We thank both reviewers for their useful comments and positive feedback. The suggested revisions substantially improved the manuscript, and we addressed all comments, point-by-point, in this document. The comments of the reviewers are shown in black and our replies in blue. We number reviewer comments for referencing purposes throughout the document (comment 1 = C1, etc.). The line, figure and table numbers are based on the updated manuscript.

# Responses to Referee #2

This paper investigates the use of ERA5 reanalysis data set, with and without dynamical downscaling to evaluate surface energy balance modeling of glaciers in western Canada. The authors look specifically at the variables that are used to calculate the energy available for melt when evaluating the different forcing sets. For downscaling they use the Weather Research Forecasting (WRF) model, and include several tests with various parametrization combinations in WRF and try to determine the best combination for the western region in Canada. I believe this paper is a nice contribution to anyone interested in glacier mass balance studies in this and similar regions and should be published after addressing the comments and suggestions.

We thank the reviewer for their useful comments and positive feedback that helped us improve the quality of the manuscript. By addressing these comments, we were able to improve the discussion of some crucial points.

## Major Comment

*R2 C1:* Page 12, line 257. Does the Fitzpatrick 2019 paper actually support neglection of heat flux from rain (or ground?) The only thing I could find was that "while readings from periods affected by precipitation on the analyser windows were removed" for eddy covariance data. That does not warrant negligible contribution from rain. I suggest removing reference to the 2019 paper. However, the Fitzpatrick 2017 paper is a nice citation. And I found this paragraph interesting: "QR provided a negligible contribution (<1%) to the total melt energy over the recorded period. However, over daily and sub-daily timescales, QR was observed to have a considerable influence on SEB and ablation during heavy rainfall."
Precipitation is one of the outputs you want to get correct in WRF. Though it might not have been important for the study window in this manuscript, I do believe a WRF parameterization combination that gets the precipitation correct (along with other metrics) is desired (and as stated in the Fitzpatrick 2017 paper, "QR was observed to have a considerable influence on SEB and ablation during heavy rainfall"). I would have liked to see which parameter combinations score the best when precipitation also is included.

We thank the reviewer for this comment. In response, we have included more references justifying the omission of rain flux and ground heat flux from our analysis, as these factors have been shown to have negligible contributions to the seasonal melt energy. We now highlight that we are primarily interested in the performance of SEB modeling for seasonal melt simulations and in exploring the suitability of ERA5 for long-term glacier melt simulations, both with and without downscaling. We agree with the reviewer that precipitation is an important variable, especially for the simulation of accumulation and therefore the glacier-wide mass balance.

Although precipitation (as rainfall) can significantly contribute to the melt energy on specific days during a melt season, as shown in Fitzpatrick et al. (2017), uncertainties associated with assessing the rain heat flux in models are relatively large and potentially grounded in unsupported assumptions (Hock, 2005; Fitzpatrick et al., 2017). However, more important for SEB modeling is precipitation in the form of snowfall, which can substantially alter the albedo and therefore the melt energy depending on the frequency of these fresh snowfall events during a melt season (e.g., Hock, 2005; Mac Dougall et al., 2011; Marshall and Miller, 2020).

We now elaborate on these limitations in more detail in the "Data and Methods" section when justifying the selection of our SEB model (lines 264-270). Following the reviewer's suggestion, we have included a comparison between modeled (from WRF and ERA5) and observed daily precipitation at our research sites (lines 355-362 and 374-379, Figures S2 and S3). We also updated Tables 5, S4 and S5 to include precipitation.

*Lines 264-270: Given our focus on the key seasonal SEB components, we neglect the ground heat flux and the heat flux from rain, since both have been shown to give negligible contributions to the total seasonal melt at mid-latitude glaciers (Sicart et al., 2005; Andreassen et al., 2008; Gillett and Cullen, 2011), as well as at our study sites (Fitzpatrick et al., 2017; Fitzpatrick, 2018; Lord-May and Radić, 2023). The rain heat flux, however, can be a substantial contributor (up to 20 %) to daily melt energy on a day with extreme rainfall (Fitzpatrick et al., 2017), but the uncertainty in the model used to assess the rain heat flux is relatively large (Hock, 2005; Fitzpatrick et al., 2017). For these reasons, we neglect the heat flux in the SEB model, but will include precipitation, both as rainfall and snowfall, in the evaluation analysis.*

*Lines 355-362: Seasonally-averaged total daily precipitation (P) from ERA5 is overestimated at some glacier sites (relative error of 130.3 % for Kaskawulsh 2019 and 84.5 % for Conrad 2016 $AWS_1$), while underestimated at other sites (relative error of –37.9 % for Conrad 2016 $AWS_2$, –33.2 % for Conrad 2015, and –18.7 % for Nordic 2014; Figures S2 and S3). On the other hand, the modeled timeseries of daily precipitation all show statistically significant correlation ($p$-value $< 0.05$) with the observed timeseries (Figure S2). The frequency of days with heavy snowfall is overestimated in the ablation zones (Castle Creek 2012, Kaskawulsh 2019) and underestimated in the accumulation zone (Conrad 2016 $AWS_2$). On average, ERA5-Land precipitation simulations perform worse than ERA5 (mean overestimation of 35.6 % in ERA5-Land vs. 20.3 % in ERA5 across all sites).*

*Lines 374-379: The modeled timeseries of daily precipitation shows statistically significant correlation ($p$-value $< 0.05$) with the observed timeseries for all sites except for Kaskawulsh glacier (Figure S2). WRF tends to overestimate the frequency of days with heavy snowfall in both the ablation zone (Castle Creek 2012, Kaskawulsh 2019) and the accumulation zone (Conrad 2016 $AWS_2$ ; Figure S3). Additionally, it also overestimates the frequency of days with light rainfall (< 2.5 mm). On average, precipitation values from WRF at 3.3 km perform significantly worse than from WRF at 1.1 km (29.0 % vs. –5.0 % relative error with the REF configuration; Table 5).*

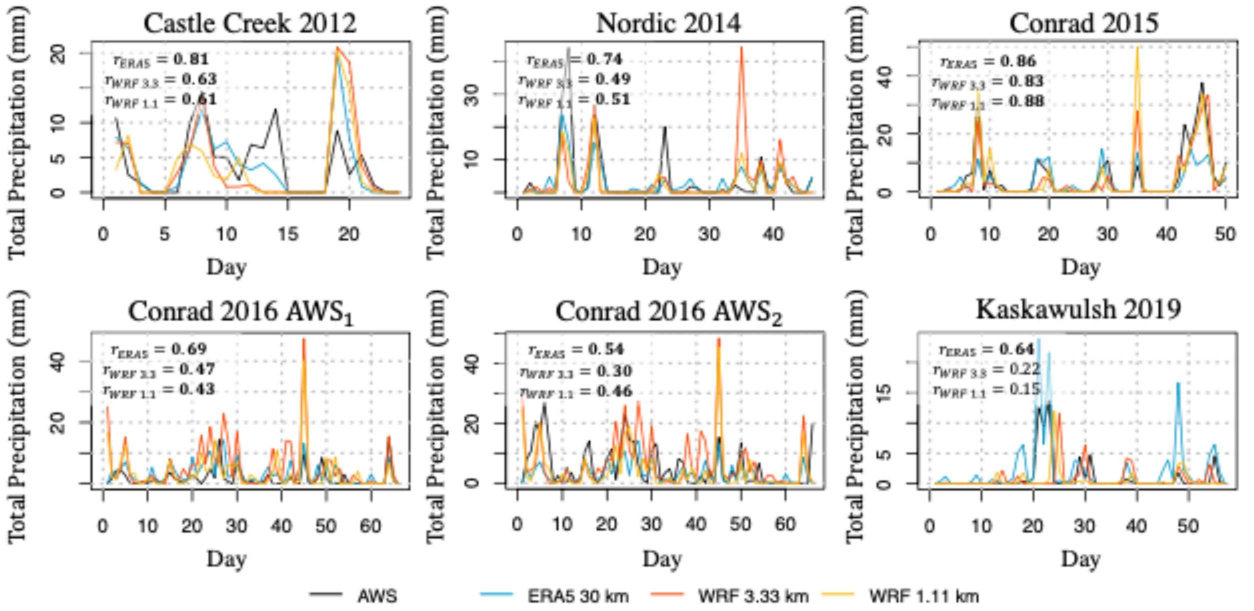
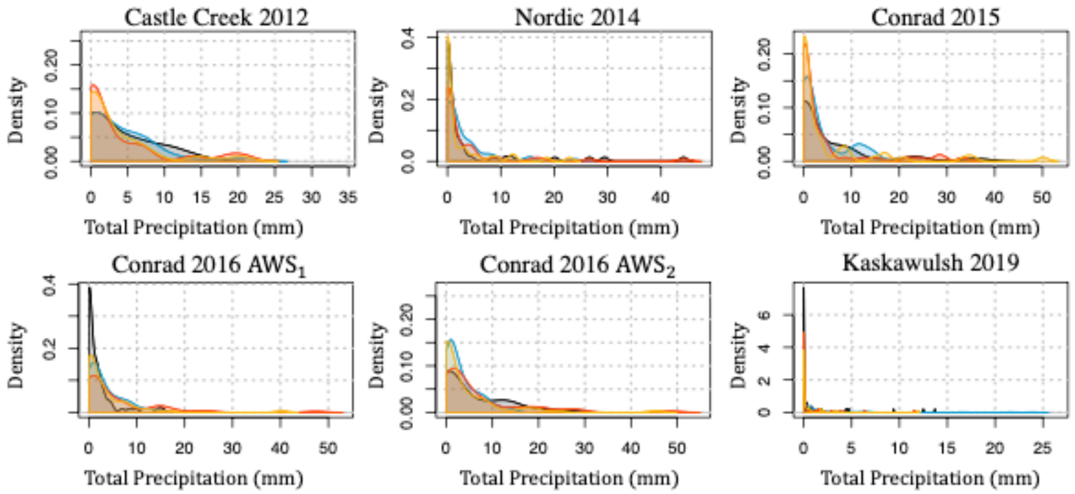
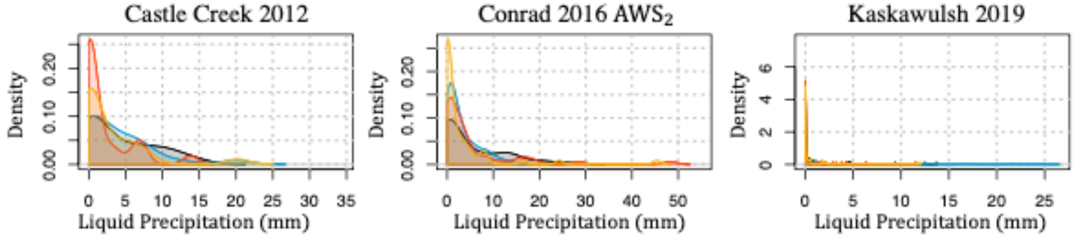


**Figure S2:** Modeled (ERA5, WRF at 3.3 km and WRF at 1.1 km) versus observed (AWS data) timeseries of daily total precipitation over the observational period. Bold values of $r_{sp}$ indicate a statistically significant correlation at the 5 % confidence level. WRF is run with the REF configuration.

**(a) Density of Total Precipitation**



**(b) Density of Rainfall**



**(c) Density of Snowfall**

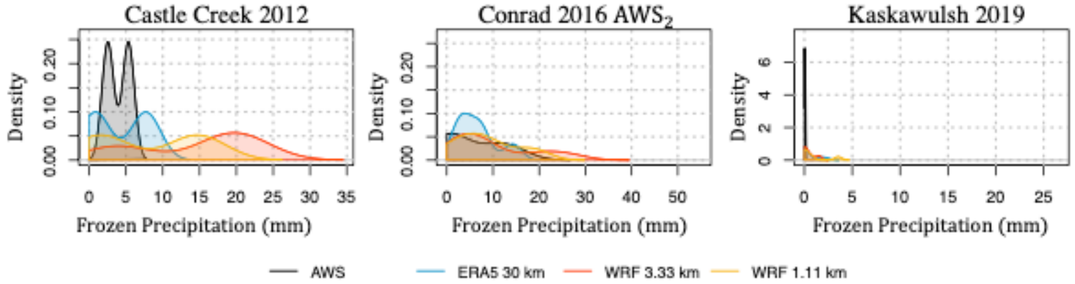


**Figure S3:** Modeled (ERA5, ERA5-Land, WRF at 3.3 km and WRF at 1.1 km) versus observed (AWS data) densities of (a) total, (b) liquid and (c) frozen daily averages of precipitation over the observational period. We differentiate between liquid and frozen precipitation using a temperature threshold of 0 °C. The densities of liquid and frozen precipitation are displayed only if there were days with snowfall during the observational period. WRF is run with REF parameterization schemes.

**Table 5.** Relative difference (%) between modeled and observed seasonally-averaged values, as well as NRMSE between modeled and observed daily values of: air temperature ($T$), relative humidity ($RH$), total precipitation ($P$), wind speed ($U$), incoming shortwave ($K_{in}$) and longwave ($L_{in}$) radiation, sensible ($Q_H$) and latent ($Q_L$) heat fluxes, and total melt energy ($Q_M$). The melt energy is estimated according to the SEB model (Eq. 1). The WRF runs are based on the three configurations of physics parameterizations: REF, minNRMSE and TOPSIS. For comparison, we also include the results of the ensemble-mean, where $T$, $RH$, $P$, $U$, $K_{in}$ and $L_{in}$ are derived as a mean across the three configurations. The turbulent heat fluxes and $Q_M$ in the ensemble-mean are derived according to the aerodynamic bulk method (Eqs. 2 and 3) and SEB model (Eq. 1), respectively. The results of each evaluation metric are shown as the mean (± one standard deviation) across the six study sites, with equal weighing of each site. For seasonally-averaged values of $Q_M$, we only take into account positive values of $Q_M$ that drive melt. Values in bold highlight the best performing model for the given variable according to the metric used.

| **Variable** | **ERA5** | **ERA5-Land** | **WRF 3.3 km** | | | | **WRF 1.1 km** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **30 km** | **9 km** | **REF** | **minNRMSE** | **TOPSIS** | **Ensemble** | **REF** | **minNRMSE** | **TOPSIS** | **Ensemble** |
| **Relative Error (%)** | | | | | | | | | | |
| $\mathbf{T}$ | 14 ± 24 | 23 ± 23 | **3** ± 37 | -21 ± 48 | 9 ± 37 | **-3** ± 40 | 6 ± 30 | -25 ± 39 | 11 ± 29 | **-3** ± 32 |
| $\mathbf{RH}$ | 30 ± 12 | 28± 12 | **-2** ± 14 | 17 ± 22 | -6 ± 12 | 3 ± 14 | -9 ± 11 | 10 ± 16 | -12 ± 9 | -4 ± 12 |
| $\mathbf{P}$ | 20 ± 70 | 36± 99 | 29 ± 88 | 46 ± 101 | **1** ± 45 | 25 ± 77 | -5 ± 51 | 12 ± 53 | -15 ± 27 | -3 ± 43 |
| $\mathbf{U}$ | -64 ± 10 | -74 ± 6 | -26 ± 18 | -42 ± 12 | **-22** ± 20 | -30 ± 15 | -44 ± 16 | -46 ± 14 | -42 ± 15 | -44 ± 14 |
| $\mathbf{K_{in}}$ | 11 ± 8 | 11 ± 8 | 12 ± 6 | -16 ± 16 | -4 ± 9 | -3 ± 9 | 19± 16 | -11 ± 18 | **1** ± 11 | 3 ± 14 |
| $\mathbf{L_{in}}$ | **0** ± 4 | -1 ± 4 | -5 ± 3 | **0** ± 4 | -3 ± 3 | -3 ± 3 | -5 ± 3 | -1 ± 3 | -3 ± 2 | -3 ± 2 |
| $\mathbf{Q_H}$ | -87 ± 11 | -95 ± 5 | -43 ± 38 | -81± 14 | **-40**± 40 | -58 ± 22 | -64 ± 16 | -82± 13 | -64 ± 16 | -73 ± 13 |
| $\mathbf{Q_L}$ | 520 ± 977 | **113** ± 335 | -613 ± 1203 | -115 ± 207 | -804 ± 1533 | -399 ± 758 | -383 ± 453 | -235 ± 298 | -531 ± 702 | -368 ± 437 |
| $\mathbf{Q_M}$ | -6 ± 7 | -10 ± 7 | **-5** ± 10 | -28 ± 10 | -15 ± 15 | -26 ± 16 | -8 ± 7 | -27 ± 10 | -19 ± 8 | -29 ± 15 |
| **NRMSE (%)** | | | | | | | | | | |
| $\mathbf{T}$ | 22 ± 10 | 24 ± 10 | 23 ± 8 | 28 ± 6 | 24 ± 8 | **21** ± 8 | **21** ± 13 | 25 ± 6 | 22 ± 14 | **21** ± 10 |
| $\mathbf{RH}$ | 50 ± 8 | 47 ± 7 | 32 ± 22 | 47 ± 20 | 32 ± 23 | 32 ± 20 | 33 ± 22 | 35 ± 13 | 35 ± 21 | **29** ± 17 |
| $\mathbf{P}$ | **21** ± 5 | 24 ± 9 | 31 ± 14 | 35 ± 16 | 30 ± 13 | 28 ± 12 | 27 ± 10 | 30 ± 13 | 27 ± 12 | 24 ± 10 |
| $\mathbf{U}$ | 55 ± 15 | 62± 16 | 40 ± 8 | 44 ± 9 | **38** ± 8 | **38** ± 9 | 43 ± 10 | 45 ± 13 | 44 ± 9 | 43 ± 10 |
| $\mathbf{K_{in}}$ | **17** ± 2 | **17** ± 3 | 26 ± 2 | 30 ± 11 | 26 ± 5 | 23 ± 4 | 29 ± 5 | 29 ± 9 | 27 ± 4 | 24 ± 2 |
| $\mathbf{L_{in}}$ | **18** ± 6 | **18** ± 6 | 30 ± 7 | 28 ± 5 | 32 ± 8 | 27 ± 6 | 31 ± 8 | 27 ± 7 | 30 ± 8 | 26 ± 8 |
| $\mathbf{Q_H}$ | 39 ± 13 | 41 ± 12 | 31 ± 8 | 35 ± 8 | **30** ± 7 | 31 ± 9 | 31 ± 9 | 35 ± 9 | 32 ± 9 | 34 ± 10 |
| $\mathbf{Q_L}$ | 25 ± 6 | 23 ± 8 | 23 ± 5 | 22 ± 6 | 24 ± 3 | 22 ± 6 | 21 ± 5 | 21 ± 6 | 22 ± 5 | **20** ± 6 |
| $\mathbf{Q_M}$ | **14** ± 4 | 15 ± 4 | 18 ± 3 | 23 ± 7 | 21 ± 4 | 19 ± 4 | 16 ± 3 | 22± 6 | 19 ± 3 | 18 ± 3 |

**Table S4:** Model performance, evaluated by $r_{sp}$ and NNSE, over the whole observational period in simulating daily: air temperature ($T$), relative humidity ($RH$), total precipitation ($P$), wind speed ($U$), incoming shortwave ($K_{in}$) and longwave ($L_{in}$) radiation, sensible ($Q_H$) and latent ($Q_L$) heat fluxes and total melt energy ($Q_M$). The melt energy is estimated according to the SEB model (Eq. 1). The WRF runs are based on three configurations of physics parameterizations: REF, minNRMSE and TOPSIS. The model performance is shown as the mean (± one standard deviation) across the six study sites, with equal weighing of each site. Values in bold highlight the best performing model for the given variable. Values in purple highlight a statistically significant correlation at the 5 % confidence level for at least four of the six glacier sites.

| Variable | ERA5 | ERA5-Land | WRF 3.3 km | | | WRF 1.1 km | | |
|---|---|---|---|---|---|---|---|---|
| | 30 km | 9 km | REF | minNRMSE | TOPSIS | REF | TOPSIS | minNRMSE |
| | | | | $\mathrm{r_{sp}}$ | | | | |
| **T** | 0.86 ± 0.19 | 0.89 ± 0.14 | 0.89 ± 0.05 | 0.86 ± 0.04 | 0.88 ± 0.03 | **0.91** ± 0.03 | 0.88 ± 0.03 | 0.90 ± 0.01 |
| **RH** | 0.74 ± 0.20 | **0.75** ± 0.15 | 0.70 ± 0.21 | 0.55 ± 0.19 | 0.72 ± 0.19 | 0.71 ± 0.20 | 0.53 ± 0.25 | 0.71 ± 0.17 |
| **P** | **0.71** ± 0.11 | **0.71** ± 0.11 | 0.49 ± 0.22 | 0.44 ± 0.28 | 0.55 ± 0.21 | 0.51 ± 0.24 | 0.53 ± 0.24 | 0.49 ± 0.26 |
| **U** | 0.19 ± 0.36 | 0.13 ± 0.32 | 0.15 ± 0.15 | 0.10 ± 0.16 | 0.16 ± 0.20 | **0.23** ± 0.20 | 0.20 ± 0.20 | 0.19 ± 0.24 |
| $\mathbf{K_{in}}$ | **0.80** ± 0.09 | **0.80** ± 0.09 | 0.60 ± 0.10 | 0.58 ± 0.19 | 0.56 ± 0.11 | 0.52 ± 0.16 | 0.55 ± 0.18 | 0.47 ± 0.14 |
| $\mathbf{L_{in}}$ | **0.79** ± 0.11 | 0.78 ± 0.12 | 0.62 ± 0.07 | 0.56 ± 0.11 | 0.41 ± 0.12 | 0.54 ± 0.07 | 0.54 ± 0.14 | 0.35 ± 0.17 |
| $\mathbf{Q_H}$ | 0.29 ± 0.36 | 0.23 ± 0.38 | 0.54 ± 0.23 | 0.47 ± 0.23 | **0.50** ± 0.24 | 0.47 ± 0.25 | 0.49 ± 0.33 | 0.40 ± 0.30 |
| $\mathbf{Q_L}$ | 0.20 ± 0.18 | 0.15 ± 0.22 | 0.55 ± 0.14 | 0.48 ± 0.19 | 0.55 ± 0.17 | 0.59 ± 0.12 | 0.53 ± 0.14 | **0.61** ± 0.10 |
| $\mathbf{Q_M}$ | **0.86** ± 0.03 | **0.86** ± 0.03 | 0.71 ± 0.08 | 0.73 ± 0.07 | 0.69 ± 0.12 | 0.76 ± 0.05 | 0.75 ± 0.07 | 0.74 ± 0.07 |
| | | | | **NNSE** (%) | | | | |
| **T** | 58 ± 25 | 54 ± 24 | 56 ± 18 | 46 ± 11 | 55 ± 18 | **62** ± 29 | 51 ± 11 | **62** ± 29 |
| **RH** | 22 ± 10 | 24 ± 11 | 47 ± 24 | 30 ± 23 | 48 ± 23 | **48** ± 24 | 38 ± 16 | 44 ± 23 |
| **P** | **57** ± 11 | 52 ± 19 | 41 ± 20 | 36 ± 18 | 43 ± 18 | 45 ± 16 | 41 ± 15 | 46 ± 20 |
| **U** | 17 ± 9 | 14 ± 6 | 27 ± 7 | 23 ± 9 | **28** ± 9 | 24 ± 7 | 23 ± 10 | 23 ± 7 |
| $\mathbf{K_{in}}$ | **69** ± 9 | 68 ± 10 | 49 ± 6 | 44 ± 20 | 49 ± 13 | 43 ± 4 | 45 ± 18 | 47 ± 10 |
| $\mathbf{L_{in}}$ | **65** ± 13 | **65** ± 14 | 41 ± 10 | 44 ± 10 | 39 ± 10 | 39 ± 10 | 45 ± 10 | 41 ± 11 |
| $\mathbf{Q_H}$ | 30 ± 12 | 27 ± 9 | 39 ± 10 | 33 ± 9 | **41** ± 10 | 38 ± 11 | 33 ± 10 | 37 ± 10 |
| $\mathbf{Q_L}$ | 42 ± 8 | 45 ± 10 | 44 ± 7 | 47 ± 8 | 41 ± 4 | **50** ± 7 | **50** ± 7 | 47 ± 5 |
| $\mathbf{Q_M}$ | **74** ± 10 | 71 ± 9 | 62 ± 6 | 53 ± 10 | 56 ± 5 | 67 ± 5 | 54± 11 | 59 ± 6 |

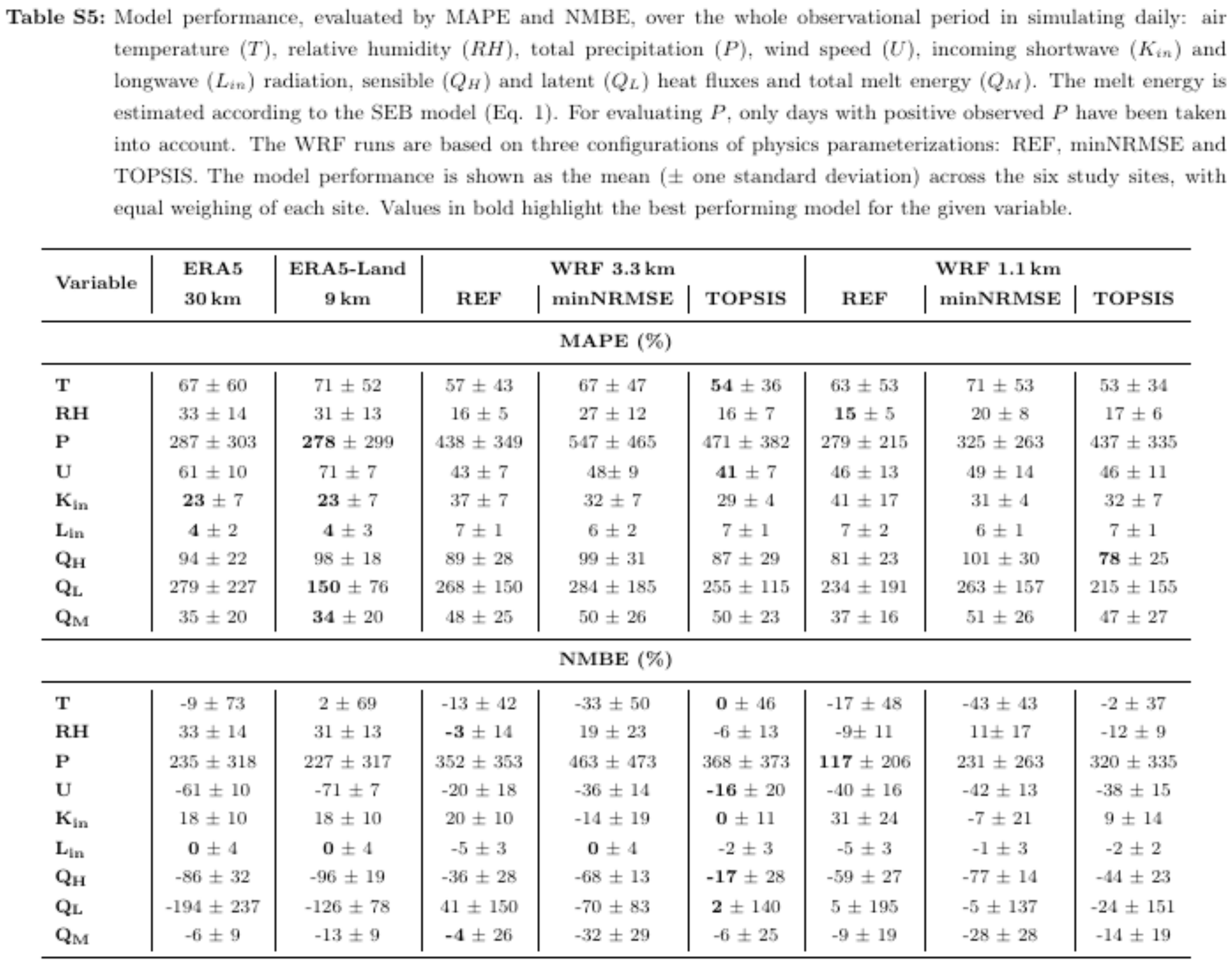


**Table S5:** Model performance, evaluated by MAPE and NMBE, over the whole observational period in simulating daily: air temperature ($T$), relative humidity ($RH$), total precipitation ($P$), wind speed ($U$), incoming shortwave ($K_{in}$) and longwave ($L_{in}$) radiation, sensible ($Q_H$) and latent ($Q_L$) heat fluxes and total melt energy ($Q_M$). The melt energy is estimated according to the SEB model (Eq. 1). For evaluating $P$, only days with positive observed $P$ have been taken into account. The WRF runs are based on three configurations of physics parameterizations: REF, minNRMSE and TOPSIS. The model performance is shown as the mean (± one standard deviation) across the six study sites, with equal weighing of each site. Values in bold highlight the best performing model for the given variable.

| **Variable** | **ERA5** | **ERA5-Land** | **WRF 3.3 km** | | | **WRF 1.1 km** | | |
|---|---|---|---|---|---|---|---|---|
| | **30 km** | **9 km** | **REF** | **minNRMSE** | **TOPSIS** | **REF** | **minNRMSE** | **TOPSIS** |
| **MAPE (%)** | | | | | | | | |
| **T** | 67 ± 60 | 71 ± 52 | 57 ± 43 | 67 ± 47 | **54** ± 36 | 63 ± 53 | 71 ± 53 | 53 ± 34 |
| **RH** | 33 ± 14 | 31 ± 13 | 16 ± 5 | 27 ± 12 | 16 ± 7 | **15** ± 5 | 20 ± 8 | 17 ± 6 |
| **P** | 287 ± 303 | **278** ± 299 | 438 ± 349 | 547 ± 465 | 471 ± 382 | 279 ± 215 | 325 ± 263 | 437 ± 335 |
| **U** | 61 ± 10 | 71 ± 7 | 43 ± 7 | 48± 9 | **41** ± 7 | 46 ± 13 | 49 ± 14 | 46 ± 11 |
| $\mathbf{K_{in}}$ | **23** ± 7 | **23** ± 7 | 37 ± 7 | 32 ± 7 | 29 ± 4 | 41 ± 17 | 31 ± 4 | 32 ± 7 |
| $\mathbf{L_{in}}$ | **4** ± 2 | **4** ± 3 | 7 ± 1 | 6 ± 2 | 7 ± 1 | 7 ± 2 | 6 ± 1 | 7 ± 1 |
| $\mathbf{Q_H}$ | 94 ± 22 | 98 ± 18 | 89 ± 28 | 99 ± 31 | 87 ± 29 | 81 ± 23 | 101 ± 30 | **78** ± 25 |
| $\mathbf{Q_L}$ | 279 ± 227 | **150** ± 76 | 268 ± 150 | 284 ± 185 | 255 ± 115 | 234 ± 191 | 263 ± 157 | 215 ± 155 |
| $\mathbf{Q_M}$ | 35 ± 20 | **34** ± 20 | 48 ± 25 | 50 ± 26 | 50 ± 23 | 37 ± 16 | 51 ± 26 | 47 ± 27 |
| **NMBE (%)** | | | | | | | | |
| **T** | -9 ± 73 | 2 ± 69 | -13 ± 42 | -33 ± 50 | **0** ± 46 | -17 ± 48 | -43 ± 43 | -2 ± 37 |
| **RH** | 33 ± 14 | 31 ± 13 | **-3** ± 14 | 19 ± 23 | -6 ± 13 | -9± 11 | 11± 17 | -12 ± 9 |
| **P** | 235 ± 318 | 227 ± 317 | 352 ± 353 | 463 ± 473 | 368 ± 373 | **117** ± 206 | 231 ± 263 | 320 ± 335 |
| **U** | -61 ± 10 | -71 ± 7 | -20 ± 18 | -36 ± 14 | **-16** ± 20 | -40 ± 16 | -42 ± 13 | -38 ± 15 |
| $\mathbf{K_{in}}$ | 18 ± 10 | 18 ± 10 | 20 ± 10 | -14 ± 19 | **0** ± 11 | 31 ± 24 | -7 ± 21 | 9 ± 14 |
| $\mathbf{L_{in}}$ | **0** ± 4 | **0** ± 4 | -5 ± 3 | **0** ± 4 | -2 ± 3 | -5 ± 3 | -1 ± 3 | -2 ± 2 |
| $\mathbf{Q_H}$ | -86 ± 32 | -96 ± 19 | -36 ± 28 | -68 ± 13 | **-17** ± 28 | -59 ± 27 | -77 ± 14 | -44 ± 23 |
| $\mathbf{Q_L}$ | -194 ± 237 | -126 ± 78 | 41 ± 150 | -70 ± 83 | **2** ± 140 | 5 ± 195 | -5 ± 137 | -24 ± 151 |
| $\mathbf{Q_M}$ | -6 ± 9 | -13 ± 9 | **-4** ± 26 | -32 ± 29 | -6 ± 25 | -9 ± 19 | -28 ± 28 | -14 ± 19 |

## Minor Comments

*R2 C2:* Page 1, line 22: I suggest rephrasing to "…and are increasingly losing a considerable amount of mass ….."

Done.

*R2 C3:* Page 2, line 39: Why does SEB models not require precipitation?

Please see our comment to R2 C1 regarding the importance of rain heat flux in seasonal SEB. We now better clarify why precipitation is not considered in the SEB model.

*R2 C4:* Page2, line 43: I suggest rephrasing to " …fewer than 100 sites worldwide, and only a handful in Western Canada…."

Done.

*R2 C5:* Page 3, section starting on line 62. Here I suggest including a citation of the recent work by Eidhammer et al 2021 (https://hess.copernicus.org/articles/25/4275/2021/), where they use the detailed snow model Crocus within the WRF-Hydro model to estimate glacier melt (and streamflow). They used 1 km downscaled WRF simulations over a glacier in Norway for four seasons.

This reference is now incorporated in the text (lines 83-86):

*More recently, Eidhammer et al. (2021) used WRF downscaling to 1 km grid spacing coupled with snowpack modeling through the WRF-Hydro model (Gochis et al., 2020), showing a good agreement between the WRF output and in-situ meteorological observations at a glacier in Norway over four years.*

*R2 C6:* Page 3: In the discussion of using dynamical downscaling, you might want to add a comment related to the paper by Lundquist et al. 2019 with the title: “Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks”. https://doi.org/10.1175/BAMS-D-19-0001.1. I think that this can add to the argument in this manuscript to use downscaling for SEB modeling.

This reference is now incorporated in the text (lines 56-60):

*An alternative to statistical is dynamical downscaling: a physics-based approach that utilizes a regional climate model (RCM), nested within a reanalysis or global climate model, to compute meteorological fields at a desirable spatial resolution, often <10 km. A well-configured high-resolution RCM outperforms radar and satellite-derived estimates of total annual rain and snowfall within mountainous regions (Lundquist et al., 2019).*

*R2 C7:* Page 3, line 22. I suggest adding a citation to the paper by Liu et al 2011 “High-Resolution Simulations of Wintertime Precipitation in the Colorado Headwaters Region: Sensitivity to Physics Parameterizations“ (https://doi.org/10.1175/MWR-D-11-00009.1). They tested several different WRF physics parameterizations over the Colorado headwaters region.

This reference is now incorporated in the text (lines 93-95):

*A relatively underexplored limitation in using WRF in glacier studies is the model’s potentially large sensitivity to the choice of physics parameterization schemes as noted in many non-glacier studies (e.g., Liu et al., 2011; Zeyaeyan et al., 2017; Gbode et al., 2019; Pervin and Gan, 2020; Shirai et al., 2022).*

*R2 C8:* Page 3, line 94. The Thompson-Eidhammer scheme (https://doi.org/10.1175/JAS-D-13- 0305.1) has also been used for Glacier studies

We have incorporated the Thompson-Eidhammer scheme in the text (lines 210-215). We now moved the description of previously used parameterization schemes to the “Data and Methods” section as per suggestion of referee #1 (R1 C7).

*Lines 210-215: For example, the most commonly used schemes in glaicer studies include RRTMG (Iacono et al., 2008), CAM (Collins et al., 2004), Dhudia (Dudhia, 1989) and Goddard (Max and Suarez, 1994; Matsui et al., 2018) for radiation, the Grell 3D Ensemble (Grell, 1993; Grell and Dévényi, 2002), the Kain-Fritsch (Kain, 2004) and the Betts-Miller-Janjić (Janjić, 1994) schemes for cumulus convection, and the Morrison two-moment (Morrison et al., 2009), the Thompson (Thompson et al., 2008) and the updated aerosol–aware Thompson-Eidhammer (Thompson and Eidhammer, 2014) schemes for microphysics.*

*R2 C9:* Page 5, Table 5, caption: What is meant by "full days"?

The table caption is rephrased to:

*Table. 1. Characteristics of the study sites. Only days with 24-hour observations have been taken into account for the observational periods.*

*R2 C10:* Page 5, line 147: I suggest rephrasing: "…the accumulation zone of the Conrad glacier in 2016."

Done.

*R2 C11:* Page 6, line 163: The way I read the sentence, the reference to Table 1 indicates that there is some information in regards to the melting surface with intermittent fresh snowfall in the Table 1. I do not think the reference to Table 1 is necessary here.

The reference to Table 1 is now removed in this line and the following lines (159-171).

*R2 C12:* Page 8, lines 2 and 3: I suggest to clarify that both 3D and 2D ERA 5 fields (I assume some 2D fields are used at initialization) are used as forcing data for the WRF model.

We revised the text (lines 179-180):

*Hourly two- and three-dimensional ERA5 reanalysis data is also used to provide initial and lateral boundary conditions to the WRF model.*

*R2 C13:* Page 8, line 193. The way I read this line, the d1 domain for all the 4 glaciers are the same. However, Figure 3 shows that d1 is different between Kaskawulsh and the other glaciers. Can you please clarify?

The domain d1 does differ between Kaskawulsh and other glaciers. Thank you for spotting this discrepancy. We revised the sentence (lines 186-190):

*We ran the WRF model, version 4.1.3, configured with four nested domains of 30 km (d1), 10 km (d2), 3.3 km (d3) and 1.1 km (d4) horizontal grid spacing, with the parent domain (d1) covering the bulk of North America and the North-East section of the Pacific Ocean (Figure 3). The domains d1 and d2 are kept the same for the three glaciers in the interior of British Columbia (Castle Creek glacier, Nordic glacier and Conrad glacier), while d3 and d4 are set differently for each of the three glaciers in order to be centered at the AWS location.*

*R2 C14:* Page 8, line 198. I assume that you mean that many physics variables are updated every 2.2 s, not outputted? And most likely, the radiation and land surface variables are probably not calculated every 2.2 seconds, but perhaps somewhere between every 5 or 30 minutes? Also make sure if the hourly outputs indeed are hourly averages. Typically, most of the WRF outputs are instantaneously outputs, with some of them being accumulated.

Yes, there are 30 minutes between radiation physics calls in our model, as recommended for our grid spacing of 30 km for the outer domain. WRF updates many physics variables every 2.2 s in the inner-most domain and, as we use WRF "tslist" (time series) output, we get WRF output for every time step (i.e., 2.2 s for the inner-most domain). We average this time series output to hourly and daily averages. The text is revised accordingly (lines 194-195):

*We use a time step of 2.2 s for the most inner domain and save the selected set of variables as hourly and daily averages.*

*R2 C15:* Page 8, line 199: Table 2 does not describe any of the output saved. I suggest remove the reference to Table 2.

We removed the reference to Table 2.

*R2 C16:* Page 9, line 208. I wonder if it would be helpful to add a delta elevation from AWS in table S1 as well. In this way, it would be easier to see the actual elevation difference.

The delta elevation for each of the input datasets (ERA5, ERA5-Land, WRF at 3.3 km, WRf at 1.1 km, SRTM/ASTER) in comparison to the AWS elevation is now added in Table S1.

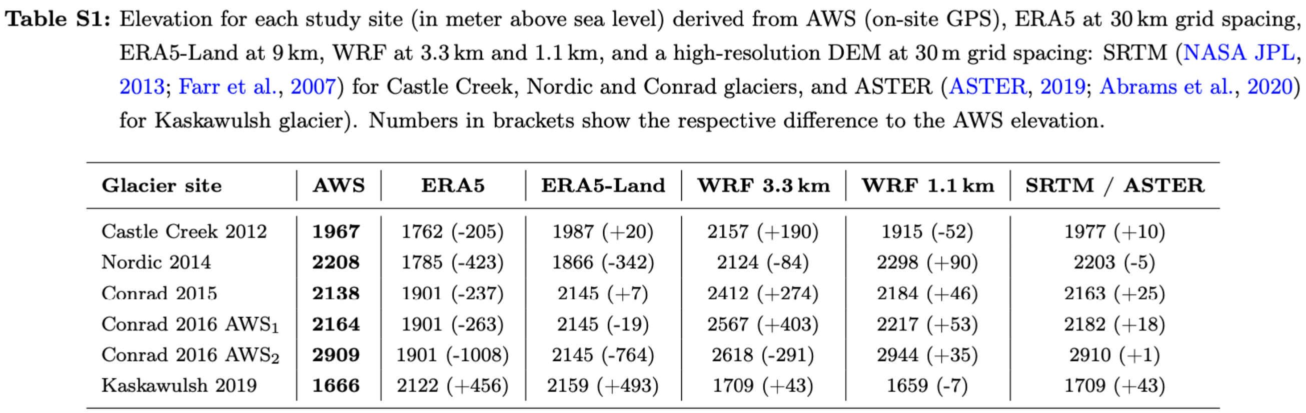


**Table S1:** Elevation for each study site (in meter above sea level) derived from AWS (on-site GPS), ERA5 at 30 km grid spacing, ERA5-Land at 9 km, WRF at 3.3 km and 1.1 km, and a high-resolution DEM at 30 m grid spacing: SRTM (NASA JPL, 2013; Farr et al., 2007) for Castle Creek, Nordic and Conrad glaciers, and ASTER (ASTER, 2019; Abrams et al., 2020) for Kaskawulsh glacier). Numbers in brackets show the respective difference to the AWS elevation.

| **Glacier site** | **AWS** | **ERA5** | **ERA5-Land** | **WRF 3.3 km** | **WRF 1.1 km** | **SRTM / ASTER** |
|---|---|---|---|---|---|---|
| Castle Creek 2012 | **1967** | 1762 (-205) | 1987 (+20) | 2157 (+190) | 1915 (-52) | 1977 (+10) |
| Nordic 2014 | **2208** | 1785 (-423) | 1866 (-342) | 2124 (-84) | 2298 (+90) | 2203 (-5) |
| Conrad 2015 | **2138** | 1901 (-237) | 2145 (+7) | 2412 (+274) | 2184 (+46) | 2163 (+25) |
| Conrad 2016 $AWS_1$ | **2164** | 1901 (-263) | 2145 (-19) | 2567 (+403) | 2217 (+53) | 2182 (+18) |
| Conrad 2016 $AWS_2$ | **2909** | 1901 (-1008) | 2145 (-764) | 2618 (-291) | 2944 (+35) | 2910 (+1) |
| Kaskawulsh 2019 | **1666** | 2122 (+456) | 2159 (+493) | 1709 (+43) | 1659 (-7) | 1709 (+43) |

*R2 C17:* Page 9, line 214. On page 3, it is stated that the Microphysics by Morrison is the most commonly used in glacier studies, but in this work, the Thompson microphysics is used. Please clarify this discrepancy. Also see page 27, line 597

Yes, the referee is correct. Both schemes have been previously used in glacier studies. The list of physics parameterization schemes for our sensitivity analysis in Table S2 now matches with the schemes described in the text (lines 209-221). During the sensitivity analysis, we tested both the Morrison 2-moment and the Thompson schemes and found that the Thompson scheme showed better performance in overall melt energy at our glacier sites.

*Lines 209-221: The WRF model comes with various options for physics parameterizations (Skamarock et al., 2019), but previous glacier studies with WRF have used some parameterization schemes more often than others. For example, the most commonly used schemes in glaicer studies include RRTMG (Iacono et al., 2008), CAM (Collins et al., 2004), Dhudia (Dudhia, 1989) and Goddard (Max and Suarez, 1994; Matsui et al., 2018) for radiation, the Grell 3D Ensemble (Grell, 1993; Grell and Dévényi, 2002), the Kain-Fritsch (Kain, 2004) and the Betts-Miller-Janjić (Janjić, 1994) schemes for cumulus convection, and the Morrison two-moment (Morrison et al., 2009), the Thompson (Thompson et al., 2008) and the updated aerosol–aware Thompson-Eidhammer (Thompson and Eidhammer, 2014) schemes for microphysics. The local-closure Mellor–Yamada–Nakanishi–Niino (MYNN) level 2.5 (Nakanishi and Niino,* *2006, 2009; Olson et al., 2019) and Mellor-Yamada-Janjic (Janjić, 1994; Mesinger, 1993) schemes, as well as the non-local closure Yonsei University (Hong et al., 2006) scheme have been most commonly used for boundary layer, and the revised MM5 (Jiménez et al., 2012) and Eta Similarity (Monin and Obukhov, 1954; Janjić, 1994, 1996, 2002) schemes for surface layer. The Noah (Tewari et al., 2004) and Noah-MP (Niu et al., 2011; Yang et al., 2011) land surface models are most commonly used in glacier studies, but the WRF simulations over non-glacierized terrain are shown to vary substantially depending on which of the two land surface models is used (Milovac et al., 2016).*

*R2 C18:* Page 13, line 277: It is stated that the goal is to evaluate daily timeseries of simulated energy available for melt. As shown in Fitzpatrick 2017, the QR was shown to have considerable influence on SEB when considering daily and sub-daily timescales. I am wondering if ignoring QR in this study is then valid?

We have revised the sentence (lines 291-296) to align more consistently with the stated objectives in the Introduction. Our primary focus is to assess the performance of ERA5 and WRF in simulating seasonal melt energy for long-term glacier melt simulations. Additionally, we explore the daily-scale performance by analyzing the timeseries of daily fluxes rather than solely relying on net (average) fluxes across the observational period.

As mentioned earlier in comment R2 C1, over a melting season, rainfall can significantly contribute (up to 20%) to daily melt energy, as demonstrated by Fitzpatrick et al. (2017). However, the uncertainty in modeling rain heat flux is relatively large and might potentially be grounded in unsupported assumptions (Hock, 2005; Fitzpatrick et al., 2017). Despite these significant contributions on daily scales, the low occurrence and duration of these extreme precipitation events does lead to a contribution of <1% to the total melt energy over the melting season, as shown by Fitzpatrick et al. (2017).

We chose not to include the modeled rain heat flux in our SEB model, as we are focused only on the key contributors to melt energy on seasonal scales. Nonetheless, we have now included a comparison between modeled (WRF/ERA5) and observed precipitation at our sites, considering the relevance of fresh snowfall events for albedo (refer to lines 355-362 and 374-379, as well as Figures S2 and S3; see R2 C1).

*R2 C19:* Page 13, line 286. The study by Eidhammer et al. 2021 also shows that the albedo, when using Noah-MP does not perform well over glaciers (especially in the ablation region).

Thank you for pointing out this study. The reference is now incorporated (lines 303-304):

*The discrepancy between WRF and observed albedo at glaciers, especially in the ablation zone, has also been noted previously in Eidhammer et al. (2021).*

*R2 C20:* Page 18, line 371: Please specify that this is the WRF REF data.

Done.

*R2 C21:* Page 19, section 3.3 I think that this section should come before section 3.2, since you are using the results from section 3.3 in describing results in section 3.2.

Done.

*R2 C22:* Page 19, line 404: Level 3, not Lavel 3

Done.

*R2 C23:* Page 23, line 490: Did you consider employing different roughness lengths over snow versus ice?

Roughness lengths do differ between snow and ice surfaces, and for most of our sites they are different by two orders of magnitude ($z_{0v} \sim 10^{-3}\mathrm{m}\ \ z_{0T} \sim z_{0q} \sim 10^{-5}\mathrm{m}$). We now make this more clear in the text (lines 283-290, Table S3):

*We use constant roughness lengths for each site, which have been adopted from the EC-derived turbulent fluxes in previous studies at these glacier sites (Table S3; Radić et al., 2017; Fitzpatrick et al., 2017, 2019; Lord-May and Radić, 2023), while for the stability corrections, based on the assessed stability conditions, we use the functions applied previously in Fitzpatrick et al. (2017). The order of magnitude in these EC-derived values for roughness lengths* ($z_{0v} = 10^{-3}\mathrm{m}\ \ z_{0T} = z_{0q} = 10^{-5}\mathrm{m}$)*.) agree with commonly assumed values for glaciers in mid-latitudes (Hock, 2005). Vapor pressure* $e_z$ *at height z is calculated from the relative humidity RH at height z using the August-Roche-Magnus formula (Alduchov and Eskridge, 1997). For more details on the bulk method and the stability corrections used in the study, readers are referred to Fitzpatrick et al. (2017).*