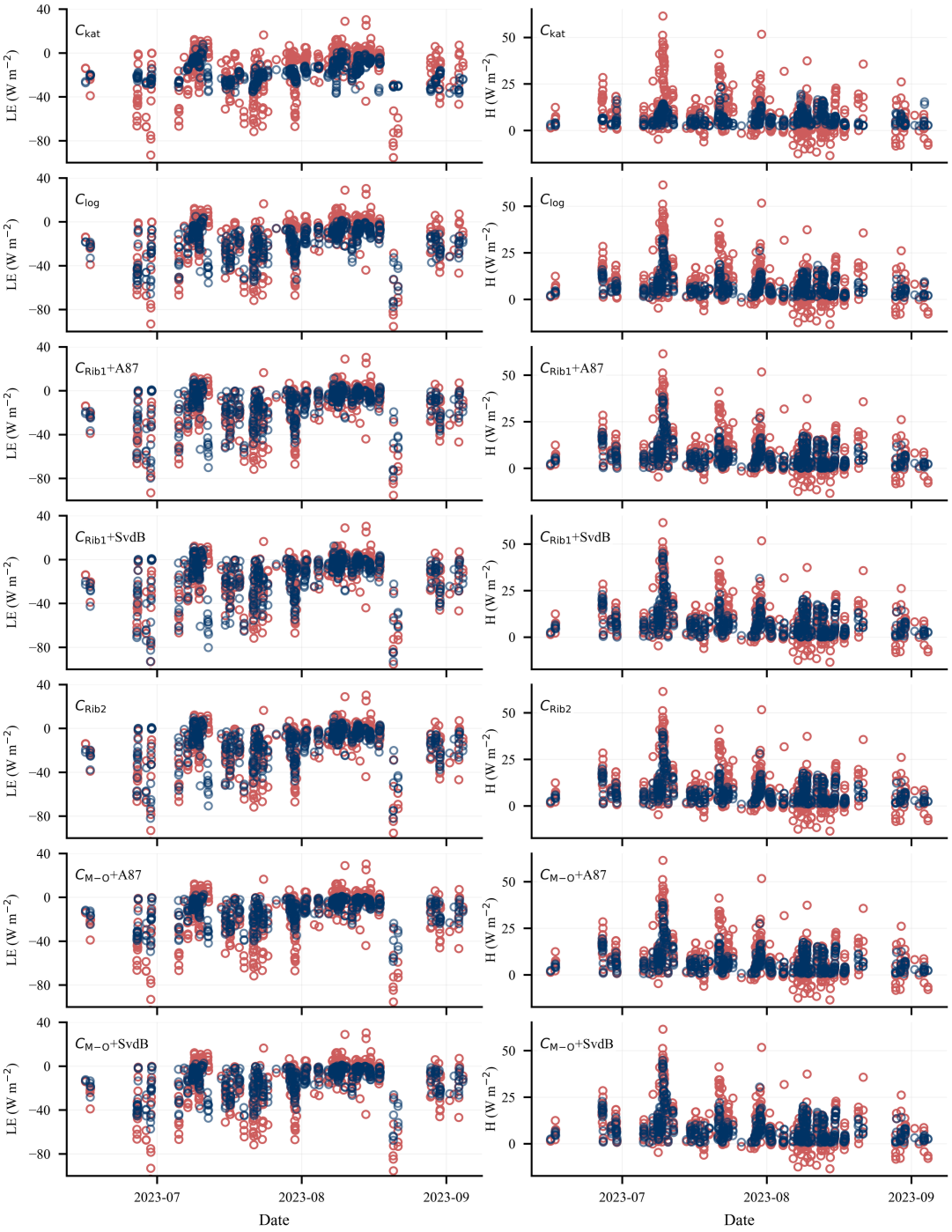


Figure 5: Time series of 30-min LE and H simulated by the five different schemes (C_{kat}, C_{log}, C_{Rib1}, C_{Rib2}, and C_{M-O}), compared with eddy covariance–derived values. The C_{Rib1}, C_{Rib2}, and C_{M-O} schemes were calculated by the observed aerodynamic roughness length (z_{0m} = 1.2 × 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.

Section 3.1 **(L228)**: "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."

L335: Across which meteorological regimes?

Answer: We thank the reviewer for pointing out that the term "meteorological regimes" was too broadly stated. In this context, the regimes refer specifically to seasonally distinct atmospheric conditions, reflecting different combinations of wind speed, atmospheric stability, temperature and humidity gradients, and glacier surface melt conditions. However, in the revised manuscript, the observed turbulent flux data used for model evaluation were reprocessed and further filtered, resulting in a shorter time span with relatively similar atmospheric conditions. Accordingly, the original Section "**4.3.2 Seasonal variability in method performance"** has been removed.

L370: I'm not sure I agree that this is a good alignment. The curves look similar qualitatively, but the relative (%) differences between measured and modelled are substantial.

Answer: We thank the reviewer for this comment and agree that visual similarity in the diurnal curves does not imply small relative differences in magnitude. Our intention in referring to "alignment" was to describe the consistency in the timing and overall shape of the diurnal cycles, rather than to suggest close quantitative agreement in absolute or relative terms.

In the revised manuscript, we recalculated the turbulent fluxes using an observation-based aerodynamic roughness length obtained through inversion. Compared with the original configuration, this substantially reduces the systematic bias between modeled and observed fluxes and improves the agreement in diurnal amplitude. In addition, to ensure statistical robustness, we processed the turbulent

flux data so that time points with few and unrepresentative observations were excluded from the diurnal analysis. This screening procedure ensured that the derived diurnal cycles were based on sufficiently representative observations.

The revised sentence **(L547)** now reads:

"**4.3.4 Evaluation of daily variations**

An evaluation of daily variations in modeled turbulent fluxes offered important diagnostic information on the simulation ability of a given parameterization. To ensure statistical robustness, days with less than 50% missing half-hourly observations (i.e., days with more than 50% valid data coverage) were first defined as effective observation days. Based on these effective days, individual time-of-day bins were then included in the diurnal averaging only when their occurrence frequency exceeded 50% of the number of effective observation days. This two-step screening ensured that the derived diurnal cycles were constructed from sufficiently representative and consistently sampled observations. This screening procedure ensured that the derived diurnal cycles were based on sufficiently representative observations.

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."

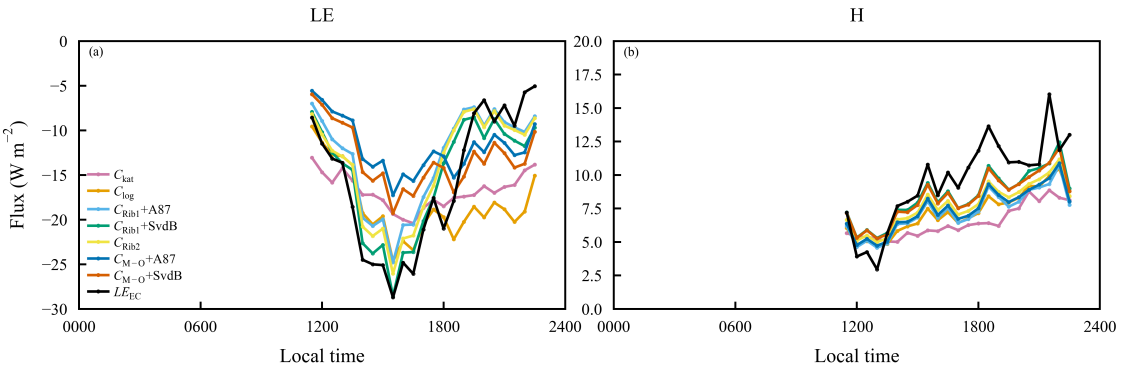


Figure 6: Mean daily cycles of modeled latent heat and sensible heat fluxes obtained from the five different schemes (C_{kat}, C_{log}, C_{Rib1}, C_{Rib2}, and C_{M-O}) compared with eddy covariance-derived values (LE_{EC} and H_{EC}).

L378: Do you mean stable in the sense of atmospheric stability?

Answer: We thank the reviewer for the clarification. No, in this context "stable" does not refer to atmospheric stability in a formal sense (e.g., as defined by the bulk Richardson number or Monin–Obukhov length). It was intended purely as a descriptive term indicating that the turbulent fluxes exhibit relatively small variability during this time period.

To avoid ambiguity, we revised the text to replace "stable" with wording that explicitly refers to reduced temporal variability or weak fluctuations in the fluxes, without implying any specific atmospheric stability regime.

L387: What do you mean by stable here? Furthermore, I do not agree that this is particularly accurate.

To avoid ambiguity, we revised the text to replace "stable" with wording that explicitly refers to reduced temporal variability or weak fluctuations in the fluxes, without implying any specific atmospheric stability regime. We have also revised the description of the results to avoid exaggeration and absolute claims.

L394: Almost all of this section needs to be reworked, and included in your methods section, see my general point. This is not a discussion of your results.

Answer: We thank the reviewer for this important comment regarding the manuscript structure. We agree that, in the original version, the separation between Methods, Results, and Discussion was not sufficiently clear, which may have obscured the distinct roles of each section.

Following this comment, we carefully reorganized the manuscript. The former Section 5.1 has now been fully relocated to the Methods section. This restructuring improves the logical flow of the Methods and ensures that all methodological descriptions are presented in a single, coherent framework.

In the revised manuscript, we introduced a new discussion on the characteristics of roughness length across different glacier types (Section 5.1). Together with the original Sections 5.2 and 5.3, these sections now collectively constitute the Discussion part of the manuscript. We would like to clarify that these sections are deliberately framed as Discussion rather than Results. Although they are directly based on the quantitative findings presented in Section 4, their primary purpose is not to introduce new results, but to elaborate on the broader implications and applications of those findings. Specifically, these sections aim to provide guidance for roughness length selection in similar studies, and to clarify how different turbulent flux parameterization schemes influence glacier energy and mass balance simulations, as well as to evaluate their applicability under varying climatic conditions. Section 5.1 highlights the current scarcity of in situ roughness length measurements for continental glaciers on the Tibetan Plateau, and emphasizes that roughness lengths derived from other glacier types cannot be directly transferred to this setting. By comparing roughness length characteristics across different glacier types, we provide a reasonable range for roughness length selection for glaciers similar to that investigated in this study. Section 5.2 focuses on how improvements in turbulent flux representation translate into enhanced performance of SEB and mass-balance models, thereby placing the results within a modeling and process-oriented framework that extends beyond simple statistical comparison. Section 5.3 further broadens this perspective by examining the behavior and limitations of the turbulent flux schemes under extreme weather conditions, highlighting their broader implications for glacier response under ongoing climate warming. Together, these discussions emphasize the applied value of the results rather than presenting new quantitative findings.

The revised sentence now reads:

3 Methods

3.1 The algorithm and quality control for roughness lengths

3.2 Bulk methods

3.3 Energy Balance Model

5 Discussion

5.1 Glacier type dependence of aerodynamic roughness length variability

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

5.3 Evaluation of turbulent flux methods under extreme weather conditions

L406: What is the relevance of the Greenland Ice Sheet?

Answer: We agree that the Greenland reference was not sufficiently motivated. We either removed the reference where it did not directly support the point being made, or reframed it explicitly to support a specific methodological/physical argument (rather than serving as a loose comparison). This improves focus and relevance.

L408: CH isn't necessarily a constant. It depends on the roughness length, which you said you weren't fixing on L139.

Answer: We thank the reviewer for this clarification. We agree that, in general, the surface exchange coefficient CH is not a fundamental constant and depends on surface roughness and atmospheric conditions through its formulation.

In this study, however, CH is treated following the original framework of Oerlemans (2000), in which CH is introduced as an effective, tunable bulk transfer coefficient rather than being explicitly recalculated from roughness lengths. Within that formulation, CH is adjusted to represent the combined effects of surface roughness and turbulent exchange efficiency under the assumptions of the bulk approach, rather than being interpreted as a universal or physically constant parameter.

We have revised the text to clarify this distinction and to avoid wording that could be interpreted as implying that CHis independent of surface roughness. The revised description now explicitly states that CH is used as a parameterized exchange coefficient following the referenced formulation, consistent with the treatment in Oerlemans (2000).

The revised sentence now reads **(L283)**:

"3.2.2. C_{log} method

The C_{log} method represents a highly simplified derivative of MOST. This method employs a constant exchange coefficient (C_h) driven by near-surface wind speed and the difference in air temperature (or humidity) and surface temperature (or humidity), providing a simplified, computationally efficient structure suitable for large-scale climate models (Oerlemans, 2000). It does not dynamically adjust atmospheric stability, and instead retains only a linear relationship for surface fluxes. Being a wind-speed-driven scheme, it performs well under less stable stratification conditions. However, the method also exhibits certain limitations. Owing to the absence of atmospheric stability parameters in its structure, it cannot identify or respond to surface inversions forming at night or in the early morning. LE and H are expressed as:

$$LE = 0.622\rho L_{s/f} C_h u(e_a - e_s)/P, \quad (10)$$

$$H = \rho C_P C_h u(T_a - T_s), \quad (11)$$

where, u is wind speed ($m \cdot s^{-1}$), and C_h is the turbulent exchange coefficient. Following (Oerlemans, 2000), turbulent fluxes are regulated through the exchange coefficient C_h. Here, C_h is treated as a calibration parameter and optimized by minimizing the mismatch between modeled and observed turbulent fluxes."

L420: The iteration on L in some models is not done to improve responsiveness. L is defined in terms of fluxes, but the fluxes are modelled in terms of L. This circular dependence necessitates an iterative scheme.

Answer: We agree and corrected the explanation. The iteration on $L$is required because $L$is defined in terms of turbulent fluxes, while the fluxes themselves depend on stability functions that depend on L. We revised the text to clearly describe this circular dependence and removed any implication that iteration is done to improve "responsiveness."

L423: I don't understand why albedo is mentioned here.

Answer: Thank you for this comment. We agree that the reference to albedo was not directly related to the point being made in the main text. We have therefore removed the unrelated references and revised the section to focus specifically on the argument under discussion.

L425: The CM−O method did not seem to perform best in all cases.

Answer: We thank the reviewer for this important clarification. We agree that the CM–O method does not outperform all other schemes under all conditions. Accordingly, in the revised manuscript we have carefully adjusted the wording to avoid overly absolute statements regarding its performance.

The revised sentence now reads:

L546: 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.

L744: 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.

L427: Most of the schemes used should have different roughness lengths. The C parameters are not constants.

Answer: We thank the reviewer for this important clarification. We agree that, in general, different turbulent flux schemes involve different roughness lengths and that the exchange coefficients (C parameters) are not fundamental constants. We have revised the text to clarify this distinction and to avoid any wording that could be interpreted as implying that roughness lengths (z0m) or exchange coefficients are assumed to be constant across schemes. The revised description now explicitly emphasizes the scheme-dependent treatment of roughness lengths and the diagnostic nature of the exchange coefficients.

L429: The iterative algorithm does not ensure energy balance closure.

Answer: We thank the reviewer for this clarification. We agree that an iterative algorithm within the MOST framework does not, by itself, ensure closure of the surface energy balance. Its role is to achieve internal consistency in the calculation of friction velocity, stability functions, and turbulent exchange coefficients, rather than to enforce energy balance closure.

To avoid this incorrect implication, we have removed the statement suggesting that the iterative algorithm ensures energy balance closure. The revised text now focuses solely on the physically consistent and diagnostic aspects of the CM–O implementation within MOST.

L430: How was this ranking done? For example, the Ckat model had lower |MBE|.

Answer: We thank the reviewer for this important question.

In the revised manuscript, the relative performance of the turbulent flux schemes is assessed under clearly defined conditions. When all methods are forced to use the same observation-based aerodynamic roughness length, the comparison focuses on schemes that explicitly depend on prescribed roughness lengths (CRib1, CRib2, and CM–O). Within this consistent framework, the three approaches based on the observed z0m (i.e., CRib1, CRib2, and CM–O) show overall comparable performance in simulating LE and H throughout the study period. When focusing on the diurnal cycle, however, CRib1 and CRib2

demonstrate better agreement with observations in reproducing the evolution of the LE diurnal pattern, whereas differences among the schemes are relatively small for H. Therefore, considering both the overall period and the diurnal variability, although the five schemes do not differ substantially, the method combining Richardson-number-based stability correction with the SvdB scalar roughness parameterization exhibits a relatively balanced performance in simulating LE and H and more accurately captures the diurnal evolution of turbulent fluxes. Such an approach can thus be regarded as a preferable option for simulating turbulent fluxes over the Dunde Glacier.

We also note, as pointed out by the reviewer, that in terms of individual metrics such as |MBE|, other schemes—particularly Ckat—may show smaller absolute bias. However, Ckat and Clog involve parameter optimization and therefore are not directly comparable to methods that rely on prescribed roughness lengths. When all schemes are allowed to optimize parameters, the performance differences among methods are reduced, and the relative advantage of the CM–O method becomes less pronounced. This is now explicitly stated in the revised text, and overly absolute wording has been removed.

Overall, the revised manuscript no longer presents a single universal ranking, but instead emphasizes conditional and framework-dependent performance differences among methods.

The revised sentence now reads:

L546: "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."

L433: I don't think interseasonal RMSE difference is a very useful statistic. If the RMSE was 100 in all seasons, it would be a terrible model, but the interseasonal RMSE difference would be 0.

Answer: We thank the reviewer for this important clarification and agree that the interseasonal RMSE difference is not a measure of absolute model performance. As correctly pointed out, a model could exhibit equally large RMSE values in all seasons and still yield an interseasonal RMSE difference of zero.

In the revised manuscript, we therefore clarify that the interseasonal RMSE difference is used solely as a supplementary metric to describe the *consistency of model performance across seasons*, rather than as an indicator of overall model accuracy. The primary evaluation of model performance is based on absolute statistics such as RMSE, mean bias, and scatter plots, which are discussed throughout Section 4.3.

To avoid misinterpretation, we have revised the text to avoid implying that a small interseasonal RMSE difference alone indicates good model performance.

L441: I do not understand what "recalibrated parameters" means in the context of this section. If this relates to EBFM, this needs to be explained in the methods. Because of this, I don't understand how Ts is being modelled.

Answer: We thank the reviewer for pointing out this ambiguity. The term "recalibrated parameters" indeed refers to parameters within the energy-balance and mass-balance framework (EBFM), and we agree that this needs to be clearly explained in the Methods.

In the revised manuscript, we have substantially expanded the description of the EBFM in Section 3.3, where all calibrated parameters and their roles in the surface energy balance are explicitly defined. In addition, Section 5.2 now clearly specifies the sources of all input variables used to drive the EBFM, including how surface temperature T_s is determined within the model framework.

These revisions are intended to ensure that the only varying component influencing the EBFM simulations in the sensitivity experiments is the choice of turbulent flux parameterization, while all other inputs and model parameters are treated consistently. This clarification resolves the ambiguity regarding the meaning of "recalibrated parameters" and provides a transparent basis for interpreting the modeled surface temperature and mass-balance responses.

The revised sentence **(L349 and L609)** now reads:

L349: "3.3 Energy Balance Model

The approximate SEB (in W m^{-2}) can be written as:

$$Q_m = S_{\text{in}} + S_{\text{out}} + L_{\text{in}} + L_{\text{out}} + H + LE + Q_{\text{rain}} + Q_{\text{sub}} \ , \quad (25)$$

where Q_m is the energy available for melt, S_{in} represents the incoming shortwave radiance, S_{out} is the reflected shortwave radiation, and L_{in} and L_{out} refer to the incoming and outgoing longwave radiation, respectively. Q_{rain} is the heat transfer due to rainfall, while Q_{sub} denotes the heat flux into the ice. In the model, the energy fluxes are expressed in a way that the surface temperature is the only unknown variable, which is determined by iteratively solving Eq. (25) with the left-hand-side set to zero. If the computed surface temperature exceeds the melting point, it is constrained to the melting point, and the energy fluxes are recalculated. In this scenario, the sum of fluxes becomes positive, and melting will occur (EBFM; Van Pelt et al., 2012, 2019). The mass balance (MB, mm w.e.), is calculated as follows:

$$MB = \int \left(\frac{Q_m}{L_m} + P_{snow} + \frac{Q_L}{L_s} \right) dt \ , \quad (26)$$

P_{snow} is the snow accumulation, Q_L is the mass exchange due to sublimation. L_m is the latent heat of melting (~ 3.34×10^5 J kg^{-1}). Since this study is based on point data and does not involve spatial data, topographical parameters were not set in EBFM during the analysis."

L609: **"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."

L451/Figure 6: Units missing.

L455/Table 4: Units missing.

Answer: We corrected Figure by adding the missing units to all axes and ensuring that variable names and units are consistent with those used in the main text and parameter table.

L457: Are the correlations correct? Comparing the measured to modelled surface height visually, correlations of r>0.9 is surprising.

Answer: Thank you for raising this concern. We verified the correlation calculations. High correlation can occur when the model captures the dominant temporal trend (e.g., seasonal evolution), even if magnitude biases remain.