

Dear Editor and Reviewer,

We sincerely thank the Editor and the Reviewer for their careful assessment of our revised manuscript and for providing us with another opportunity to improve the manuscript. The remaining comments have been carefully considered and addressed in the revised manuscript.

A detailed point-by-point response is provided below. The Reviewer's comments are reproduced first, followed by our answers and the corresponding revisions made in the manuscript.

Specific comments:

Calibration of turbulent flux models:

The new strategy to calibrate roughness lengths or similar in the turbulent flux models against observations is an improvement. It is clear how a best value for z_{0m} was found, but it is not described how the constants in the C_{kat} and C_{log} methods were calibrated. What range of values was tested? And what measure (RMSE, bias, R) was used to select a best value? Please add this information in Sect. 3.2.

Answer: Thank you for identifying this omission. We agree that the previous version did not provide sufficient information on the calibration of the empirical parameters used in the C_{kat} and C_{log} methods. We have therefore expanded Sect. 3.2 to describe the calibrated parameters, parameter-search procedure, selection criterion, and resulting optimal values.

For the C_{kat} method, the calibrated parameters were C_{turb} and C_{turb2} , whereas C_h was calibrated for the C_{log} method. The calibration followed a coarse-to-fine parameter-search procedure. Broad initial ranges were defined according to the orders of

magnitude reported in the original publications, and the search intervals were then progressively narrowed around parameter combinations associated with lower errors.

For details, please refer to Sect. 3.2, lines 267–282 of the revised manuscript:

Model performance was 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 were 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).

For the two methods in which parameter calibration does not involve roughness length, namely C_{kat} and C_{log} , the empirical parameters were recalibrated against the quality-controlled 30-min EC observations. Specifically, C_{turb} and C_{turb2} were calibrated for the C_{kat} method, whereas C_h was calibrated for the C_{log} method (see Sects. 3.2.1 and 3.2.2 for details). The calibration followed a coarse-to-fine parameter-search procedure. First, the approximate orders of magnitude of the parameters were determined from values reported in previous studies, and relatively broad parameter intervals around these values were explored. The parameter space was then progressively narrowed around combinations that produced smaller discrepancies between the modeled and observed fluxes. The objective function was defined as the sum of the root mean square errors (RMSEs) of the modeled latent and sensible heat fluxes relative to the corresponding EC observations:

$$J(\theta) = \text{RMSE}[LE_{\text{mod}}(\theta), LE_{\text{obs}}] + \text{RMSE}[H_{\text{mod}}(\theta), H_{\text{obs}}] \quad (7)$$

where $\theta = (C_{\text{turb}}, C_{\text{turb2}})$ for the C_{kat} method and $\theta = C_h$ for the C_{log} method. The search intervals were iteratively refined until the parameter combination yielding the minimum value of J was identified. The resulting optimal values were $C_{\text{turb}} = 0.0001$, $C_{\text{turb2}} = 0.007$, and $C_h = 0.0005$. RMSE was therefore used as the criterion for parameter selection, whereas MBE and MAD were used only for the subsequent evaluation of model performance.

Results with and without calibration:

It could be made more clear at the start of Sect. 4.3 that all results in that section are based on the calibrated turbulent fluxes. Furthermore, it is unclear what Table 3 is showing. It is not cited in the text and it contains different values than the similar Table 4. Maybe Table 3 simply needs to be removed (?).

Answer: We agree with the Reviewer. We have revised the opening paragraph of Sect. 4.3 to explicitly state that the results presented in this section are based on recalibrated or otherwise observation-constrained turbulent-flux configurations.

The revised text clarifies that the C_{kat} and C_{log} methods use empirical parameters recalibrated against the quality-controlled 30-min EC observations. The C_{Rib1} , C_{Rib2} , and $C_{\text{M-O}}$ methods use the representative momentum roughness length derived from the EC observations, $z_{0m} = 1.2 \times 10^{-4}$ m. Where required, the scalar roughness lengths are determined using the A87 and SvdB parameterizations.

We also agree that the former Table 3 was unclear and overlapped with the information provided in Table 4. Table 3 has therefore been removed. The qualitative result concerning the underestimation of sensible heat flux under near-neutral conditions has been retained in a shortened form in the text.

For details, please refer to the opening of Sect. 4.3, lines 471–500 of the revised manuscript:

4.3 Evaluation of different turbulent flux parameterizations

In this section, we evaluate the performance of five turbulent-flux methods (C_{kat} , C_{log} , C_{Rib1} , C_{Rib2} , and $C_{\text{M-O}}$) across different timescales using observation-constrained model configurations. Specifically, the C_{kat} and C_{log} methods were applied using empirical parameters recalibrated against the quality-controlled 30-min EC observations. In contrast, the C_{Rib1} , C_{Rib2} , and $C_{\text{M-O}}$ methods were applied using the representative momentum roughness length ($z_{0m} = 1.2 \times 10^{-4}$ m) derived from the EC observations. Where required, the scalar roughness lengths were determined using the A87 and SvdB parameterizations. Unless otherwise stated, all results presented in this section are based on these recalibrated or

observation-constrained model configurations. To ensure that the assessment was based on high-quality data, only the quality-controlled 30-min EC turbulent-flux observations were used for validation.

4.3.1 Model–observation correspondence

The scatter plots revealed differences among the evaluated turbulent-flux configurations in simulating turbulent heat fluxes over the Dundee Glacier (Fig. 4). Overall, all methods exhibited relatively high accuracy for both latent and sensible heat fluxes when the absolute flux magnitudes were small, with the modeled values showing good agreement with the observations. However, as the absolute magnitude of the fluxes increased, the simulation errors tended to increase, accompanied by a gradual decline in model performance. This behavior indicated that the evaluated parameterization schemes had greater difficulty reproducing large-magnitude turbulent fluxes, thereby reducing the reliability of the simulations under high-flux conditions.

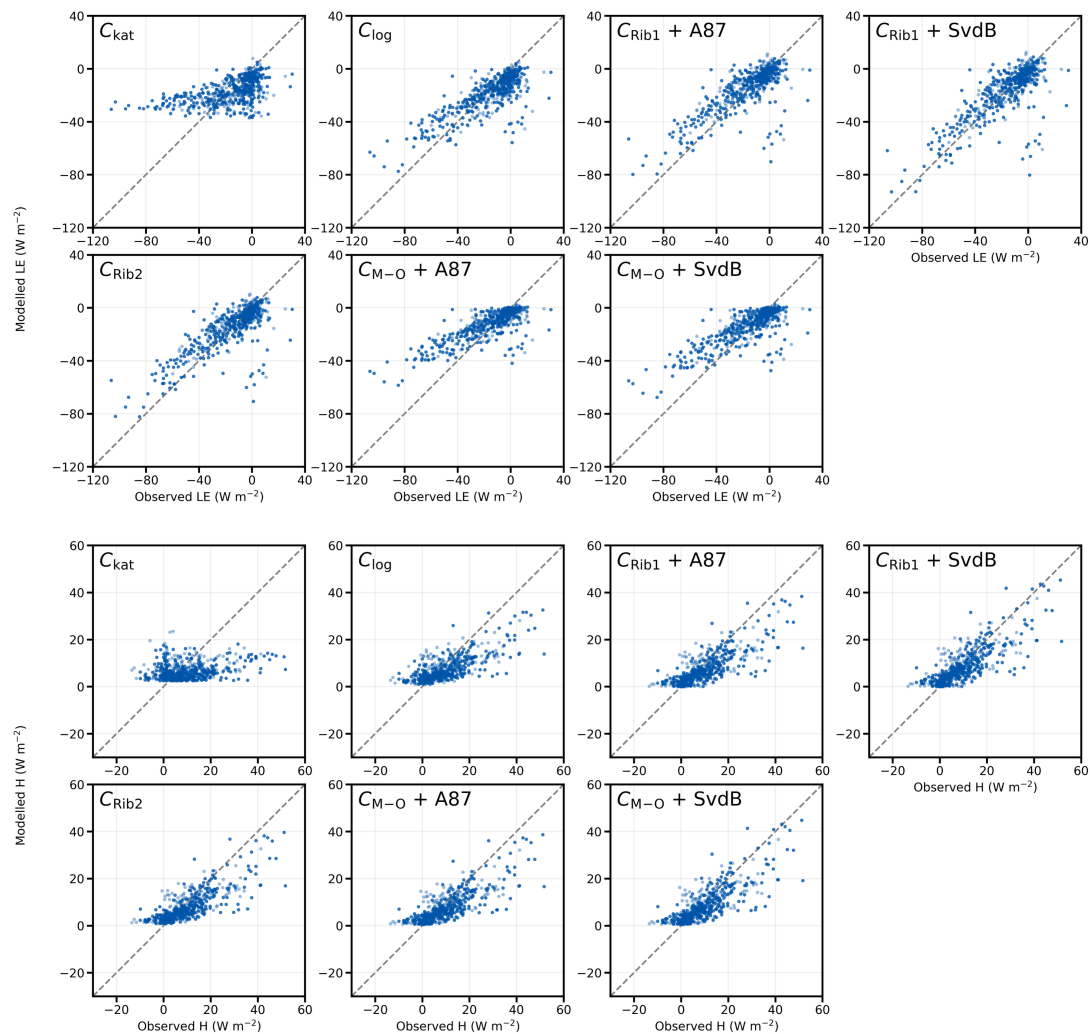


Figure 4: Comparison of 30-min observed (eddy-covariance-derived) and modeled H and LE for the five turbulent flux methods.

Further analysis showed that, under near-neutral conditions, all parameterization schemes generally underestimated H. This systematic underestimation may partly be related to uncertainties in the surface–air temperature gradient, including potential errors in the estimated glacier surface temperature, rather than arising solely from random variability. These results indicate potential limitations of bulk aerodynamic methods in simulating turbulent fluxes over glacier surfaces and suggest that modeled H under near-neutral conditions should be interpreted with caution.

Answer: Thank you for this helpful suggestion. We revised Figure 5 to improve the visibility of the temporal variability and the comparison between modeled and observed turbulent fluxes. Specifically, the symbol sizes were substantially reduced, and the figure was expanded horizontally to occupy the full-page width. We also retained common axis ranges within each column and adjusted the panel labels, spacing, and legend to improve readability. Because the turbulent-flux observations became temporally discontinuous after quality control and data screening, we used scatter points rather than connecting the data with continuous lines, thereby avoiding the visual implication of continuous observations across data gaps.

The revised caption reads:

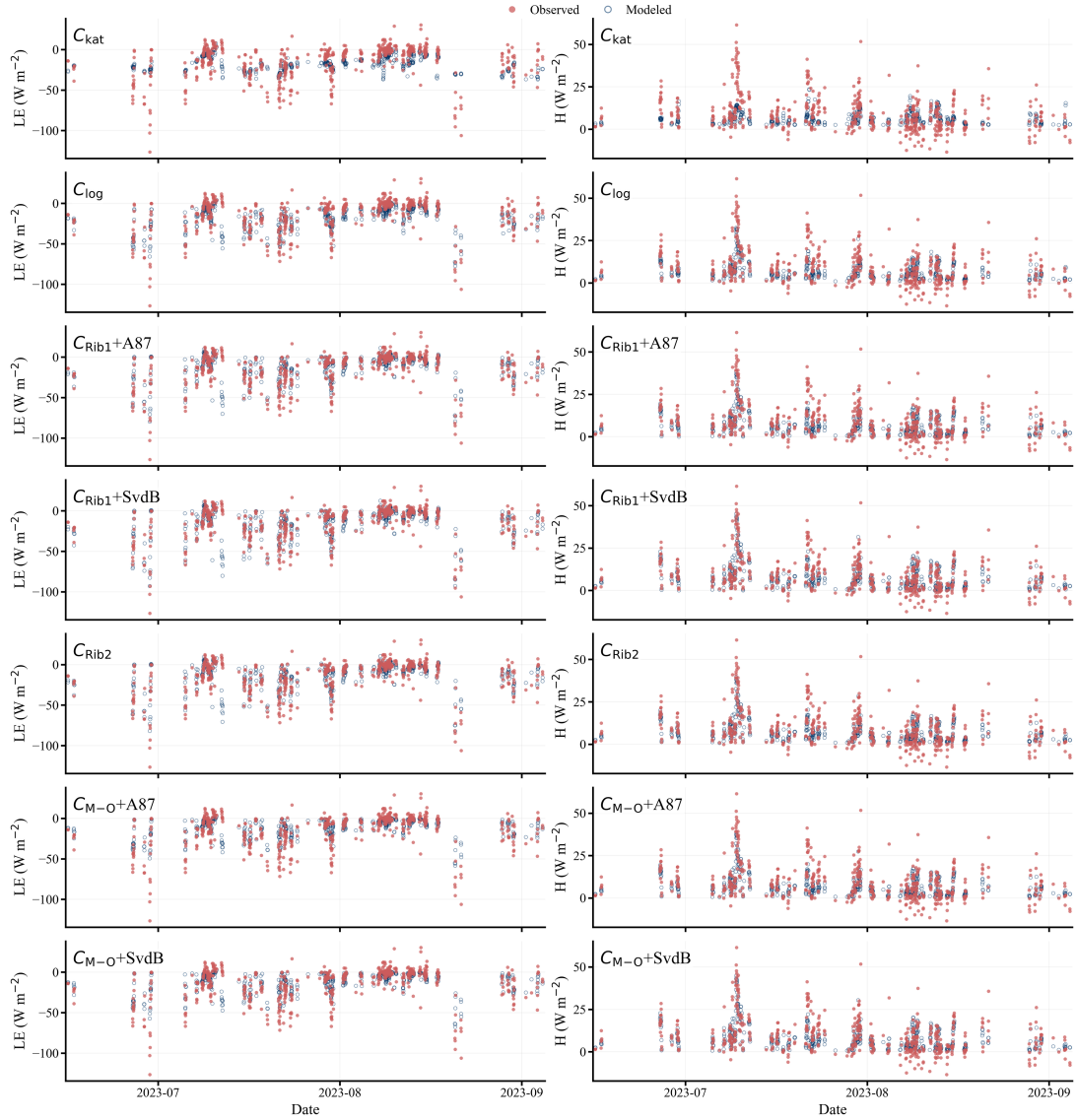


Figure 5: Time series of 30-min LE and H simulated by the five different schemes (C_{kat} , C_{log} , C_{Rib1} , C_{Rib2} , and C_{M-O}), compared with eddy covariance-derived values. The C_{Rib1} , C_{Rib2} , and C_{M-O} schemes were calculated using the observed aerodynamic roughness length ($z_{0m}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

Equation (16): If I am correct z_{0m} and z_{0q} are the same, so please select one term for those.

Answer: Thank you for identifying this notation inconsistency. We understand the Reviewer's comment to refer to the use of both z_{0u} and z_{0m} for the momentum

roughness length. These two symbols represented the same physical quantity in the previous version. We have therefore standardized the notation throughout the manuscript and now use z_{0m} consistently to denote the momentum roughness length. Accordingly, the notation has been corrected in Eq. (17) and the associated text of the revised manuscript; Eq. (17) corresponds to Eq. (16) in the previous version.

For details, please refer to Eq. (17), lines 331 of the revised manuscript:

where $\bar{u}$, $\bar{q}$, and $\bar{T}$ represent the mean wind speed, specific humidity, and air temperature, respectively.

The logarithmic mean heights ($z_{m,t,q}$) are defined as:

$$z_{m,t,q} = \frac{z - z_{0m,t,q}}{\ln\left(\frac{z}{z_{0m,t,q}}\right)} \quad (17)$$

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

Equation (23): Similarly, z_{0e} should probably be z_{0q} (?).

Answer: We agree. The symbol z_{0e} was inconsistent with the notation used elsewhere in the manuscript. It has been replaced by z_{0q} and throughout the corresponding text. The scalar roughness length for water vapour is now consistently denoted by z_{0q} . Accordingly, the notation has been corrected in Eq. (24) and the associated text of the revised manuscript; Eq. (24) corresponds to Eq. (23) in the previous version.

For details, please refer to Eq. (24), lines 364 of the revised manuscript:

$$LE = L_{s/f} \frac{0.622\rho}{P_0} \frac{k^2}{\left[\ln\left(\frac{z}{z_{0m}}\right) - \psi_M\left(\frac{z}{L}\right)\right] \left[\ln\left(\frac{z}{z_{0q}}\right) - \psi_E\left(\frac{z}{L}\right)\right]} u(e_a - e_s), \quad (24)$$

Equation (22): z_{-1} is not defined.

Answer: Thank you for pointing this out. We have now defined z_1 immediately after Eq. (23) as the wind-speed measurement height above the glacier surface. Accordingly, the notation has been corrected in Eq. (23) and the associated text of the revised manuscript; Eq. (23) corresponds to Eq. (22) in the previous version.

For details, please refer to Eq. (23), lines 351 of the revised manuscript:

$$Ri_b = \frac{gz_1}{u^2} \left\{ \frac{T_a - T_s}{T_a} + \frac{q - q_{sat}(T_s, P)}{q + \epsilon / (1 - \epsilon)} \right\}, \quad (23)$$

where ϵ is defined as the ratio of the molecular weights of water to dry air, with a value of 0.622, and z_1 denotes the measurement height of wind speed above the glacier surface.

Table 2: C_p should be with small p . Furthermore, C_h is listed twice. One occurrence should probably be C_H .

Answer: Thank you for pointing this out. We have corrected the notation of the specific heat capacity of air to C_p , with a lowercase subscript p . Regarding the two entries for C_h , these are not duplicate symbols or typographical errors. They represent the same turbulent exchange coefficient used in the C_{log} method, but with two different parameter values: one is the reference value reported in the original publication, whereas the other is the recalibrated value obtained in this study. To avoid confusion, we have retained the symbol C_h for both entries and revised their descriptions in Table 2 to clearly distinguish the “reference value” from the “recalibrated value.”

For details, please refer to Table 2, lines 370 of the revised manuscript:

C_p	heat capacity of air	1005	$J\ kg^{-1}\ K^{-1}$	Standard constant
	turbulent exchange			
C_h	coefficient (recalibrated in this study)	0.0005	—	this study
	turbulent exchange			
C_h	coefficient (reference value)	0.00127	—	Oerlemans, 2000

Section 3.3:

It remains unclear with what (standard) parameter settings the mass balance model was run or what version of the model was used. Please include a reference. This is important for reproducibility of the presented results.

Answer: We agree that the model version, parameter settings, and experimental configuration need to be described more explicitly to ensure reproducibility. We have therefore substantially revised Sect. 3.3.

The revised manuscript now identifies the publicly available MATLAB implementation of the coupled energy balance–snow and firn model used in this study and cites the corresponding EBFM publications.

For details, please refer to Sect. 3.3, lines 372–399 of the revised manuscript:

3.3 Energy Balance Model

We used the publicly available MATLAB implementation of the coupled energy balance–snow and firn model (EBFM; Van Pelt et al., 2012, 2019), obtained from the EBFM-glacier repository at commit 620f835. The surface energy balance (SEB; W m^{-2}) is expressed as

$$Q_m = S_{\text{in}} + S_{\text{out}} + L_{\text{in}} + L_{\text{out}} + H + LE + Q_{\text{rain}} + Q_{\text{sub}} , \quad (26)$$

where Q_m is the energy available for melt; S_{in} and S_{out} are the incoming and reflected shortwave radiation, respectively; L_{in} and L_{out} are the incoming and outgoing longwave radiation, respectively; Q_{rain} is the heat supplied by rainfall; and Q_{sub} is the subsurface heat flux. The energy-balance terms are formulated such that surface temperature is the only unknown variable. Surface temperature was determined iteratively by solving Eq. (26), with melt energy initially set to zero. If the calculated surface temperature exceeds the melting point, it is constrained to the melting point and the energy-balance terms are recalculated. Any resulting positive energy residual is then used for melt.

The mass balance (MB, mm w.e.) is calculated as follows:

$$MB = \int \left(\frac{Q_m}{L_m} + P_{\text{snow}} + \frac{Q_L}{L_s} \right) dt , \quad (27)$$

P_{snow} is the snow accumulation, Q_L is the mass exchange due to sublimation. L_m is the latent heat of melting ($\sim 3.34 \times 10^5 \text{ J kg}^{-1}$), and L_s is the latent heat of sublimation.

The model was applied at the AWS site as a point-scale simulation, and spatial interpolation and topographic-correction routines were therefore not used. Incoming and outgoing shortwave and longwave radiation, surface albedo, snow depth, and rainfall were prescribed directly from the AWS observations. Except for the settings described here, all EBFM parameters were retained at the default values supplied with the public code. A precipitation-phase threshold of 273.75 K was used to distinguish snowfall from rainfall. The turbulent-flux parameters were adjusted only in the experiments designed to evaluate the effects of recalibrating the turbulent-flux methods on SEB model performance. In the baseline simulations, these parameters were assigned the standard values reported in the original publications of the respective turbulent-flux methods. In the recalibrated simulations, the parameter values that minimized the discrepancies between the modeled and observed turbulent fluxes were used for the corresponding methods. All other model settings and forcing data were kept identical between the paired simulations.

Section 5.2:

- The results of the mass balance modelling are currently presented in the Discussion chapter. Please reorganize this, so that mass balance modelling results go to Sect. 4 and the discussion of the results remains in Sect. 5.

Answer: We agree and have reorganized the manuscript accordingly. The quantitative SEB and mass-balance modelling results, together with Figure 7 and Table 5, have been moved from the former Sect. 5.2 to a new Results subsection: “4.4 Effects of turbulent-flux recalibration on SEB model performance”. The new Sect. 4.4 presents the experimental design and quantitative results for turbulent heat fluxes, surface temperature, surface height, and cumulative mass balance. Sect. 5.2 now focuses only on interpretation.

For details, please refer to Sect. 4.4, lines 595–659, and Sect. 5.2, lines 707–743, of the revised manuscript:

4.4 Effects of turbulent-flux recalibration on SEB model performance

To assess the influence of turbulent-flux parameter recalibration on glacier mass-balance simulations, three representative turbulent-flux configurations— $C_{\log}$, C_{Rib2} , and $C_{\text{M-O}} + \text{SvdB}$ were implemented in the coupled energy balance–snow and firn model at the point scale. These configurations were selected because they represent three structurally distinct categories of turbulent-flux parameterization: a simplified bulk formulation without explicit stability corrections, a bulk-Richardson-number-based stability-correction method, and a full Monin–Obukhov similarity-theory method combined with a scalar roughness-length parameterization, respectively. The selection was intended to examine the effects of recalibration across methods with different formulations and levels of complexity while keeping the coupled-model analysis concise and focused. Because the remaining configurations showed broadly comparable overall performance in the preceding turbulent-flux evaluation, the subsequent mass-balance analysis focused on these representative configurations rather than providing an exhaustive comparison of all seven evaluated configurations.

For each selected configuration, paired simulations were conducted using the original and recalibrated turbulent-flux parameter values. In the baseline simulations, the turbulent-flux parameters were assigned the standard values reported in the original publications of the respective methods. In the recalibrated simulations, the corresponding parameter values that minimized the discrepancies between the modeled and observed turbulent fluxes were used. All other model settings and forcing data were kept identical between the paired simulations. Incoming and outgoing longwave radiation, incoming and outgoing shortwave radiation, snow depth, surface albedo, and rainfall were prescribed using the in situ observations at AWS1 to ensure a controlled experimental setup.

Under this framework, the coupled model simulated glacier surface temperature, surface height (SH), and cumulative mass balance (CMB). Because continuous stake-based mass-balance observations were unavailable for the exact modelling period, an observation-constrained reference simulation was constructed to provide a continuous reference CMB time series. In this simulation, the measured meteorological variables and quality-controlled EC turbulent fluxes were prescribed as model forcing. Gaps in the EC turbulent-flux record were filled using the recalibrated C_{Rib2} configuration, whose parameter values minimized the discrepancies between the modeled and observed turbulent fluxes. The CMB simulated under this observation-constrained setup was therefore used as the reference CMB for evaluating the paired simulations.

The reliability of the observation-constrained reference simulation was assessed using independent

measurements. The simulated surface temperature and surface-height variations agreed closely with the corresponding in situ observations. In addition, the independently measured point surface mass balance at AWS1 was -303 mm w.e. from 26 April to 31 December 2023, as indicated by the open star in Fig. 7, which was close to the final observation-constrained reference CMB of -298.6 mm w.e. Although the stake-observation period extended beyond the modelling period, little additional mass-balance change occurred after mid-to-late October because of the persistently cold conditions. The close agreement between the two estimates supports the use of the observation-constrained reference CMB for model evaluation.

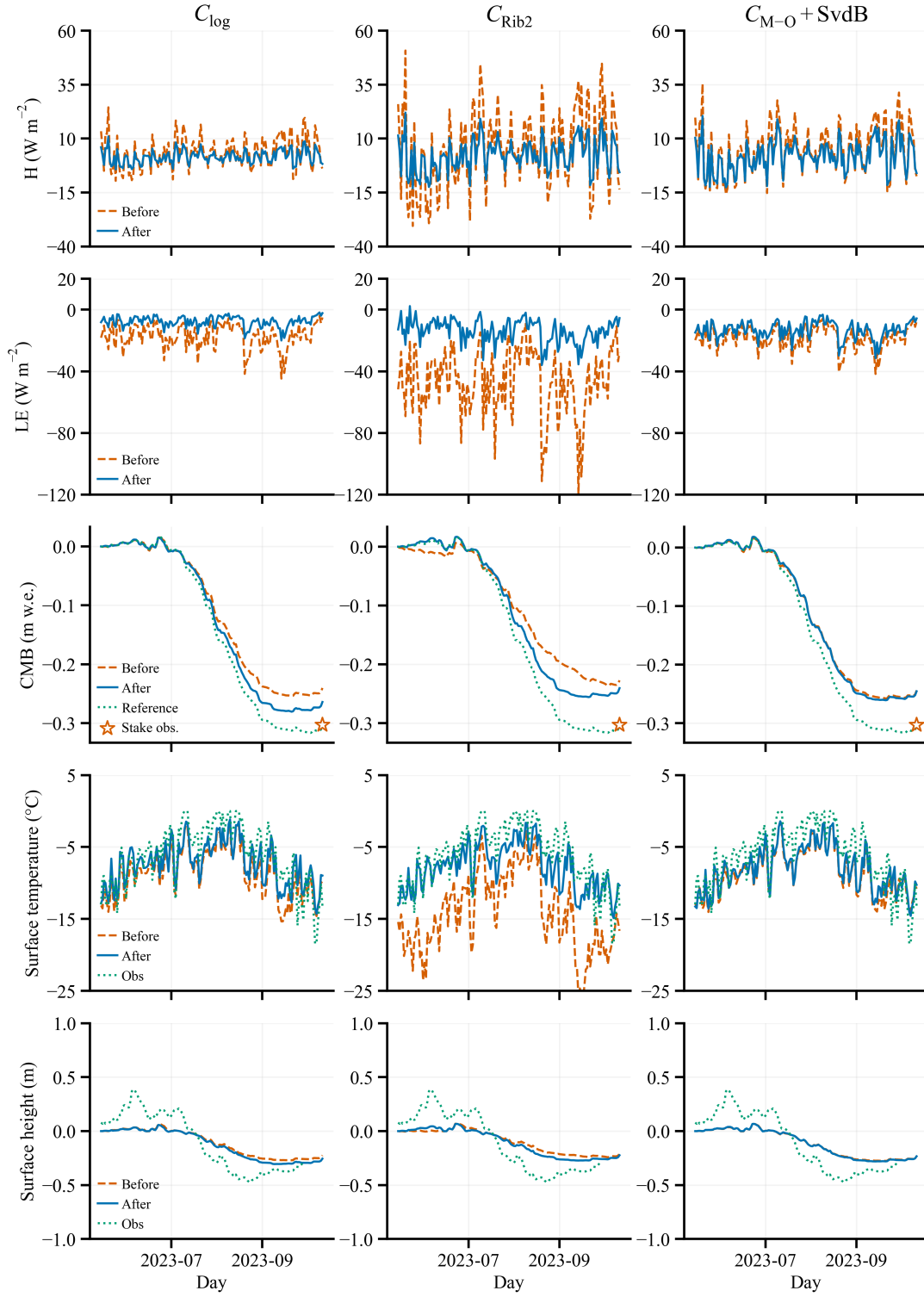


Figure 7: Time series of daily mean turbulent heat fluxes, cumulative mass balance (CMB), surface temperature, and surface height over the Dundee Glacier during the study period, based on the original parameter values (Before), recalibrated parameter values (After), and observations or reference data. “Obs” denotes the EC turbulent-flux observations and the in situ surface-temperature and surface-height observations, whereas “Reference” denotes the observation-constrained CMB generated using EBFM. The

open star in the CMB panels indicates the independently measured stake-based point surface mass balance of -303 mm w.e.

Table 4: Comparative statistics between observations and surface energy balance model simulations of surface temperature (T_s), surface height (SH), and cumulative mass balance (CMB) at the Dundee Glacier, using both original (before) and recalibrated (after) parameters.

Method	Before			After		
	T_s R/RMSE	SH R/RMSE/MAD	CMB (mm w.e.)	T_s R/RMSE	SH R/RMSE/MAD	CMB (mm w.e.)
$C_{\log}$	0.84	0.93		0.85	0.93	
	3.3°C	0.16 m	-238.4	2.5°C	0.15 m	-260.3
		0.14 m			0.14 m	
C_{Rib2}	0.78	0.91		0.87	0.94	
	8.8°C	0.18 m	-225.8	2.8°C	0.15 m	-238.6
		0.16 m			0.14 m	
$C_{M-O} +$ SvdB	0.86	0.94		0.88	0.94	
	3.2°C	0.16 m	-242.4	2.6°C	0.14 m	-249.5
		0.14 m			0.14 m	

We first ran the SEB models using the original parameter values for the three turbulent flux methods (the $C_{\log}$, C_{Rib2} , and $C_{M-O} + \text{SvdB}$ methods). These gave RMSE values for T_s of 3.3, 8.8, and 3.2 °C, respectively. The R values between modeled and measured T_s exceeded 0.78 for all three schemes. The RMSE values for surface height change ranged from 0.16 to 0.18 m, while the CMB estimates were -238.4 , -225.8 , and -242.4 mm w.e., respectively, all deviating markedly from the observation-constrained reference value of -298.6 mm w.e. Following parameter calibration, all three schemes exhibited improvements in model performance. The C_{Rib2} method showed the greatest improvement, giving a CMB value of -238.6 mm w.e., equating to an error reduction of 17.6% relative to the uncalibrated result. In parallel, the RMSE values for T_s and surface height decreased by 68.2% and 16.7%, respectively. The models based on the C_{M-O} and C_{Rib2} methods achieved relatively high overall accuracy after optimization, with R values of 0.88 and 0.87 for T_s , respectively, and 0.94 for surface height in both cases (Table 4).

5.2 Implications of turbulent-flux recalibration for SEB model performance

The paired simulations demonstrate that uncertainties in turbulent-flux parameterization can propagate into the simulated surface temperature, surface-height change, and cumulative mass balance. Because all other forcing data and model settings were held constant, the improvements in the recalibrated simulations can be attributed primarily to the revised turbulent-flux parameters. The simultaneous improvement in T_s , surface height, and CMB indicates that observationally constraining turbulent heat exchange can improve not only the instantaneous surface energy balance but also the integrated mass-balance response.

The magnitude of the improvement differed among the three representative configurations. The particularly large reduction in surface-temperature RMSE for C_{Rib2} suggests that its original parameter setting was less suitable for the atmospheric and glacial conditions of the Dundee Glacier and was therefore more sensitive to recalibration. By contrast, the original C_{log} and $C_{M-O} + SvdB$ configurations already produced comparatively smaller surface-temperature errors, leaving less scope for improvement. Nevertheless, recalibration improved the CMB estimates for all three configurations, indicating that even relatively modest changes in turbulent heat exchange can accumulate over time and affect the modeled mass balance.

Despite these improvements, all recalibrated configurations still underestimated the magnitude of the observation-constrained reference cumulative mass loss. This residual discrepancy indicates that uncertainties in turbulent heat fluxes are not the only source of error in the point-scale SEB simulations. Potential contributions may include uncertainties in surface temperature, precipitation partitioning, subsurface heat storage, snow and ice density, and the representation of snow and firn processes. Therefore, recalibrating turbulent-flux parameters improves SEB model performance but does not eliminate uncertainties arising from other model components and observational constraints.

In addition, the modeled surface-height time series exhibited smaller short-term variability than the observations, particularly following snowfall events. This difference is likely related to the absence of a seasonal snow-compaction scheme in the version of EBFM used in this study. Rather than explicitly simulating the rapid densification of low-density fresh snow immediately after snowfall, this model version assigns a relatively high initial density to fresh snow and subsequently treats its compaction in a manner more representative of firn. Consequently, rapid post-snowfall changes in surface height may be dampened in the simulations. In addition to this model limitation, wind-driven snow redistribution may also contribute to the larger short-term variability observed in the field. Wind erosion, transport, and

redemption can produce rapid local changes in snow depth and surface height, thereby increasing the variability of the observed surface-height record.

Because the three configurations represent distinct formulations and levels of structural complexity, the improvements obtained across all three indicate that the benefits of observationally constrained turbulent-flux parameters are not limited to a single method type, although the magnitude of the response remains method dependent.

- It remains a mystery why only three (out of seven) of the turbulent flux methods were included in the mass balance modelling, and why these three were selected. Please indicate this.

Answer: Thank you for requesting clarification. We have added a detailed explanation at the beginning of Sect. 4.4. The three configurations were selected for three main reasons. First, they represent structurally distinct types of turbulent-flux parameterization, including a simplified bulk formulation without explicit stability corrections, a bulk-Richardson-number-based stability-correction method, and a full Monin–Obukhov similarity-theory method combined with a scalar roughness-length parameterization. Second, restricting the coupled SEB and mass-balance experiments to a small number of representative configurations helped keep the analysis concise and focused. Third, the overall performance of the remaining configurations was broadly comparable to that of the selected configurations in the preceding turbulent-flux evaluation. We therefore focused the subsequent analysis on these representative methods and their results rather than extending the mass-balance experiments to all seven configurations. We have clarified these considerations in Sect. 4.4.

For details, please refer to Sect. 4.4, lines 595–607, of the revised manuscript:

4.4 Effects of turbulent-flux recalibration on SEB model performance

To assess the influence of turbulent-flux parameter recalibration on glacier mass-balance simulations, three representative turbulent-flux configurations— $C_{\log}$, $C_{\text{Rib}2}$, and $C_{\text{M-O}} + \text{SvdB}$ were

implemented in the coupled energy balance–snow and firn model at the point scale. These configurations were selected because they represent three structurally distinct categories of turbulent-flux parameterization: a simplified bulk formulation without explicit stability corrections, a bulk-Richardson-number-based stability-correction method, and a full Monin–Obukhov similarity-theory method combined with a scalar roughness-length parameterization, respectively. The selection was intended to examine the effects of recalibration across methods with different formulations and levels of complexity while keeping the coupled-model analysis concise and focused. Because the remaining configurations showed broadly comparable overall performance in the preceding turbulent-flux evaluation, the subsequent mass-balance analysis focused on these representative configurations rather than providing an exhaustive comparison of all seven evaluated configurations.

- The cumulative mass balance observations and surface height measurements are currently not described (unless I missed it). This information should be included in Chapter 2. The analysis in Sect. 5.2 for example depends on what density was used to convert cumulative surface height change observations to CMB estimates.

Answer: We agree that the surface-height and point mass-balance measurements required a clearer description. We have therefore added the corresponding information to Sect. 2.2 and Table 1.

For details, please refer to Sect. 2.2, lines 144-163 of the revised manuscript:

2.2. Meteorological data

Meteorological variables, including air temperature, air pressure, relative humidity, wind speed and direction, incoming and outgoing shortwave radiation, and incoming and outgoing longwave radiation, were primarily measured at the study site using a Campbell-ClimaVUE50 automatic weather station (AWS1) located at 5,317 m a.s.l. Glacier surface height was measured independently using a Campbell SR50 sonic ranging sensor installed at AWS1. The sensor records the distance between the instrument and the glacier surface, from which temporal changes in surface height were derived. Occasional gaps in the AWS1 records were filled using linear relationships established from concurrent observations at the

nearby auxiliary station (AWS2). The glacier surface temperature used in this study was derived from the observed outgoing longwave radiation, assuming a constant surface emissivity of 0.98. Details of the sensors used to measure each variable are provided in Table 1.

The point surface mass balance at the Dundee Glacier was measured using a single stake installed near AWS1 following the glaciological method described by Paterson (1994). During the observation period from 26 April to 31 December 2023, snow pits were excavated during periods of snow cover to measure snow layer density and stratigraphy. The measurements included the stake height above the glacier surface; the thickness and density of snow, firn, and superimposed ice layers, where present; and the stratigraphic structure of the snow-pit profiles (Pu et al., 2008). These measurements were used to calculate the net point surface mass balance over the observation period.

Table1 (L174) :

Glacier surface height	(m)	Campbell SR50	0.01 m	-	2.05 m
---------------------------	-----	---------------	--------	---	--------

We have also clarified that the continuous CMB time series used to evaluate the paired simulations was not obtained by converting the SR50 surface-height measurements using an assumed constant density. Instead, because continuous stake-based mass-balance observations were not available for the exact model period, an observation-constrained EBFM reference simulation was constructed.

In this reference simulation, measured meteorological variables and quality-controlled EC turbulent fluxes were used as model forcing. Gaps in the EC turbulent-flux record were filled using the recalibrated C_{Rib2} configuration, with parameter values optimized by minimizing the sum of the RMSEs between the modeled and observed sensible and latent heat fluxes. The CMB produced by this simulation is therefore referred to as the “observation-constrained reference CMB,” rather than “observed CMB.”

The reliability of this reference CMB was independently evaluated using the stake-based point mass balance. The stake measurements yielded a cumulative point mass

balance of -303mm w.e. from 26 April to 31 December 2023, which was close to the observation-constrained reference CMB of -298.6mm w.e.

For details, please refer to Sect. 4.4, lines 616–633 of the revised manuscript:

Under this framework, the coupled model simulated glacier surface temperature, surface height (SH), and cumulative mass balance (CMB). Because continuous stake-based mass-balance observations were unavailable for the exact modelling period, an observation-constrained reference simulation was constructed to provide a continuous reference CMB time series. In this simulation, the measured meteorological variables and quality-controlled EC turbulent fluxes were prescribed as model forcing. Gaps in the EC turbulent-flux record were filled using the recalibrated C_{Rib2} configuration, whose parameter values minimized the discrepancies between the modeled and observed turbulent fluxes. The CMB simulated under this observation-constrained setup was therefore used as the reference CMB for evaluating the paired simulations.

The reliability of the observation-constrained reference simulation was assessed using independent measurements. The simulated surface temperature and surface-height variations agreed closely with the corresponding in situ observations. In addition, the independently measured point surface mass balance at AWS1 was -303 mm w.e. from 26 April to 31 December 2023, as indicated by the open star in Fig. 7, which was close to the final observation-constrained reference CMB of -298.6 mm w.e. Although the stake-observation period extended beyond the modelling period, little additional mass-balance change occurred after mid-to-late October because of the persistently cold conditions. The close agreement between the two estimates supports the use of the observation-constrained reference CMB for model evaluation.

Figure 7:

- The lines are too thick, which makes that there is a lot of overlap between them.

Answer: We agree and have reduced the line widths throughout Figure 7. This change decreases overlap among the time series and improves the visibility of differences

between simulations using the original and recalibrated parameters and the observations or reference data.

- Are the fluxes daily mean values? This information should be included in the caption.

Answer: Yes, the turbulent heat fluxes shown in Figure 7 are daily mean values. We have added this information explicitly to the revised caption.

- For comparison purposes it would be helpful to have the same y-axis range per row. This is currently not the case for the fluxes and surface temperature.

Answer: We agree and have standardized the vertical-axis ranges within each row of Figure 7. The three turbulent-flux panels now use the same y-axis range and tick intervals, and the three surface-temperature panels likewise use a common y-axis range. This revision enables direct comparison among the three selected turbulent-flux configurations.

- It is confusing that the green line is showing something else than observations in the uppermost panels. Please choose another colour in the uppermost panels for clarity.

Answer: We agree and have revised the colour scheme in Figure 7. Green is no longer used for modeled turbulent heat fluxes in the upper panels.

The upper panels now use separate red and blue colour families for sensible and latent heat fluxes. Lighter shades represent simulations using the original parameter values, whereas darker shades represent simulations using the recalibrated parameter values. Green is reserved for observational or reference time series in the lower panels. This consistent colour scheme avoids confusion between modeled and observed data.

For details, please refer to Fig 7, lines 635–642 of the revised manuscript:

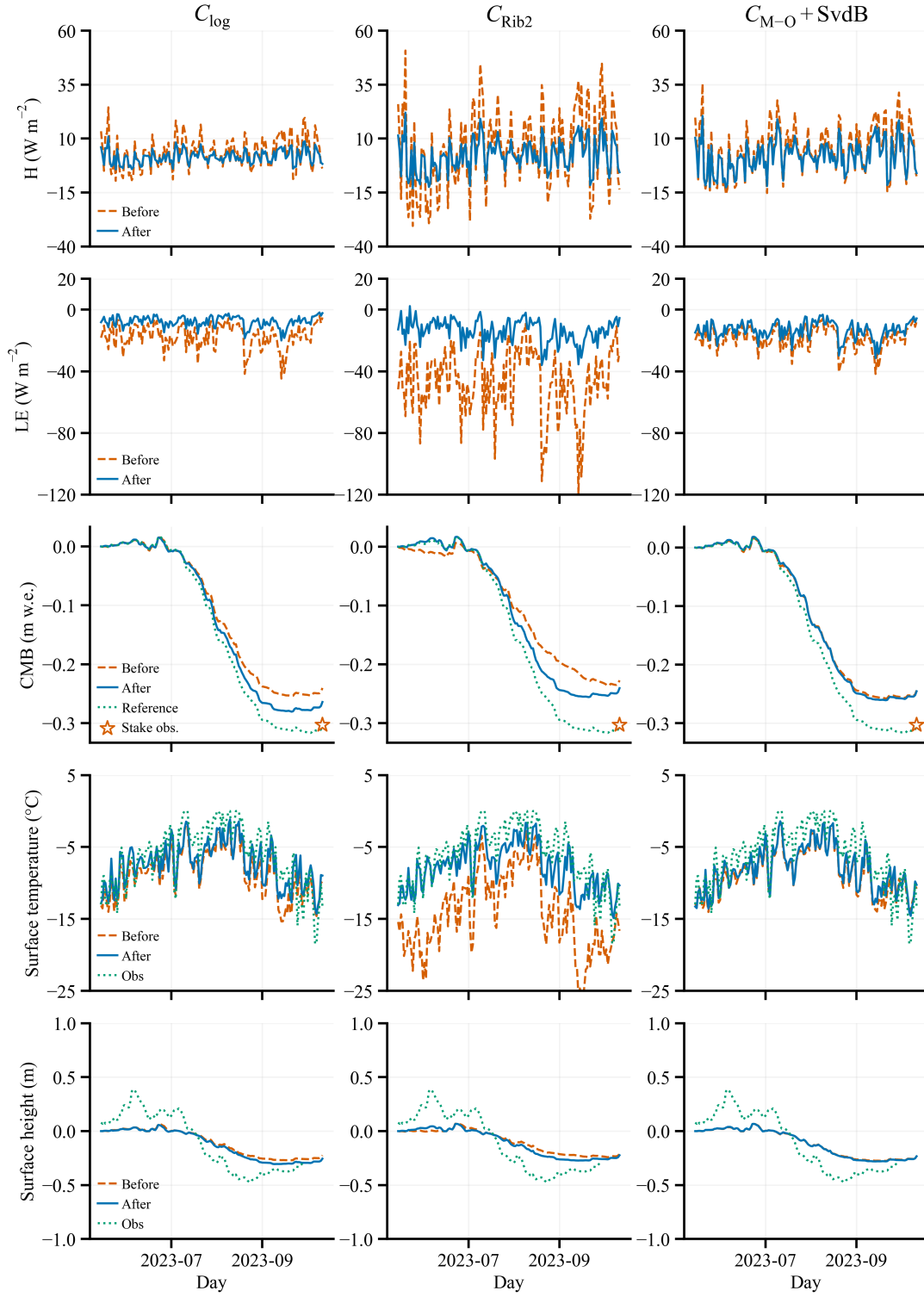


Figure 7: Time series of daily mean turbulent heat fluxes, cumulative mass balance (CMB), surface temperature, and surface height over the Dundee Glacier during the study period, based on the original parameter values (Before), recalibrated parameter values (After), and observations or reference data. “Obs” denotes the EC turbulent-flux observations and the in situ surface-temperature and surface-height observations, whereas “Reference” denotes the observation-constrained CMB generated using EBFM. The

open star in the CMB panels indicates the independently measured stake-based point surface mass balance of -303 mm w.e.

- It could be discussed in the text that the smaller surface height variability in the modelled timeseries (lowermost panels in Fig. 7) is a result of the lack of a seasonal snow compaction model in the used version of EBFM. Instead of simulating quick densification of low-density snow directly after a snowfall event, EBFM directly sets a much higher “fresh snow density” and simulates compaction as if it is firn. The newer versions of EBFM (<https://github.com/EBFMorg/EBFM>) also contain a seasonal snow compaction model.

Answer: Thank you for this important explanation. We have added a new paragraph to Sect. 5.2 discussing the smaller short-term variability of the modeled surface-height time series. Following the Reviewer’s suggestion, we now clarify that this discrepancy is partly related to the absence of a seasonal snow-compaction scheme in the version of EBFM used in this study. In addition, wind-driven snow redistribution may also contribute to the larger short-term variability observed in the field, because erosion, transport, and redeposition of snow can produce rapid local changes in snow depth and surface height that are not represented in the present point-scale simulations.

For details, please refer to Sect. 5.2, lines 723–723 of the revised manuscript:

Despite these improvements, all recalibrated configurations still underestimated the magnitude of the observation-constrained reference cumulative mass loss. This residual discrepancy indicates that uncertainties in turbulent heat fluxes are not the only source of error in the point-scale SEB simulations. Potential contributions may include uncertainties in surface temperature, precipitation partitioning, subsurface heat storage, snow and ice density, and the representation of snow and firn processes. Therefore, recalibrating turbulent-flux parameters improves SEB model performance but does not eliminate uncertainties arising from other model components and observational constraints.

In addition, the modeled surface-height time series exhibited smaller short-term variability than the observations, particularly following snowfall events. This difference is likely related to the absence of a

seasonal snow-compaction scheme in the version of EBFM used in this study. Rather than explicitly simulating the rapid densification of low-density fresh snow immediately after snowfall, this model version assigns a relatively high initial density to fresh snow and subsequently treats its compaction in a manner more representative of firn. Consequently, rapid post-snowfall changes in surface height may be dampened in the simulations. In addition to this model limitation, wind-driven snow redistribution may also contribute to the larger short-term variability observed in the field. Wind erosion, transport, and redeposition can produce rapid local changes in snow depth and surface height, thereby increasing the variability of the observed surface-height record.

Because the three configurations represent distinct formulations and levels of structural complexity, the improvements obtained across all three indicate that the benefits of observationally constrained turbulent-flux parameters are not limited to a single method type, although the magnitude of the response remains method dependent.

We sincerely thank the Editor and the Reviewer again for their careful and constructive comments. We believe that these revisions have improved the clarity, reproducibility, organization, and scientific interpretation of the manuscript, and we hope that the revised manuscript is now suitable for publication.

Sincerely,

Yichan Xu, on behalf of all authors