DCG-MIP: The Debris-Covered Glacier melt Model Intercomparison exPeriment

Francesca Pellicciotti*[1,2#], Adrià Fontrodona-Bach*[1,2#], David R. Rounce[3], Catriona L. Fyffe[1,4#], Leif S. Anderson[5], Álvaro Ayala[6], Ben W. Brock[4], Pascal Buri[2#,7], Stefan Fugger[2#], Koji Fujita[8], Prateek Gantayat[1,9#], Alexander R. Groos[10,11], Walter Immerzeel[12], Marin Kneib[2,13], Christoph Mayer[14], Shelley MacDonell[6,15], Michael McCarthy[1,2], James McPhee[16,17], Evan Miles[2,18,19], Heather Purdie[20], Ekaterina Rets[21], Akiko Sakai[8], Thomas E. Shaw[1], Jakob Steiner[12,22,23], Patrick Wagnon[24], Alex Winter-Billington[25]

*These authors contributed equally to this work.
#Formerly at
[1]Institute of Science and Technology Austria, Klosterneuburg, Austria;
[2]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
[3]Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[4]School of Geography and Natural Sciences, Northumbria University, Newcastle-Upon-Tyne, UK
[5]Department of Geology and Geophysics, University of Utah, Salt Lake City, UT, USA
[6]Centro de Estudios Avanzados en Zonas Áridas (CEAZA), La Serena, Chile
[7]Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA
[8]Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan
[9]Lancaster Environment Center, Lancaster University, Bailrigg, Lancaster, UK
[10]Institute of Geography, University of Bern, Bern, Switzerland
[11]Institute of Geography, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[12]Department of Physical Geography, Utrecht University, Utrecht, The Netherlands
[13]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland
[14]Geodesy and Glaciology, Bavarian Academy of Sciences and Humanities, Munich, Germany
[15]Waterways Centre, University of Canterbury, Christchurch, New Zealand
[16]Department of Civil Engineering, University of Chile, Santiago, Chile
[17]Advanced Mining Technology Center, University of Chile, Santiago, Chile
[18]Glaciology and Geomorphodynamics Group, University of Zurich, Zurich, Switzerland
[19]Department of Geosciences, University of Fribourg, Fribourg, Switzerland
[20]School of Earth & Environment, University of Canterbury, Christchurch, New Zealand
[21]Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland
[22]Himalayan University Consortium, Lalitpur, Nepal
[23]Institute of Geography and Regional Science, University of Graz, Graz, Austria
[24]Univ. Grenoble Alpes, CNRS, IRD, IGE, Grenoble, France
[25]Te Puna Pātiotio Antarctic Research Centre, Te Herenga Waka Victoria University of Wellington, New Zealand

Correspondence to: Francesca Pellicciotti (francesca.pellicciotti@ista.ac.at) and Adrià Fontrodona-Bach (adria.fontrodona-bach@ista.ac.at)

Abstract. In a warming world of glacier changes, the scientific community has dedicated increasing attention to debris-covered glaciers and their response to climate. A variety of models with distinct complexity and data requirements have been developed and widely used to simulate melt under debris at different sites and scales, but their skills have never been compared. As part of the activities of the International Association of Cryospheric Science (IACS) Debris Covered Glacier Working Group, we present an intercomparison exercise aimed at advancing our understanding of model skills in simulating ice melt under a debris layer. We compare 154 models with different complexity at nine sites in the European Alps, Caucasus, Chilean Andes, Nepalese Himalaya and the Southern Alps of New Zealand, over one melt season. We run the models with measured meteorological data from automatic weather stations and estimated or measured debris properties. We consider four main model categories: i) energy balance models that calculate melt by solving the physics of heat transfer to the debris layer, but require a high amount of input data; ii) a simplified energy balance model; iii) ~~an~~ enhanced temperature-index models; and iv) simple empirical temperature-index models that have been extensively used given their low data requirement but require

calibration of their empirical parameters. Model performance is evaluated using on-site measurements of sub-debris melt (for all models) and surface temperature (for models based on the surface energy balance). Our results show that physically-based energy balance models and empirical temperature-index models perform in a distinct manner. At the one end of the spectrum, simple temperature index models are accurate when recalibrated or when using site-specific literature parameters, and show poor results when parameters are uncalibrated. At the other end, energy balance models show a range of performance: the most accurate energy balance models are those with the highest degree of complexity at the atmosphere-debris interface. An important data gap emerged from our experiment: the poor performance of all models at three sites was related to the poor knowledge of debris properties, and specifically of thermal conductivity. Future work should focus on both: i) consistent data acquisition to evaluate existing models and support new model developments; ii) advancing models by accounting for processes such as debris-snow interactions, moisture in the debris and refreezing. We suggest that a systematic effort of model development using a common model framework could be carried out in phase II of the Working Group.

1. Introduction

Glacier ice is often covered by a continuous or discontinuous layer of rock debris, which can vary in thickness from a few centimetres to several metres (Østrem, 1959; Kirkbride and Dugmore, 2003; Reid et al., 2012; Juen et al., 2014; Rounce and McKinney, 2014; Fyffe et al., 2020). Such debris-covered ice is extensive in many mountain ranges around the world (Scherler et al., 2018, Herreid and Pellicciotti, 2020). In a warming climate, debris cover has been observed to increase in area and thickness (Deline, 2005; Stokes et al., 2007; Bhambri et al., 2011; Thakuri et al., 2014; Mölg et al., 2019; Tielidze et al., 2020; Xie et al., 2020; Anderson et al., 2021) as a result of melt-out and accumulation of englacial debris at the glacier surface (Kirkbride and Deline, 2013; Anderson and Anderson, 2018) as well as increased debris input from surrounding slopes and lateral moraines destabilised by glacier debuttressing (van Woerkom et al. 2019) and permafrost degradation (Gruber et al., 2017). During sustained periods of negative glacier mass balance, debris cover expands laterally from medial moraines and upstream as debris-rich ice is brought to the surface (Anderson, 2000; Jouvet et al., 2011; Rowan et al., 2015).

The role of supraglacial debris in modulating glacier response to climate across scales is an open topic of research. We broadly understand debris to reduce melt rates when thicker than a few centimetres (Østrem, 1959; 1965; Kirkbride and Dugmore, 2003) and to potentially increase melt rates when thinner (Østrem, 1959; Mattson et al., 1993) or patchy (Fyffe et al., 2020). The relationship between debris thickness and sub-debris melt is commonly referred to as an Østrem curve, which has ~~has~~ been established through field observations (Østrem, 1959; 1965; Khan, 1989; Mattson et al., 1993; Konovalov, 2000; Popovnin and Rozova, 2002; Lukas et al., 2005; Mihalcea et al., 2006; Nicholson and Benn, 2006; Hagg et al., 2008) and numerical simulations with energy balance models at the point scale (Reid and Brock, 2010; Wang et al., 2011; Brook et al., 2013; Lejeune et al., 2013; Evatt et al., 2015). Studies that document melt across debris-covered glacier surfaces beyond the point scale are scarcer (Reid et al., 2012; Vincent et al., 2016; Anderson et al., 2021; Steiner et al., 2021) and our understanding of glacier-scale ablation patterns is more limited.

Research on debris-covered glaciers has seen an enormous growth in the last decade. Novel lines of research include the first global mapping efforts of debris areal extent (Scherler et al., 2018; Herreid and Pellicciotti, 2020); determining the thickness of debris covering glaciers at local and regional scales (Schauwecker et al., 2015; Groos et al., 2017; McCarthy et al., 2017, 2022; Nicholson et al., 2018; Rounce et al., 2018; Rounce et al., 2021); understanding how debris is transported through the ice and affects glacier flow and geometry (Rowan et al., 2015; Anderson and Anderson, 2016; Banerjee, 2017; Wirbel et al., 2018; Scherler and Egholm, 2020; Kirkbride et al., 2023; Margirier et al., 2025); identifying the distinct large scale thinning patterns of debris-covered glaciers as compared to debris-free glaciers (Kääb et al., 2012, Brun et al., 2019); advancing our understanding of debris-covered glacier meteorology (Brock et al., 2010; Shaw et al., 2016; Steiner and Pellicciotti, 2016; Yang et al., 2017; Steiner et al., 2018; Bonekamp et al., 2020; Nicholson and Stiperski, 2020), surface properties (Nicholson

and Benn, 2013; Rounce et al., 2015; Miles et al., 2017; Quincey et al., 2017) and hydrology (Fyffe et al., 2020; Miles et al., 2020); and insights into the processes controlling debris-covered glacier mass balance and the role that surface features such as ice cliffs and ponds play in amplifying mass balance locally and at the glacier scale (Sakai et al., 2000, 2002; Han et al., 2010; Immerzeel et al., 2014; Reid and Brock, 2014; Buri et al., 2016a,2016b,2018; Thompson et al., 2016; Salerno et al., 2017; Miles et al., 2016, 2018; Brun et al., 2016, 2018; Watson et al., 2018; Mölg et al., 2019; Anderson et al., 2021).

Some of these new lines of research have exploited satellite observations of increasing resolution (Brun et al., 2018) as well as surveys from Uncrewed Aerial Vehicles (Immerzeel et al., 2014; Kraaijenbrink et al., 2016, 2018; Fyffe et al., 2020; Westoby et al., 2020; Bisset et al., 2022; Messmer and Groos, 2024), which allow processes of glacier mass loss, debris evolution and dynamics to be understood at high resolution. Others have focused on model developments (e.g. Buri and Pellicciotti, 2018; Potter et al., 2020) and new theoretical advances (e.g. Nicholson and Stiperski, 2020). Despite these tremendous advances, some of the basic aspects of debris-covered glacier processes and modelling remain elusive (e.g. debris sourcing and evolution over scales, numerical reconstruction of debris thickness across spatial and temporal scales, the future trajectory of debris covered glaciers at local and global scales). Numerous models have emerged to represent some aspects of this complexity (e.g. Buri et al., 2016a,b; Rowan et al., 2015; Wirbel et al., 2018) but in the case of ablation of ice covered by debris, our understanding of key processes is still lacking.

Models developed to simulate the ablation of debris-covered ice can be broadly grouped into physically-based energy balance models and empirical temperature-index models, with a number of intermediate models between the two categories. Energy balance models estimate the energy fluxes at the interface between the debris and atmosphere, within the debris and at the interface between the debris and ice. As a result, they are able to explain the physical processes causing melt. These models have been primarily applied at the point scale using data from on-site automatic weather stations (Nicholson and Benn, 2006; Reid and Brock, 2010; Lejeune et al., 2013; Rounce et al., 2015; Giese et al., 2020), where they can be forced with meteorological data measured within the glacier boundary layer. Energy balance models have also been applied using off-glacier and re-analysis data products (e.g. Rounce et al., 2015). While energy balance models are physically realistic relative to temperature index models, they require more input meteorological data, as well as knowledge of debris properties and physical parameters used to calculate the main energy fluxes. These debris and atmospheric parameters (such as debris thermal conductivity, debris porosity, debris albedo, surface roughness length and heat transfer coefficients) are difficult to constrain spatially and temporally even for individual glaciers and short (sub-annual) periods, especially at remote sites outside of the European Alps, as ~~where~~ most previous research has been carried out in the European Alps (e.g. Nicholson et al., 2006; Brock et al., 2010). Energy balance models are thus less commonly applied at the glacier scale (e.g. Fyffe et al., 2014; Reid et al., 2012; Groos et al., 2017; Shaw et al., 2016). It is also accepted by now that sophisticated models forced with low-quality input data will produce poor simulations (Machguth et al., 2008; Anslow et al., 2008; MacDougall and Flowers, 2011; Gabbi et al., 2015; Shaw et al., 2016).

On the other side of the spectrum are temperature-index (or degree-day) models. These models, initially developed for clean ice, assume a linear relationship between the melt rate and air temperature above a given temperature threshold, typically near 0°C, such that the melt can be estimated using a multiplicative factor called the degree-day factor (Hock, 2003). Because most debris-covered areas are mantled in relatively thick debris which reduces ablation, the most common approach to account for debris in these models has been to reduce the degree-day factor, thereby reducing the melt rates (e.g. Immerzeel et al., 2012, 2013; Shea et al., 2015). However, complexity can be added by including parameterizations for other factors such as the radiation component (Carenzo et al., 2016). Degree-day factors for different debris thicknesses have been calculated from sub-debris melt rates and air temperature measurements (e.g. Kayastha et al., 2000; Mihalcea et al., 2006; Hagg et al., 2008; Wei et al., 2010; Brook et al., 2013; Juen et al., 2014), but knowledge of their spatial variation remains a challenge that limits this approach. Constraining the variation of degree-day factors in space and in time is an area of active research, which has

generated a number of variants of this approach (Anderson and Anderson, 2016; Carenzo et al., 2016; Winter-Billington et al., 2020). Degree-day factors cannot be measured directly in the field, and rely on calibration with in-situ measurements, challenging their transferability to other sites. Despite this, temperature-index models have seen successful applications at the glacier and regional scale because they are simple, computationally efficient and require only air temperature (occasionally incoming shortwave radiation) as input to model melt and a low number of parameters (e.g., Kraaijenbrink et al., 2017). In most cases temperature-index models are applied at daily or coarser temporal resolution.

Different types of models respond to distinct needs, data availability and purposes. Often, numerical model development has balanced complexity with applicability. Energy balance models provide an accurate representation of the complex physical processes driving melt under debris at the expense of high data requirements, while temperature-index models provide a wider applicability at the expense of simplicity in process representation. Crucially, modelling skills, model structures and approaches have rarely been systematically evaluated across a range of sites for debris-covered glaciers.

Given the growing recognition that debris cover plays an important role in glacier mass balance and evolution, a Debris Covered Glaciers Working Group was established within the International Association of Cryospheric Sciences (IACS) to foster knowledge sharing and address key knowledge gaps within the community. In this context, we have designed and carried out a model experiment to compare several types of models with different degrees of complexity for modelling melt under debris. Our motivation is to understand whether models of differing complexity agree, under which conditions each of them performs well, and to identify areas where model development is needed. The comparison is carried out at the point scale of automatic weather stations installed on nine debris-covered glaciers using 154 models that cover a wide spectrum of model structure and complexity, and we compare them in a systematic manner at all sites. Our specific objectives are:

1. to assess model performance at different sites
2. to identify the strengths and limitations of the different model categories
3. to advance our understanding of the impact of model choice on the accuracy and uncertainty of simulated melt.
4. to attribute differences in model performance to model physics, assumptions and/or inaccurate data

We first present the design of the intercomparison experiment, describe the study sites and input data, and provide an overview of the models. We then present and discuss the results of the intercomparison, the implications for modelling and data collection experiments, and conclude with recommendations for future work.

2. Experimental setup

Our intercomparison was conducted as an open experiment. Nine data provider groups and 13 modelling groups responded to the call for participation. Modellers were provided with standardised meteorological data at hourly resolution, debris properties, and validation datasets that included sub-debris melt rate from ablation stakes and/or from ultrasonic depth gauge measurements, and debris surface temperature data (where available) for 9 study sites (Table 1). Some of the temperature index models used the sub-debris melt and/or surface temperature data for calibration, and the calibration strategy for individual models was left to the modellers (see Sect. 4.1 on model calibration).

At each site, we carried out two melt modelling experiments:

1. With measured debris thickness; and
2. With variable debris thickness (with values of 1, 2, 4, 6, 8, 10, 12, 15, 20, 30, 50 and 100 centimetres) to derive Østrem curves.

We quantified uncertainty for each experiment using Monte Carlo simulations where we varied the debris properties of the energy balance models and the parameters of the temperature-index models (see Sect. 4.2 on model uncertainty).

Each model was run for the period of time that data was available, which varied by site (Table 1). We restricted the model simulations to the period without continuous snow cover on the debris surface to avoid issues associated with the choices each modeller made associated with snow on top of the debris, which would affect the comparison of the sub-debris melt. Nonetheless, each modeller had to decide how to deal with short, occasional periods of snowfall (see model descriptions in the Supplement Sect. 2).

Given that data were available for one melt season only, no spin-up was possible for energy balance models. Given that no snow was present at the beginning of our experiment, that the time required to spin up within-debris temperature should be in the order of hours or a few days, and that only one model allows ice-temperatures to go below zero, we expect the lack of a spin up period to have a minimal effect on the study.

3. Data and study sites

The experiments were performed at nine study sites, which included three in the European Alps, two in High Mountain Asia, two in the Chilean Andes, one in the Caucasus and one in the Southern Alps of New Zealand (Fig. 1; Table 1). Study sites were selected based on availability of a complete set of meteorological data, debris properties and validation data. The nine glacier sites span a large range of elevations and climates, as well as debris thickness and morphologies (Fig. 2). Nevertheless, we recognise that our sites do not include some critical regions where debris is abundant: including Alaska, Greenland, Peru and the tropical Andes, Patagonia and the Western regions of High Mountain Asia such as the Karakoram (Fig. 1).

For each study site, the following data were provided:

- Automatic Weather Station (forcing) data: air temperature (°C), relative humidity (%), wind speed (m s^{-1}) and direction (°), air pressure (hPa), shortwave and longwave radiation (incoming and outgoing, W m^{-2}), precipitation (mm hr^{-1}), snow depth (cm), height of meteorological sensors (m). These data were provided at hourly resolution. Wind direction was not used by any model.
- Debris thickness (h_d) measured at the site.
- Debris properties that were measured, derived, assumed or optimised in a previous modelling exercise: surface roughness length (z_0), thermal conductivity (k_d), porosity (φ) and emissivity (ε) (Table S1).
- Validation data: surface height change measurements, from either ablation stakes or ultrasonic depth gauge readings, and debris surface temperature measurements derived from outgoing longwave radiation.
- Metadata of the site.

The Supplement provides additional data and information from the study sites: a photo from each of the nine sites (Fig. S1), the values of debris properties at each site and whether they were measured, optimised, estimated or assumed (Table S1), the uncertainty of the validation data measurements (Table S2), and a summary of mean measured meteorological data at each site (Table S3).

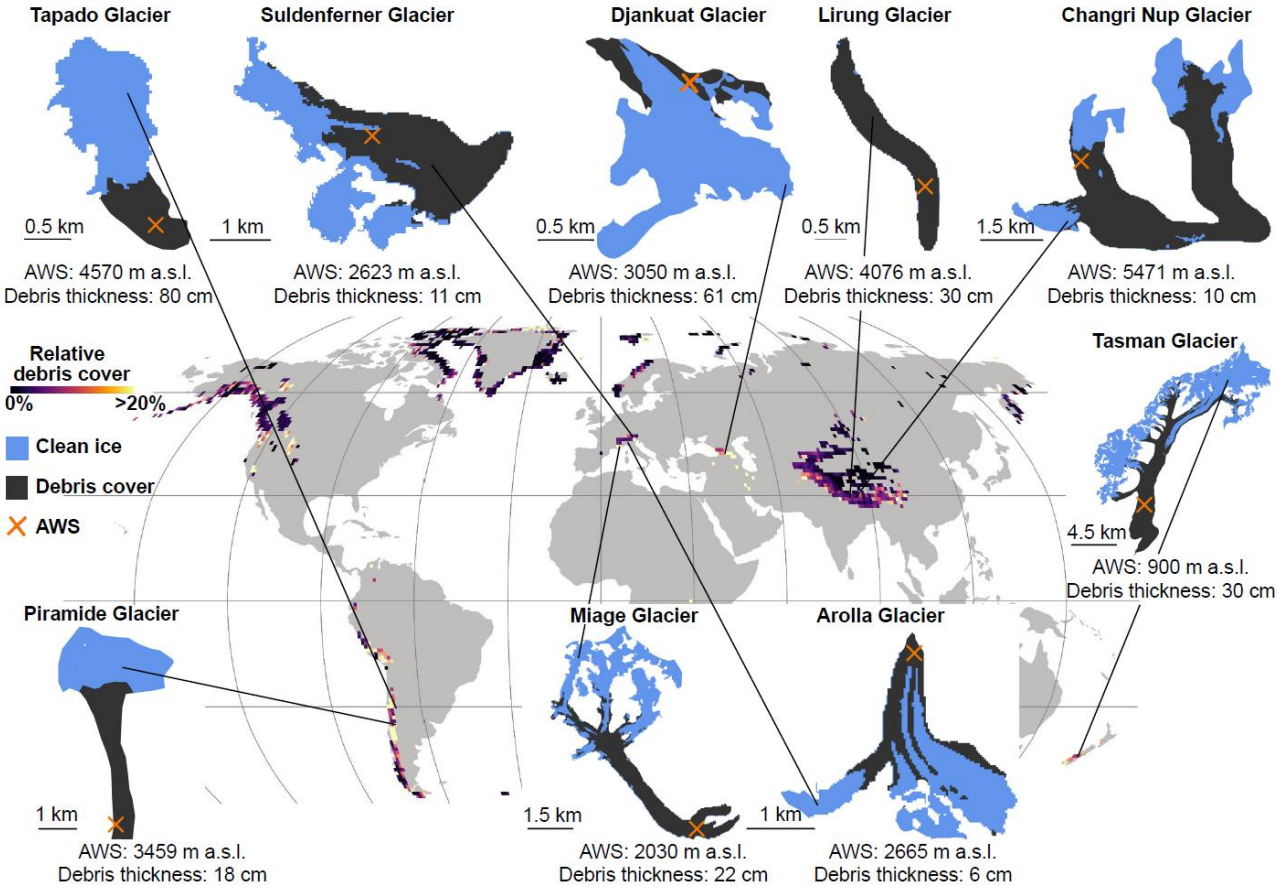
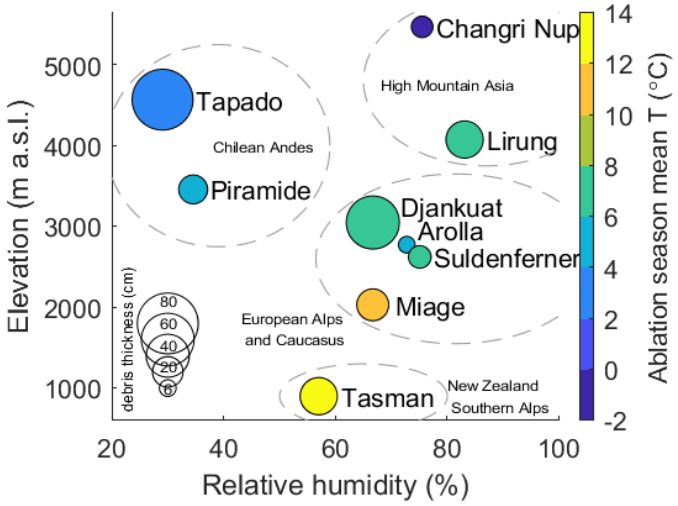


Figure 1. Location of the study sites and glacier maps with the position of the automatic weather station (indicated by a cross) on the debris-covered part (dark grey area) of each glacier (blue shade indicates clean ice). The background colours show relative debris cover per 1x1° tile, from Scherler et al. (2018).



205

Figure 2. Characteristics of the study sites. Elevation, mean relative humidity, mean air temperature and debris thickness at the nine study sites, grouped by four main geographic regions. Circle sizes denote debris thickness at the location of the respective weather station. Mean air temperature and relative humidity are calculated as the average over the simulation period of the model intercomparison. Note that periods are of different duration at different sites (see Table 1).

210

Table 1. Overview of study sites. Validation data indicates what kind of melt observations are used at each site: ultrasonic depth gauge (UDG), ablation stakes, a draw-wire, and debris surface temperature (T_s). h_d = debris thickness. The latitude and longitude coordinates refer to the locations of the automatic weather stations.

Glacier	Latitude	Longitude	Elevation (m asl)	Start date	End date	Simulation length (days)	h_d (cm)	Validation data
Arolla	45.9814	7.5244	2665	04/07/2010	12/09/2010	71	6	UDG, Stakes, T_s
Changri Nup	27.9925	86.7799	5471	16/04/2016	24/07/2016	100	10	UDG, T_s
Djankuat	42.7596	43.1997	3050	17/06/2007	07/09/2007	83	61	UDG, T_s
Lirung	28.2326	85.5621	4076	04/05/2014	24/10/2014	174	30	UDG, Stakes, T_s
Miage	45.7823	6.8763	2030	13/06/2005	07/09/2005	87	22	Stakes, T_s
Piramide	-33.5896	-69.8907	3459	21/10/2014	30/04/2015	192	18	Stakes, T_s
Suldenferner	46.4958	10.5692	2623	19/07/2016	11/09/2016	55	11	UDG, T_s
Tapado	-30.1565	-69.9224	4570	13/12/2014	22/03/2015	100	80	UDG, T_s
Tasman	-43.6286	170.2030	900	06/10/2017	09/04/2018	186	30	Draw-wire, Stakes, T_s

4. Models

~~Fourteen~~ Fifteen models are part of our intercomparison experiment. When accounting for both calibrated and uncalibrated model runs, the total number of approaches increases to seventeen. Of the 15~~4~~ models, ~~nine~~ ten were published prior to the call, three were a modification of a published model, and two were unpublished. The 15~~4~~ models span a range of model complexity from energy balance models to temperature index models including intermediate models. The ~~two~~ intermediate models are different enough to be included in two distinct categories: a model that is close to the energy balance approach (a simplified energy balance) and ~~a~~ two models that advance~~s~~ on the standard temperature-index approach (enhanced temperature index models). We thus group all models in four categories: eight energy balance models, one simplified energy balance model, two ~~one~~ enhanced temperature index model and four temperature index models (Fig. 3). Temperature index model simulations were provided both in the uncalibrated and calibrated version where possible.

We rank models by complexity assuming that the temperature-index models are the simplest and the energy balance models the most sophisticated ones in the sense that they include the most complete representation of the physics of the processes leading to melt under debris. A general description of the model approaches can be found below, and an overview of participating models is shown in Table 2 and Fig. 3. A more detailed description of each model can be found in the Supplement, which fully documents unpublished approaches and includes sufficient detail of published models to enable understanding of the models' main characteristics and of the differences that are relevant for this intercomparison.

The models will hereafter be referred to by the model acronyms in Table 2. We used the model acronym provided in previous publications, or, when no name was provided, by the first three letters of the first author's last name followed by the publication year.

Table 2. Overview of the models used in this intercomparison, sorted alphabetically within each model category. Note models $GRO17_{A/B}$ are together as they contain the same information in this table. The Supplement describes further model details. EB = energy balance model; SEB = simplified energy balance model; ETI = enhanced temperature index model; TI = temperature index model. Ta = Air temperature, p = air pressure, RH = Relative humidity, FF = Wind speed, S↓, S↑ = Incoming/outgoing shortwave radiation, L↓ = Incoming longwave radiation, P = precipitation. NA = not applicable.

Model name	Model type	Meteorological inputs	Temporal Resolution	Number of debris layers (N) Thickness of debris layer (h_l)	Model references
A-Melt	EB	T_a, RH, FF, S↓, S↑, L↓, P	hourly	N = 1	Rets and Kireeva (2010), Elagina et al. (2025)
d2EB	EB	T_a, RH, FF, S↓, S↑, L↓, p	hourly	N layers of 1 cm each	Reid and Brock (2010), Steiner et al. (2018, 2021)
DEB_{CF}	EB	T_a, RH, FF, S↓, S↑, L↓, P, p	hourly	N = $h_d/10$	Reid and Brock (2010), Fyffe et al. (2014)
DEB_{PG}	EB	T_a, RH, FF, S↓, S↑, L↓, p	hourly	N = $h_d/3$ if $h_d \leq 6$, $h_l = 2{-}2.05$ if $h_d > 6$	Reid and Brock (2010)
$GRO17_{A/B}$	EB	T_a, RH, FF, S↓, S↑, L↓, p	hourly	N = 1	Evatt et al. (2015), Groos et al. (2017a, b)
ROU15	EB	T_a, RH, FF, S↓, S↑, L↓, P	hourly	N = $h_d/10$	Rounce et al. (2015)
THRED	EB	T_a, RH, FF, S↓, S↑, L↓, P	daily	N = 1	Fujita and Sakai (2014)
MCC19	SEB	T_a, S↓, S↑	hourly	N = $h_d/10$ if $h_d < 10$ cm, N layers of 1 cm if $h_d > 10$ cm	McCarthy (2025)
$DETI_m$	ETI	T_a, S↓, S↑	hourly	NA	Carenzo et al. (2016), modified
KO2	ETI	T_a , S↓, S↑	daily	NA	Winter-Billington et al. (2020)
DDF_{debris}	TI	T_a	hourly	NA	Kayastha et al. (2000)
Hyper-fit	TI	T_a	hourly	NA	Anderson and Anderson (2016), Anderson et al. (2021)
KM1	TI	T_a	daily	NA	Winter-Billington et al. (2020)
KP1	TI	T_a	daily	NA	Winter-Billington et al. (2020)

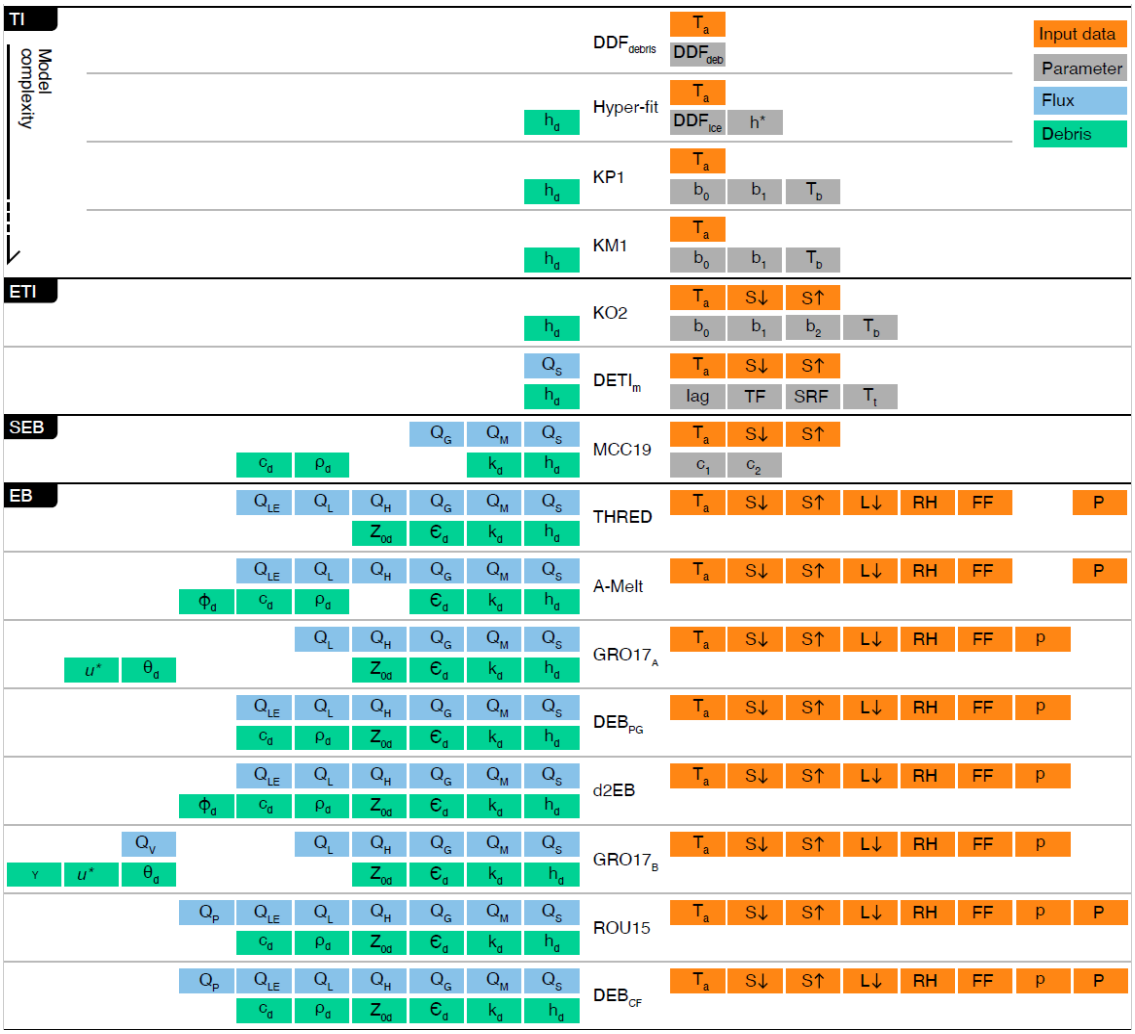


Figure 3. A graphical representation of model complexity, from top (simplest models) to bottom (most sophisticated models), for all four model categories considered in this intercomparison experiment (TI: temperature index models; ETI: enhanced temperature index model; SEB: simplified energy balance model; EB: energy balance models). Model complexity includes the required meteorological data (orange), empirical parameters (grey), debris properties (green) and energy fluxes (blue). Symbols for the meteorological input data are as in Table 2. Abbreviations and symbols for the fluxes, input data and parameters are as in the text. More details about models are in the Supplement.

Energy balance models

General model approach

Energy balance models calculate sub-debris melt by solving two main equations: i) the heat exchange at the debris-atmosphere interface and ii) the heat conduction of this surface energy into the debris, until the energy reaches the debris-ice interface and is transferred to the ice. If the ice is at 0°C, this energy is used for melt; otherwise, energy is used to warm the ice towards 0°C. This assumes that no other energy transfer occurs within the debris.

The general debris surface energy balance equation, following the notation of Reid and Brock (2010), is:

$$S\downarrow + S\uparrow + L\downarrow + L\uparrow(T_s) + H(T_s) + LE(T_s) + G(T_s) + P(T_s) = 0 \quad (1)$$

where *S is the net shortwave radiation*, $L\downarrow$ *and* $L\uparrow$ are the incoming and emitted longwave radiation, H is the turbulent sensible heat flux, LE is the turbulent latent heat flux, P is the heat flux due to precipitation, G is the heat conducted into the debris

(equivalent to G_1, heat conducted into the first debris layer, in Fig. 4) and T_s is the surface temperature. The fluxes that are a function of the surface temperature T_s are explicitly indicated as such.

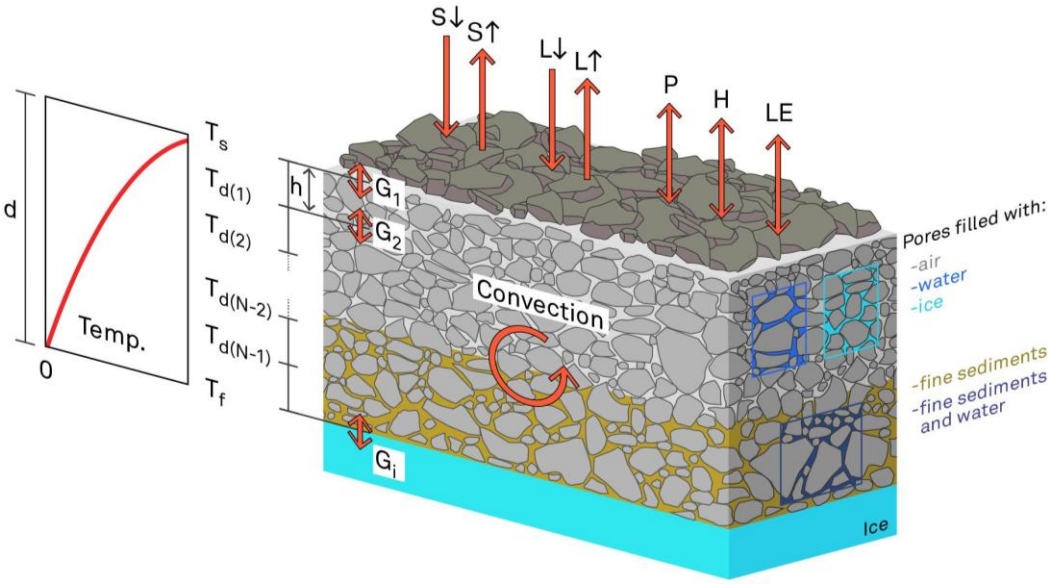


Figure 4. Scheme of energy fluxes at the interface air-debris, within debris and between debris and ice. Fluxes are considered as positive when directed towards the surface and negative when away from the surface. The debris is discretised into N layers of height (h) each. Symbols for the fluxes are as in the text.

The conductive heat flux G is the heat transferred through the debris layer to reach the ice and melt it, and it depends on the properties of the debris layer. This flux is calculated with the heat conduction equation through the debris (Eq. S7), and most models solve it by iteratively computing the debris surface temperature to close the energy balance, unless they assume a linear temperature profile or steady-state conditions, in which case the simpler linear Eq. S3 is used. When solving iteratively, most models, building on Reid and Brock (2010), use an iterative Newton–Raphson method to calculate surface temperature (Eq. S6), where the debris temperature is calculated for N layers of thickness *h*, with boundary conditions defined by the surface temperature, *Ts*, and the temperature of the debris/ice interface, which is assumed to stay at $T_f = 0°C$ for all models except one (A-melt).

The two main equations of the surface energy balance and heat conduction through the debris are similar for most models. Models however diverge in several aspects: i) the actual number of fluxes that are included; ii) the way individual fluxes are calculated, which can be more or less sophisticated; iii) the assumption about the temperature gradient in the debris; iv) the way the debris layer is discretised to calculate the conduction flux (i.e. the number of layers and their thickness); v) the numerical scheme to solve the two coupled equations above (see Supplement); vi) the ability to treat the interaction of debris with snow; and vii) the model temporal resolution (daily versus hourly).

EB model complexity

We use those aspects of model characteristics and representation of the debris domain to arrange energy balance models along an axis of complexity, from the simplest to the most sophisticated (Fig. 3). We also include in our arrangement the temporal resolution (hourly versus daily, with hourly models regarded as more complex). Since in this intercomparison ablation seasons with no snow cover (or only occasional snow cover) were chosen, the ability of the models to deal with snow is not taken into account. The overall definition of model complexity that includes also temperature-index and intermediate models is discussed at the end of this section (and also illustrated in Fig. 3).

All energy balance models ~~calculate the~~ directly use the provided observed net shortwave radiation flux and calculate the longwave radiative fluxes and the turbulent sensible heat flux at the surface (Fig. 3). All models except one (GRO17) include the turbulent latent heat flux at the debris-surface, and two models include the heat flux due to rain (ROU15 and DEB_{CF}). Models differ substantially in how turbulent heat fluxes are calculated. Building on Kuzmin (1961) and Nicholson and Benn (2006), most models use simplified bulk approaches with constant turbulent exchange coefficients (Steiner et al., 2018, 2021; Fujita and Sakai, 2014) and do not take into account atmospheric stability, thus assuming neutral conditions (Table S15 and S16). Only two models (DEB_{CF} and DEB_{PG}) account for the stability of the atmosphere using non-dimensional stability functions for momentum and heat expressed as functions of the Richardson number (Reid and Brock, 2010; Fyffe et al., 2014). A second major difference is the way the relative humidity of the debris surface is treated to calculate the latent heat flux. Since no data on the water content within the debris were available at any of the sites, modellers either neglected this flux or made assumptions on the actual relative humidity of the air at the debris surface (RH_s) based on the relative humidity of the air or precipitation occurrence (Table S16). These vary from assuming that the surface is saturated when it rains (DEB_{CF}, ROU15) to assuming that RH_s=100% when the air relative humidity at the measurement height (RH_a) is 100% (DEB_{PG}).

Models also differ in how the debris layer is represented. Four models ($GRO17_A$, $GRO17_B$, THRED and A-Melt) assume the debris can be treated as a single layer with a linear temperature gradient between the debris surface and the underlying ice (Table 2). This is an assumption that has been made for simulations with a time step of 24 hours, informed by temperature measurements showing that, although the profile is nonlinear at various times throughout the day, it is approximately linear on a 24 hour averaged basis (Nicholson and Benn, 2006; Brock et al., 2010; Reid and Brock., 2010). The only other study that used this assumption, Brock et al. (2010), assumes a linear temperature profile in simulations at 1 hour time-step, but introduces a 'debris heat storage' flux to account for debris warming during the day and cooling at night. This assumption simplifies the calculation of heat conduction into the debris (Eq. S7), thereby considerably reducing the computational costs.

Four models (DEB_{CF}, ROU15, THRED, A-Melt) represent snow on the debris surface by calculating snowmelt until the debris is exposed again (Table S19). Even here, differences are evident: some models accumulate snow on the ground using the precipitation and air temperature record (THRED) while other models use the snow depth record to calculate snowmelt as long as snow is on the ground (DEB_{CF}, ROU15).

There are other smaller differences between models. All models assume the ice to be at melting point, except for A-Melt (see Sect. 2 in the Supplement). All models use the thermal conductivity value provided for each site (Table S1) for the heat conduction flux calculations and the debris emissivity provided for the calculation of the outgoing longwave radiation flux. All models used the aerodynamic surface roughness length provided for the calculation of the turbulent heat fluxes, except the A-Melt model which parameterised it based on Kumin (1961).

Inclusion of convection within the debris

All models, with the exception of $GRO17_B$, assume that energy conduction into the debris is the only mechanism by which heat is transported through the debris to the underlying ice. The turbulent latent heat flux within the debris - and its variation with depth - was introduced by Evatt et al. (2015) to reproduce the peak in melt rate associated with the critical debris thickness (Østrem, 1959; Kirkbride and Dugmore, 2003; Reznichenko et al., 2010), and address a potential contradiction between model predictions and field observations: some field-derived Østrem curves show an initial increase in melt rates (compared to clean ice) up to a thickness of a few cms, after which melt decreases and reduces below the clean ice melt rate at a critical debris thickness (Østrem, 1991; Kirkbride and Dugmore, 2003). The only other work that attempted to reproduce this behaviour (Reid and Brock, 2010) attributed it to the patchiness of debris for a thin, non-uniform debris layer, and proposed a patchiness parameter to mimic the increase in melt rate for thin debris. Evatt et al. (2015) instead included air flow through the porous debris layer, and accounted for the energy exchange between the moving air and the ice at the bottom of the debris layer that

takes the form of either condensation or evaporation (turbulent latent heat flux). The airflow within the debris layer is attenuated with debris depth, causing a reduction in the evaporative heat flux as the debris thickens. This initially increases the melt rate, as less latent energy is used for evaporation and more energy becomes available for melting. However, as the debris layer continues to thicken, its insulating effect eventually dominates, leading to a reduction in the melt rate. The $GRO17_B$ model builds on the Evatt et al. (2015) model to reproduce these processes within a porous debris layer. The model requires knowledge of porosity and grain size, which determine the friction velocity and wind speed attenuation parameters (Fig 3).

Simplified energy balance model

One simplified energy balance model (MCC19; McCarthy, 2025) is part of this intercomparison (Supplement Sect. 2.2). At the debris surface, the model computes the net shortwave radiative flux and the conductive heat flux, and represents the remaining fluxes using the air and debris temperature difference together with two free parameters (Fig. 3, Table S12) (cf. Oerlemans, 2001). The model conducts heat through the debris layer using the one-dimensional heat equation, where the boundary condition at the ice surface is the temperature of melting ice (following Reid and Brock, 2010), and the boundary condition at the debris surface is the simplified debris-surface energy balance. The model therefore only requires air temperature and incoming shortwave radiation as meteorological forcing and debris parameters (conductivity, heat capacity and density) to solve the heat conduction equation, and has two parameters that need to be calibrated. Melt is calculated from the conductive heat flux at the base of the debris layer.

Enhanced temperature index models

The debris enhanced temperature index (DETI; Carenzo et al., 2016) model was developed as a model of intermediate complexity between a temperature index model and an energy balance model, building on similar developments for clean ice (the ETI model, Pellicciotti et al., 2005). It includes the shortwave radiation balance, and a term dependent on air temperature that represents empirically all other fluxes in the energy balance equations. The model's empirical parameters are a function of debris thickness, to account for the time needed to transfer energy from the surface to the ice, and were derived through functional relationships between the shortwave radiation flux and temperature with sub-debris melt simulated by an energy balance model at different thicknesses. It is designed to run at hourly resolution.

Winter-Billington et al. (2020) introduced an enhanced temperature-index model based on a mixed-effects approach. Fitted using data from 27 glaciers, the model predicts degree-day factors as an exponential function of debris thickness and, combined with air temperature data and net shortwave radiation as a second fixed effect, estimates melt at a daily time step. The mixed-effects framework allows this model to be applied to new sites without recalibration, while providing prediction uncertainty based on the original training data. With two fixed-effect predictors, the model KO2 is considered an enhanced temperature-index model under our classification scheme.

Temperature index models

Temperature-index models assume that melt is linearly dependent on air temperature and use a degree-day factor to estimate melt. The degree-day factor is generally calibrated to reproduce observed melt under debris (e.g. Kayastha et al., 2000), and likely cannot be transferred to sites with a different debris thickness or different climates ~~(Winter-Billington et al., 2020)~~. Anderson and Anderson (2016) developed a sub-debris melt model (Hyper-fit; Anderson et al., 2021) where a degree-day factor for clean ice is used to estimate a hypothetical bare ice melt rate at each site. To estimate sub-debris melt, the bare ice melt rate is reduced based on local debris thickness and a characteristic debris thickness length scale, h*. The characteristic length scale controls how rapidly sub-debris melt asymptotes toward zero melt as debris thickens, via a hyperbolic relationship. The parameter h* can be calculated as a function of debris properties (conductivity and porosity, and ambient conditions) but

the model performs best by constraining h* directly using empirical debris-thickness melt data. The model has two parameters: DDF_{ice} and h* (Fig. 3, Supplement Sect. 2.4).

Winter-Billington et al. (2020) introduced two modifications of the temperature-index model ~~. by Kayastha et al. (2000), both~~ designed for daily simulations. Models KM1 and KP1 differ from KO2 in that they do not use shortwave radiation and use debris thickness as the sole fixed effect. The models KM1 and KP1 share the same structure but are considered different models because they differ in their training dataset and parameter values, as well as calibration scheme (see Sect. 4.1). ~~In the first one (KP1/KM1), the degree-day factor is computed as a function of debris thickness through an empirical relationship with seven parameters, while in the second model (KO2) the degree-day factor is a function of both debris thickness and potential incoming shortwave radiation that has four empirical parameters.~~ The last model included in this intercomparison is the DDF_{debris}, the simplest degree-day factor approach calibrated for sub-debris melt reduction (Fig. 3).

Sorting models by complexity

Models have different levels of complexity based on the number of input data they require, the number of fluxes they calculate, the physical realism of the equations used to calculate the fluxes, the assumptions made, the numerical schemes used, the temporal resolution, the number of parameters and debris properties required and the vertical discretisation of the debris layer. We have previously identified four model categories with varying levels of complexity. Sorting the complexity of models within each model category is less straightforward. In order to use a definition that seeks to be the least subjective as possible, we quantify the model complexity based on the sum of the total number of input data required, fluxes calculated, empirical parameters and debris properties required. This is illustrated in Fig. 3. Based on this definition, the most complex models are DEB_{CF} and ROU15, with a total sum of 21 (8 input data, 7 fluxes, 6 debris properties), but we regard DEB_{CF} as more complex because it accounts for atmospheric stability corrections in the calculation of the turbulent fluxes. The simplest model is the DDF_{debris}. We arrange the models along this axis of complexity in all figures throughout the manuscript.

4.1. Model calibration

The energy balance models do not require calibration and were run with the input meteorological forcing and debris properties provided per site. The only energy balance model that adopted a calibration strategy was $GRO17_B$ to account for the latent heat flux within the debris and at the debris-ice interface in a realistic manner. Calculation of these fluxes requires porosity and grain size, and lack of site-specific accurate field observations of these debris properties led to unrealistic turbulent heat fluxes and poor simulation outcomes. Therefore, the turbulent heat fluxes from $GRO17_{A/B}$ were calibrated against the approach of Nicholson and Benn (2006), to exclude combinations of model parameters (low friction velocity and rapid wind speed attenuation) that lead to unrealistic turbulent heat fluxes (i.e. converging to zero). The simplified energy balance model, one enhanced temperature index model and other temperature index models were all calibrated.

For some of the latter models ($DETI_m$, ~~KM1~~ and Hyper-fit), the uncalibrated versions were also run to assess their transferability. The definition of uncalibrated was left to each modeller, so that in some cases literature parameters from the sites were used as long as they were not optimised for the same time period. The choice of how to calibrate any model was considered part of the model setup, and was left to the individual modellers (parameters, goodness-of-fit metrics and target variable). The entire period was used for calibration because the data series were not long enough to split into separate calibration and validation subperiods. A brief overview is provided below and full details of the calibration procedures are in the model descriptions in the Supplement.

The simplified energy balance model was calibrated using both the surface temperature and sub-debris melt data provided, following a multiparameter, multiobjective optimisation approach (after Rye et al. 2010). The $DETI_m$ model was calibrated

against the DEB_{CF} simulations, as in the original publication, while uncalibrated runs used the original model parameters from Carenzo et al. (2016), which were obtained for Miage glacier. The calibrated version of Hyper-fit used hourly cumulative melt data to optimise the characteristic length scale, h*, for each specific site and time period. The uncalibrated version of Hyper-fit used previously-published~~, independent parameters~~ melt and debris thickness data from six of the nine sites, outside the study period of this experiment, to estimate values for h* (see Table S24). For the other three sites, the global mean h* value (Anderson and Anderson, 2016) was used to represent h*. Therefore, these Hyper-fit model runs were not defined as uncalibrated but as estimated (E). ~~The KP1 model (in which the degree-day factor depends on only the debris thickness) was run in its calibrated (KP1) and uncalibrated (KM1) version, with the calibrated version optimising the threshold temperature T_0 using cumulative melt data while all other empirical coefficients are as in Winter-Billington et al. (2020). The KO2 model (in which the degree-day factor depends on both debris thickness and net shortwave radiation) is uncalibrated, as it uses the empirical coefficient from Winter-Billington et al. (2020).~~ In models KP1, KM1 and KO2, the parameters b_0, b_1 and b_2 were not recalibrated. The model parameter values are applicable to any site without recalibration by definition of the mixed-effects modelling approach (Winter-Billington et al., 2020). However, the value of the melt onset threshold air temperature (T_b) that was used to calculate positive degree days (PDD) as input to model KM1 was recalibrated for each site. Due to the original data used to fit the models, it was not possible to recalibrate the value of T_b to compute PDD for input to models KP1 or KO2 (see Winter-Billington et al. (2020) for details), and therefore these models are considered uncalibrated.

4.2. Model uncertainty

Additional simulations were performed for the two experiments (see Sect. 2) and for all energy balance models to quantify uncertainty associated with debris properties. At each site, 100 samples were randomly taken from a uniformly distributed +/- 10% uncertainty range applied to surface roughness, debris thermal conductivity and debris porosity, while a +/- 5% range was applied to emissivity. These standardised uncertainty ranges match those provided for most properties at most sites, and are the same ranges used by Reid and Brock (2010). The parameters and corresponding uncertainty ranges are provided in Table S1. The standard deviation of the 100 simulations was used to estimate the uncertainty of modelled melt.

In an attempt to use an approach as similar as possible to the uncertainty in debris properties, all temperature index models used +/-10% range around the calibration range of model parameters at each of the nine sites, and 100 randomly sampled combinations of parameters within that range. Uncalibrated models used a +/-10% of the applied parameter values.

4.3. Model Evaluation

All models are evaluated against measured sub-debris ice melt, while only the energy balance models and the simplified energy balance model are assessed against the debris surface temperature too, as the other models do not calculate surface temperature. Model performance was only evaluated over one ablation season because data were not available for additional seasons at most sites. This limited our capacity to assess model robustness over multiple seasons. For the models requiring calibration, i.e. temperature-index, enhanced temperature-index and simplified energy-balance models (MCC19, DETI, Hyper-fit, KM1, DDF_{debris}), in particular, this does not allow a separate validation period, as all models requiring calibration were optimised for the entire period of data available.

4.3.1. Evaluation against melt observations

Models were evaluated against observed sub-debris melt using daily surface height change measurements for Arolla, Changri Nup, Djankuat, Lirung, Suldenferner, Tapado and Tasman, and three discrete measurements of ablation stakes over the ablation season for Miage and Pirámide. The percentage final melt error, simply defined as the difference between modelled and observed melt at the end of the simulation period as a percentage of the observed melt, was used to evaluate the modelled

melt. This metric allows for comparison across all sites regardless of the resolution and type of validation data available. We evaluate performance across models and sites based on the median percentage error, mean absolute error and the interquartile range (spread). We follow a similar method to Farinotti et al. (2017) and rank the models' performance based on their ranking for median, mean absolute error and interquartile range errors.

4.3.2. Evaluation against debris surface temperature

Hourly and daily debris surface temperature was used to evaluate the models using the root mean square error (RMSE) and bias between modelled and observed surface temperature. The Nash-Sutcliffe Efficiency was also used to further evaluate the hourly performance of models because this metric is most sensitive and meaningful for clearly defined daily cycles. The observed debris surface temperature was derived at all sites from the measured outgoing and incoming longwave radiation (which were available at all sites), following Stefan Boltzmann law as:

$$T_s = \left(\frac{L\uparrow - ((1-\varepsilon) \cdot L\downarrow)}{\varepsilon \cdot \sigma} \right)^{1/4} \quad (2)$$

where T_s is the surface temperature of the debris (in K), ε is the emissivity of the debris, σ is the Stefan-Boltzmann constant (in W m^{-2} K^{-4}), and $L\uparrow$ and $L\downarrow$ the outgoing and incoming longwave radiation (in W m^{-2}), respectively.

5. Results

5.1. Performance of model ensemble at sites

Fig. 5a shows the ensemble mean daily melt compared to the observations at each site for all models considered. As expected, sites with thinner debris show higher daily melt rates than sites with thicker debris, with the exception of Changri Nup due to its low ablation season temperature (Fig. 2). The observed melt is within the ensemble range of modelled melt at all sites, but variations among sites and groups of models are strong. Relative errors and the spread of model performance become larger with increasing debris thickness, with the exception of Tapado. The ensemble performance of models can also be observed in Fig. S2, where the continuous cumulative melt for the entire period is shown for each site and model.

Three groups of sites are evident. First are the alpine sites of Arolla and Suldenferner, where models' performance is high and consistent for most energy balance and temperature-index models except for some uncalibrated ones. The models' median error is lowest at Suldenferner (-1.1%), followed by Arolla (-3.7%) (Table 3). For these sites, models are consistent and show the smallest spread in terms of interquartile range (14.7% for Suldenferner and 15.2% for Arolla, Table 3).

Second, at Changri Nup, Miage, Pirámide and Tapado (three of the four highest sites in elevation), models' performance is relatively high, with -7.5% median error for Tapado, 7.2% for Pirámide and a higher -15.8% at Changri Nup and 23.7% at Miage. These sites show a larger spread in model performance, with 24.2% at Changri Nup, 34.5% at Miage, 37.8% at Pirámide and 76.7% at Tapado, although the latter is because of the low absolute melt at this site, which makes the relative errors larger.

Finally, three sites stand out as characterised by low performance, high errors for most models and large spread among models: Lirung, Tasman and Djankuat (with median melt errors of 55.2%, 57.9% and 90.8% respectively, Table 3). However, the consistency of the large median errors across model groups differs at these three sites. At Tasman, the energy balance models show high consistency and are grouped together but highly overestimate the observed melt, while the temperature-index models perform better but have a much larger spread. On Lirung, the energy balance models also perform poorly and consistently overestimate melt, while the temperature-index models consistently show a smaller error but large spread. Finally, on Djankuat, all four groups of models perform poorly and in a comparable manner, with the energy-balance models having most of the largest errors and the temperature-index models the lowest errors. In general, these poor-performance sites have

both high debris thickness (30, 30 and 61 cm for Lirung, Tasman and Djankuat, respectively, Fig. 5 and Table 1), and very high debris thermal conductivity (Table S1).

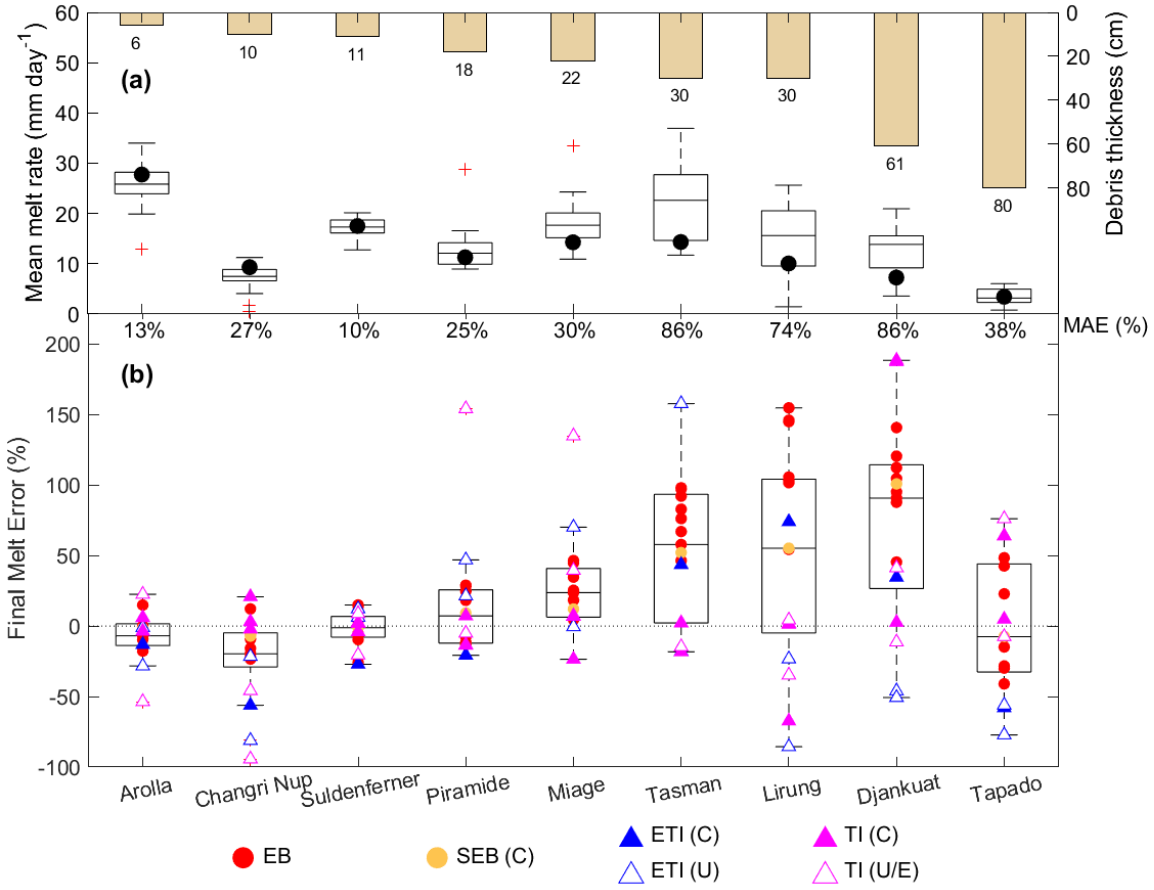


Figure 5. Performance of the ensemble of models. (a) Modelled mean daily melt by the model ensemble (boxplot) and observed mean daily melt (black dot). Red crosses indicate outliers, defined as more than 1.5 times outside of the interquartile range. The model ensemble mean absolute error (MAE) is expressed as a percentage and shown for each site. Bar plots at the top indicate the debris thickness at each site. (b) Final melt error (in %, as defined in Section 4.3.1) at each site, for each group of models, and for calibrated and uncalibrated/estimated models separately. Sites are ordered by debris thickness.

Validation against hourly surface temperature shows both similarities and differences from the validation against melt observations (Table S27, ~~Fig. S3~~ Fig. 8b and Fig. S34). Arolla, an Alpine site with good performance against melt observations, has the best performance (smallest RMSE) against debris surface temperature, with a median RMSE of 3.1°C. One of the worst performing sites against melt observations, Lirung, also has the worst performance against surface temperature, with a RMSE of 7.1°C, and the lowest Nash-Sutcliffe Efficiency (Fig. S34). Tasman is a distinct site: it shows a high performance against surface temperature with a RMSE of 3.8°C (Table S27), and a high and consistent NSE across models (Fig. S34), in contrast to the poor agreement with observed melt (Fig. 5 and Table 3). Contrasting to the high melt performance, most simulations in Tapado show a high RMSE, but also a high NSE, indicating the daily cycle of temperature is well reproduced despite a temperature bias (~~Fig. S3~~ Fig. 8b and S34). The rest of the sites (Changri Nup, Djankuat, Miage, Piramide and Suldenferner) have a similar RMSE between 4.2 and 4.9°C (Table S27), despite their different performances against melt. Model consistency for surface temperature at sites is variable, very high in Arolla and Tasman, and low on Lirung and Tapado (~~Fig. S3~~ Fig. 8b).

Table 3. Modelled melt error across models and across sites in percentage. Models are ordered by complexity as defined in Figure 3. The last three columns correspond to the median (Med.), interquartile range (IQR) and mean absolute error (MAE) across sites per model, and the last three rows correspond to the median, IQR and MAE across models per site. We provide the IQR as a measure of the spread in model performance and the MAE as a measure of absolute error. The values of median, IQR and MAE in the bottom right of the table indicate the overall respective value across sites and models altogether. "C" indicates calibrated, and "U" indicates uncalibrated. Values between -20% and +20% are shown in **bold**.

Cat.	Model	ARO	CNU	DJA	LIR	MIA	PIR	SDF	TAP	TAS	Med.(%)	IQR (%)	MAE (%)
EB	DEB_{CF}	**-17.7**	-23.3	45.4	103.7	**4.6**	**-7.2**	-25.7	-41.0	46.5	**-7.2**	69.6	35.0
EB	ROU15	**-9.6**	-22.4	87.8	103.1	**18.4**	**-7.5**	**-5.8**	**-7.7**	67.1	**-5.8**	80.4	36.6
EB	$GRO17_{B}$	**-6.9**	-23.4	120.6	154.8	46.5	29.0	**15.0**	48.5	98.1	46.5	94.2	60.3
EB	d2EB	**-16.0**	**-15.8**	90.8	105.7	24.2	**-11.7**	**-5.3**	-28.3	57.9	**-5.3**	82.0	39.5
EB	DEB_{PG}	**14.9**	**-19.7**	95.2	54.4	23.7	**18.3**	**-9.5**	-29.8	76.3	**18.3**	72.0	38.0
EB	$GRO17_{A}$	**0.2**	**12.2**	112.4	145.1	44.8	28.5	**13.9**	42.7	97.3	42.7	87.6	55.2
EB	A-Melt	**-9.2**	**-9.2**	104.6	101.8	25.4	**-13.2**	**-7.3**	**-14.8**	92.2	**-7.3**	104.8	42.0
EB	THRED	**-7.1**	**-5.7**	140.8	146.1	34.7	24.8	**4.0**	22.9	82.9	24.8	95.8	52.1
SEB	MCC19-C	**-1.6**	**-7.1**	100.9	55.2	**12.1**	**9.3**	**-1.1**	**-7.4**	52.2	**9.3**	55.9	27.4
ETI	DETIm-C	**-13.0**	-56.2	34.6	74.0	**-0.5**	-20.7	-27.1	-58.0	43.6	**-13.0**	71.2	36.4
ETI	DETIm-U	**-1.0**	-21.5	-46.0	-23.3	**-0.5**	21.4	**6.2**	-56.1	**-18.0**	**-18.0**	30.2	21.6
ETI	KO2-U	-28.3	-81.1	-50.7	-85.5	70.2	46.9	**12.2**	-77.2	157.8	-28.3	130.9	67.8
TI	KM1-C	**-3.7**	**-2.1**	**2.6**	-67.3	-23.5	**7.2**	**-4.0**	**4.9**	**-18.2**	**-3.7**	22.7	**14.8**
TI	KP1-U	-53.6	-94.3	**-11.3**	-34.8	134.9	154.2	-20.5	**-7.5**	542.6	**-11.3**	179.2	117.1
TI	Hyper-fit-C	**6.1**	**3.1**	188.5	**1.2**	**7.3**	**-13.6**	**2.5**	63.9	**2.1**	**3.1**	**19.5**	32.0
TI	Hyper-fit-E	22.6	-45.8	41.3	**4.3**	39.5	**-5.0**	**8.8**	76.2	**-14.7**	**8.8**	47.4	28.7
TI	DDF_{debris}-C	**6.0**	20.8	187.6	**1.7**	**6.9**	**-13.3**	**1.9**	64.1	**2.3**	**6.0**	29.8	33.8
-	Median (%)	**-3.7**	**-15.8**	90.8	55.2	23.7	**7.2**	**-1.1**	**-7.5**	57.9	**4.3**	-	-
-	IQR (%)	**15.4**	24.2	87.8	109.1	34.5	37.8	**14.7**	76.7	91.2	-	59.5	-
-	MAE (%)	**12.8**	27.3	85.9	74.2	30.5	25.4	**10.1**	38.3	86.5	-	-	43.4

Model performance against *daily* surface temperature is higher but results from averaging out of errors in the hourly time series (Table S27, ~~Fig. S3~~ Fig. 8b). At this time scale, the overall pattern of performance across sites does not change substantially, but tends to be more uniform, suggesting that aggregation to daily resolution, by smoothing out variability and errors, is not appropriate to identify models' skills.

Consistency of model ensemble performance

A key aspect of energy balance model performance is the accuracy in simulating both sub-debris melt and debris surface temperature. To evaluate this, we plot the melt error against the daily bias in surface temperature for each site separately (Fig. 6). A high model performance across both validation variables is indicated by models centred around the origin, and a low scatter indicates high consistency across models.

At most sites, the majority of models diverge from the origin (Fig. 6). A consistent warm bias in surface temperature is evident at most sites, with the exceptions of Piramide and Tasman, which exhibit a cold bias across models (Fig. 6). Models perform consistently well at Arolla and Suldenferner (low scatter) for both melt and surface temperature (Fig. 6e,g). At Piramide and Tapado, models are distributed across three different quadrants, indicating a variability in melt errors despite consistent surface temperature biases (Fig. 6f,h). Conversely, at Lirung and Djankuat, models consistently fall in the same quadrant, albeit with high scatter and high errors for melt at both sites and for surface temperature at Lirung (Fig. 6 c,d). At Tasman, models show a rather high consistency, with a high overestimation of melt and a cold temperature bias for most models (quadrant IV, Fig. 6i). The observed consistency in surface temperature biases can also be observed in Fig. S4~~5~~, where all but two models show a positive temperature bias at most sites. The clustering in Fig. 6 also contrasts with the scatter in Fig. S4~~5~~, and provides a

clear demonstration that the sites' characteristics or data quality likely control the spread of performance more than the model physics (see discussion).

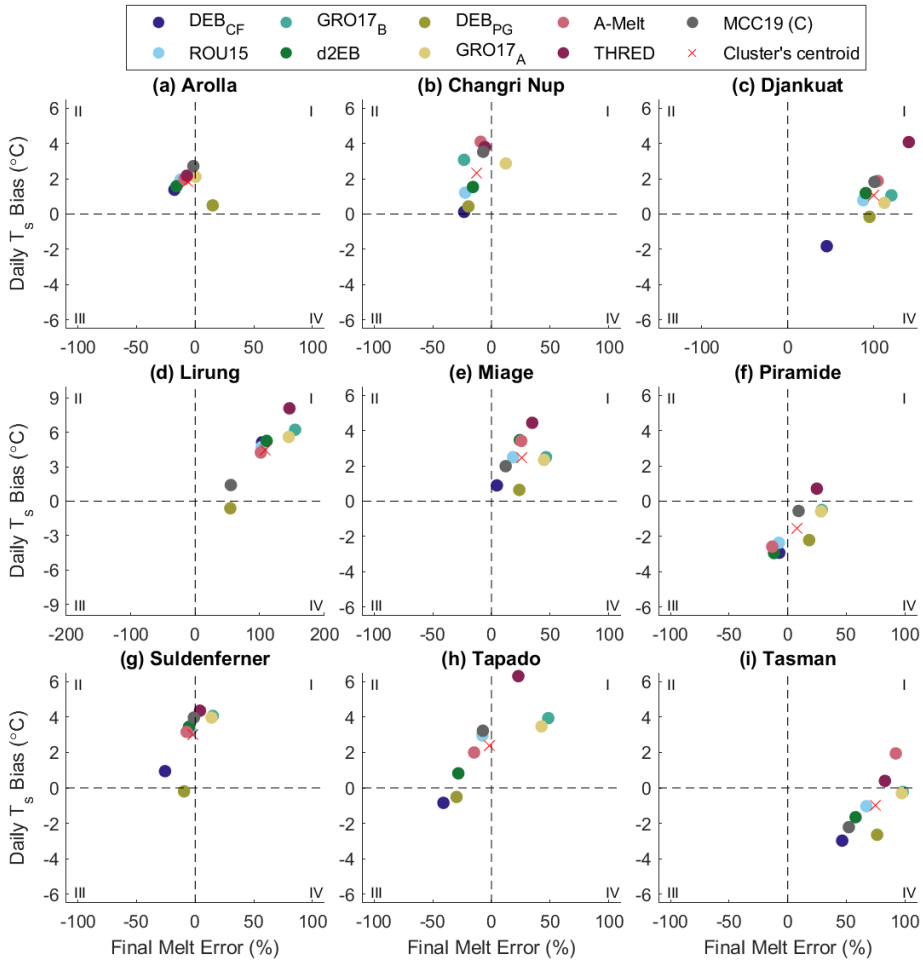


Figure 6. Consistency of model performance across the two validation datasets. For each site (a-i), each energy balance model (and simplified energy balance) is scattered based on their daily temperature bias and melt error. Dashed lines correspond to the zero line for both axes, and separate the plot in four quadrants. Quadrants above the horizontal dashed line indicate overestimation of surface temperature. Quadrants to the right of the vertical dashed line indicate overestimation of melt. Note Djankuat and Lirung have wider axes ranges due to the high errors at those sites.

5.2. Performance of individual models

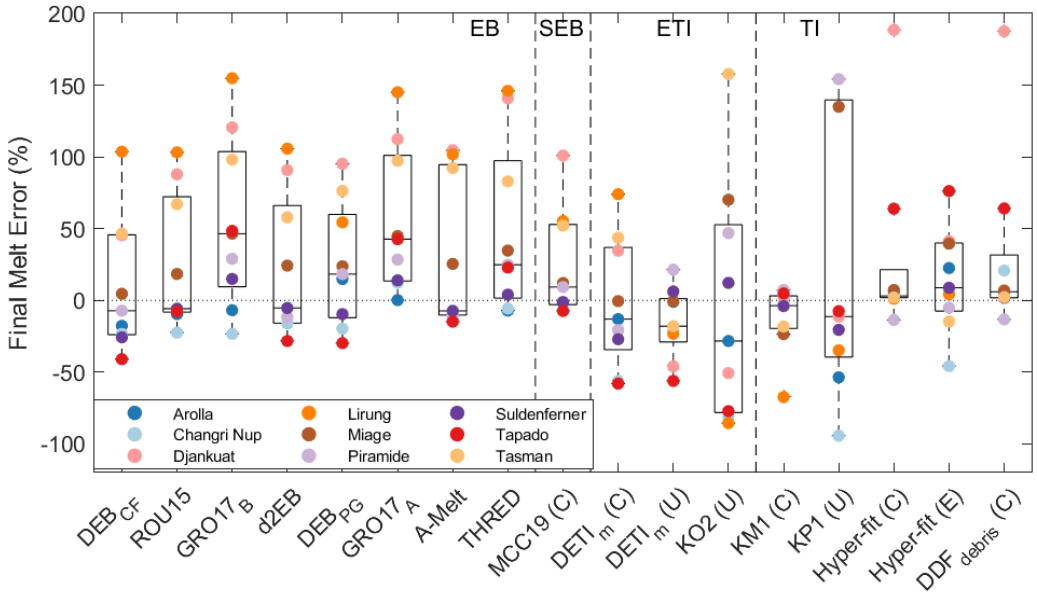


Figure 7. Final melt error (as defined in Section 4.3.1) for individual models at all sites expressed as interquartile box plots. Models are grouped in the four categories that we use in the paper (separated by dashed lines) and sorted from more complex (left) to less complex (right). Both calibrated (C) and uncalibrated (U) temperature index models are shown.

We evaluate the performance of individual models across sites against melt for all models (using the final melt error, Fig. 7, Tables 3 and 4), and against debris surface temperature for models solving the energy balance (using the root-mean-square-error, Fig. 8, Fig. S5~~6~~). Overall, models tend to overestimate melt, with a median final melt error of 4.3% and mean absolute error of 43.4% across all model runs (Table 3), dominated by the three sites where melt is consistently overestimated. However, there are important differences between models and sites. When removing the effect of the three worst performing sites, and the uncalibrated models, the median error across all model runs is -2.8%, and the mean absolute error 17.5%.

The highest ranked model (the calibrated KM1~~Hyper-fit~~) and the lowest ranked model in Table 4 (uncalibrated KO2) are temperature-index based models, highlighting that they perform very well when calibrated and show a much wider spread when uncalibrated. The calibrated temperature index models (calibrated Hyper-fit, DDF_{debris} and KM1) have a low error and high consistency, here regarded as the spread (interquartile range) among sites. They are the first three models ranked on Table 4. The uncalibrated temperature index models have higher errors and lower consistency. The Hyper-fit model performs well with estimated parameters ~~when uncalibrated~~ (Table 3) because it uses literature parameters for the specific study sites (even if not tuned for this experiment). KM1, KP1 and KO2 tend to underestimate melt, while the DDF_{debris} and the Hyper-fit models tend to slightly overestimate it. These differences may result from differing parameter calibration choices. The $DETI_m$ model performance is unusual, as the uncalibrated version performs almost as well as the calibrated version on average, but has a smaller spread across sites. The calibrated model strongly overestimates melt at Lirung, Tasman and Djankuat, as it follows the performance of its reference EB model (DEB_{CF} model, Sect. 4.2 and Supplement), and its performance reflects the performance of the DEB_{CF} model and thus ranks in the middle on Table 4.

Clear differences are also evident among energy balance models~~.~~, which show an overall overestimation of melt and relatively large errors (20.6% median, 44.8% MAE) when all sites are considered, and a slight underestimation of melt and much smaller errors (-6.3% median, 18.7% MAE) when the three sites with poorly constrained debris properties are not considered. The models that perform best in terms of the smallest median error are DEB_{CF} (-7.2% error, Table 3), ROU15 (-5.8%, Table 3) and d2EB (-5.3%, Table 3). The ones with the strongest consistency (taken as the IQR) across sites are DEB_{CF} (69.6%, Table 3) and MCC19 (55.9%, Table 3), which is a calibrated model. This makes DEB_{CF} the highest ranked energy balance model in

Table 4, followed closely by ROU15 and d2EB. The energy balance model with the lowest performance in terms of median error is $GRO17_B$ (46.5% error, Table 3), which is also the lowest ranked energy balance model in Table 4, followed closely by THRED and $GRO17_A$. Figure S6 presents an alternative visualization of model rankings, demonstrating that the method of visualization does not alter the overall results or ranking. The figure produces results and interpretations that are consistent with those reported in Table 4.

Table 4. Ranking of model performance based on median, interquartile range (IQR), and mean absolute error (MAE). Models are ranked on each metric and sorted by the mean rank.

Model Type	Model name	Rank Median	Rank IQR	Rank MAE	Overall Rank
TI	KM1 (C)	2	2	1	1.7
TI	Hyper-fit (C)	1	1	5	2.3
TI	DDF_{debris} (C)	5	3	6	4.7
TI	Hyper-fit (E)	8	5	4	5.7
SEB	MCC19 (C)	9	6	3	6.0
ETI	$DETI_m$ (U)	12	4	2	6.0
EB	DEB_{CF}	6	7	7	6.7
EB	ROU15	4	10	9	7.7
EB	d2EB	3	11	11	8.3
ETI	$DETI_m$ (C)	11	8	8	9.0
EB	DEB_{PG}	13	9	10	10.7
EB	A-Melt	7	15	12	11.3
EB	THRED	14	14	13	13.7
EB	$GRO17_A$	16	12	14	14.0
TI	KP1 (U)	10	17	17	14.7
EB	$GRO17_B$	17	13	15	15.0
ETI	KO2 (U)	15	16	16	15.7

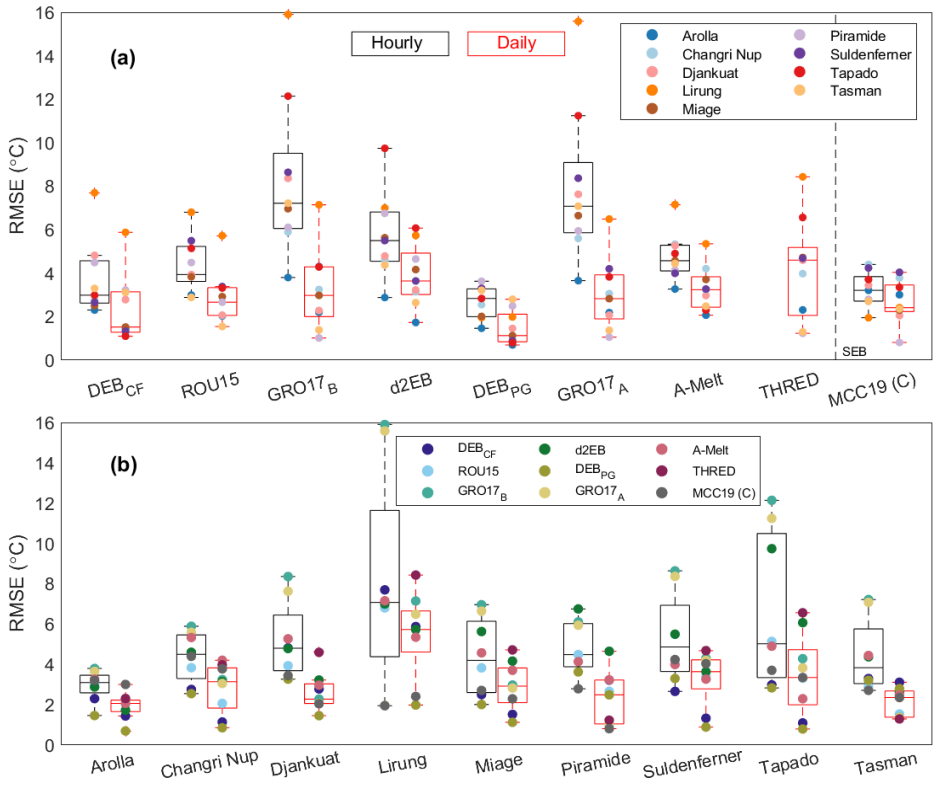


Figure 8. Validation of the energy balance model simulations (including SEB) against surface temperature, ~~at all~~ across models (a) and sites (b). Statistical boxes for each model show the distribution of the Root Mean Square Error of surface temperature at each site. Black boxes are for hourly data, red boxes are for daily data. Note that the THRED model is run and validated at daily resolution only.

The energy balance models (including the simplified energy balance model) were also evaluated against the surface temperature data (Fig. 8, Table S27). The RMSE is provided both at the hourly scale (of most simulations) and at the daily scale to include the THRED model. Important differences between energy balance models are apparent. At the **hourly scale**, DEB_{CF} and DEB_{PG} perform particularly well, with low median RMSE, of 3.0ºC and 2.8ºC (Fig. 8, Table S27), respectively, as well as high NSE of 0.81 and 0.88 respectively (Fig. S5~~6~~), and a low spread and thus high consistency across sites. ROU15 and A-Melt have a slightly higher median RMSE of 3.9ºC and 4.6ºC, respectively. The $GRO17_A$, $GRO17_B$ and d2EB models show the highest RMSE, of 7.2ºC, 7.1ºC and 5.5ºC, and lowest NSE of 0.15, 0.27 and 0.52, respectively (Fig. S5~~6~~), and a much larger spread across sites.

At the **daily scale**, the patterns of errors remain the same, but with overall lower RMSEs (Fig. 8, Table S27). The THRED model has the highest median RMSE (4.6ºC). The simplified energy balance model performs well at both temporal scales (RMSE of 3.2ºC and 2.4ºC, respectively), but this is to be expected given that it is calibrated against surface temperature. Sites with poor performance for more than one model are Tapado, Lirung and Piramide.

5.3. Model uncertainty

We depict the uncertainty in model results due to debris properties and model parameters (see Sect. 4.3) in Fig. 9. Model sensitivity to parameter uncertainty is small (around 5% for most models) compared to the modelled total error for most models considered and at most sites; particularly so for the energy balance models. For the energy balance models, the model sensitivity to uncertainty in debris properties propagates into final modelled melt uncertainty in a manner that is consistent both among models and sites. For empirical models, the comparison is more difficult, as each model has different parameters that control the melt calculations in a distinct manner; parameters have different meanings and units, and the plausible parameter ranges vary between parameters and are more difficult to define. Even with these caveats, however, model sensitivity due to estimated parameter uncertainty is smaller than the model error for most temperature index models. In general, empirical models have higher uncertainty, and the uncertainty is more variable among models, reflecting models' differences from one another. The uncertainty is slightly higher for the KM1/KP1 models, followed by the $DETI_m$ model, while the lowest uncertainty is for the DDF_{debris}, followed by the Hyper-fit models.

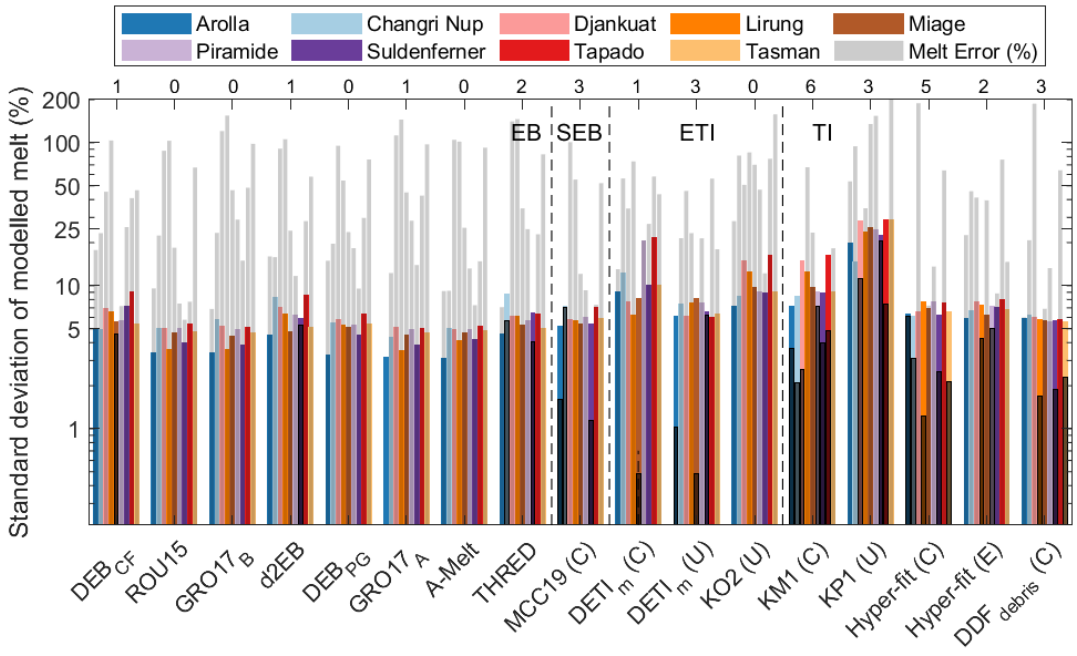


Figure 9. Model sensitivity to parameter uncertainty (standard deviation of the Monte Carlo 100 model runs in %) for all models. The model error is depicted on the same scale as grey bars (total melt error in %, as shown in Fig. 5 and 7), for each model and site. Note the logarithmic y-axis. Where the model error is less than the uncertainty value, the grey bar denoting melt error overlaps the

uncertainty bar and is therefore displayed darker. The number above each model bar denotes the number of sites (out of nine sites) in which the melt error is lower than the model uncertainty.

5.4. Østrem curves

All models until now have been evaluated based on their performance at the location of the automatic weather stations for the measured debris thickness. Here, we plot simulated melt as a function of debris thickness (as described in Sect. 2) (Fig. 10) and consider how divergent the models are when used to calculate melt at different thicknesses, which has implications for simulations of melt at the scale of an entire glacier, i.e. across a spatial domain of varying thickness.

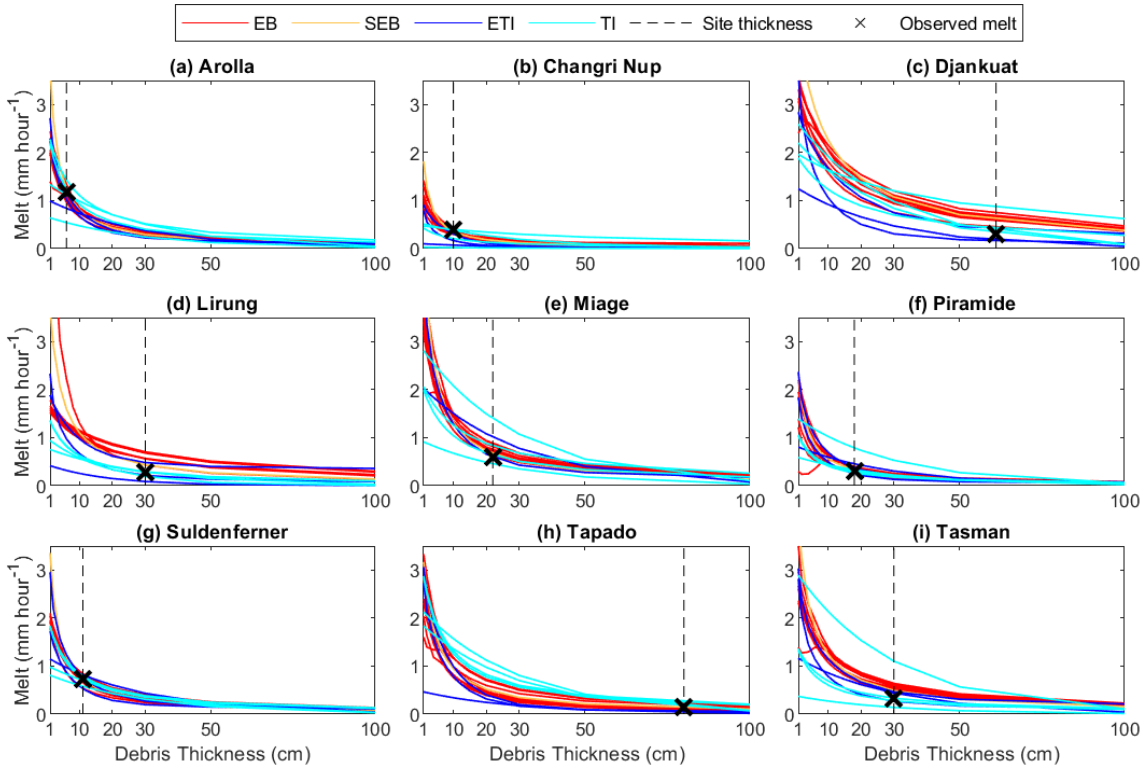


Figure 10. Østrem curves for all models, aggregated into model groups. The actual thickness at each site is indicated by a vertical line, and the observed melt by a cross.

Fig. 10 shows the Østrem curves simulated by each model, grouped by model type, and it highlights clear differences between models and sites (see also Fig. S8). The temperature-index models produce a larger spread for the same value of debris thickness than the other models and generally exhibit a more linear behaviour than the rest of the ensemble (Fig. S8), suggesting smaller sensitivity to debris thickness. This is not the case for the ETI and SEB, which simulate shapes that are close to those of the energy balance models. For most sites, and especially at Lirung, the spread among models increases dramatically for thin debris, implying that the choice of model is crucial for melt simulations for thin debris. It also suggests that we are not able to constrain the Østrem curves for thin debris, partly because this experiment is conducted with the input data and debris properties of the original location with thicker debris, which do not represent the properties of a very thin debris layer. Surface albedo in particular remained constant with debris thickness but should increase for thin debris as this becomes patchy and the debris cover area decreases (Azzoni et al., 2016). Models also diverge for thicker debris at Djankuat (Fig. 10~~b and~~ c).

For most models and sites, the slope of the curve is steeper for thin debris than for thick debris (Fig. 10, S8), suggesting that models are more sensitive where the debris is thinner. This points to the importance of correctly representing processes where debris is thin (Fyffe et al., 2020), where we expect a quicker response to changing meteorological conditions, quicker drying or moistening of debris, and a larger role of surface roughness. Only the GRO17$_B$ model is able to reproduce the peak in melt

occurring for thin debris based on the data provided and the experimental set up (Fig. S8). This melt enhancement is mostly visible at Tasman and Piramide and occurs only for thin debris.

6. Discussion

6.1. Performance across sites and importance of debris properties and input data quality

6.1.1. Sites with high performance

A clear result of our analysis is that model performance varies considerably between sites. Models perform well and in a consistent manner at the three European Alps sites: at Arolla and Suldenferner, with a consistent, high performance, and at Miage, with slightly larger errors (Sect. 5.3). For energy balance models, this might be due to a combination of three aspects: i) most energy balance models have been developed for initial application to those sites, and thus might be better suited to represent processes that dominate there; ii) the ease of access to these sites facilitates field visits, instrument maintenance and data quality checks, so that the quality of input and validation data might be higher; and iii) debris properties are better constrained at those sites as they have been measured there (e.g. Brock et al., 2010; Reid and Brock., 2010). At Miage, in particular, an extensive effort of field measurements since 2005 (Brock et al., 2010) has made this glacier one of the few where debris properties have been measured or directly derived from measurements (Table 1 and Table S1). Miage Glacier is thus a "literature" site, the properties of which have been used by numerous other studies (e.g. Carenzo et al., 2016) and for a number of sites in this intercomparison (Table S1).

6.1.2. Sites with poor performance

In contrast to the European sites, three sites stand out as low performance sites: Djankuat, Lirung and Tasman. The intermediate and energy balance models cannot reproduce the observed melt at any of these three sites (Fig. 5) nor surface temperature at Lirung (~~Fig. S3~~ Fig. 8b). The only models that achieve a good or reasonable performance are the calibrated temperature index models (Fig. 5, Table 3), which tune their parameter(s) to maximise agreement with observed melt; it remains to be investigated whether their performance would remain high over a distinct validation period. The uncalibrated Hyper-fit model performed well, but it used literature parameters from the same site. It is not straightforward to disentangle the reasons leading to reduced model skills at these three sites. A low model performance can be associated with either poor data (input and/or validation data), poor parameter values (debris properties), or a poor model, (i.e. lacking or failing in the representation of processes that are important at those sites). At the three sites, the clustering of model performance shown in Fig. 6 suggests that either all models miss a crucial process that is important at those sites or there is a common problem with the validation, forcing or debris properties data, which affects equally all models except for the calibrated temperature-index models.

6.1.3 Lirung study site and the difficulties in deriving thermal conductivity

Lirung is one of the few sites with a conductivity value estimated from field data. The debris thermal conductivity value however is very high (Table S1). It was derived from thermistor records of temperature at variable depths within the debris using three approaches: i) the method by Nawako and Young (1982) and Brock et al. (2010), which assumes a linear vertical temperature gradient within the debris; ii) the approach by Conway and Rasmussen (2000) based on the diffusion equation; and iii) the approach of Anderson (1998), which assumes a sinusoidal variation of temperature in time and an exponential decay of temperature in space. All methods were applied to data collected in 2013 at an AWS location (unpublished data), and the value provided for this intercomparison experiment was the average of the three values. The values obtained with each approach differed considerably and were higher than many literature values, but since there was no way to establish which method was best, the average was provided. This points to an irony, that at one of the only sites where data were collected to

estimate thermal conductivity, the values obtained through a devoted calculation may be inappropriate for modelling, pointing to a discrepancy between conductivity derived from field data and values needed by EB models, something that has also been suggested recently by Melo-Velasco et al. (2025). This suggests that: i) we do not know yet which is the best method to measure or derive debris conductivity in the field, directly or from other field observations (Melo-Velasco et al., 2025), and ii) simpler methods providing bulk values, such as the one by Brock et al. (2010), might be more suitable for the existing energy balance models (e.g. the DEB model of Reid and Brock, 2010), which have been developed for conductivity values derived in this way. Future efforts should therefore seek to devise methods to estimate debris thermal conductivity accurately and in a manner that is consistent with their use in EB models.

It should also be noted that on Lirung there is a difference of 14 cm between the debris thickness at the location of the automatic weather station (h = 30 cm) and the location of the ultrasonic depth gauge used as validation site (h = 44 cm). This might explain some of the divergence of models' outputs from the observations. As observed in the Østrem curves for Lirung (Fig. 10d), melt rates are lower (for all models) if thicker debris is used. We thus run one of the best performing energy balance models, DEB_{CF}, with the debris thickness at the ultrasonic depth gauge location, with the debris conductivity and surface roughness of Miage (k = 1.04 W m^{-1} K^{-1} and z_0=0.016 m, Brock et al. 2010), which were used by other sites as well (Table S1), and with a combination of these changed debris properties. The model simulations with debris thickness changed from 30 cm to 44 cm differ by 22.2% and show a better agreement with the observations (Table 5). The difference from the standard simulation is highest (42.7%) when we additionally consider the conductivity value from Miage, and slightly lower (38.5%) when we also consider surface roughness from Miage (Table 5). This combination reduces the total melt error of the DEB_{CF} model at Lirung from 103.8% to only 16.6%, demonstrating the large impact of inaccurate debris properties. The thicker debris decreases melt rate considerably (and delays its peak), and the smaller conductivity value also considerably reduces and delays the peak melt (Fig. S9). It is therefore a combination of at least these two factors (heterogeneous debris thickness between the automatic weather station and validation site, and uncertainty in the site conductivity) that likely explains the poor performance of all models at this site. Our sensitivity analysis thus confirms that deviations between the debris thickness at the location of the automatic weather station and the location of the sub-debris melt observations can have substantial implications for sub-debris melt modelling assessments.

Table 5. Sensitivity test of melt rates by the DEB_{CF} model at Lirung to substantially modified debris properties. Simulations with $h_d = 30$ cm (thickness at the automatic weather station location) and $h_d = 44$ cm (thickness at the ultrasonic depth gauge location), and with debris conductivity and surface roughness from Lirung ($k = 1.55$ and $z_0 = 0.035$) and Miage ($k_d = 1.04$ and $z_0 = 0.016$), as well as combinations of them. Debris properties underlined in *italics* denote the standard run (experiment 1) from this model intercomparison. Debris properties in **bold** denote the largest difference with the standard simulations.

Debris thickness, h_d (cm)	*30*				**44**			
Thermal conductivity, k (W m^{-2} K^{-1})	*1.55*		1.04		1.55		**1.04**	
Surface roughness, z_0 (m)	*0.035*	0.016	0.035	0.016	0.035	0.016	**0.035**	0.016
Total modelled melt (m w.e.)	*2.34*	2.48	1.78	1.91	1.82	1.94	**1.34**	1.44
Total modelled melt error (%)	*103.7*	116.4	55	66.8	58.4	69	**16.6**	26
Difference from standard simulation (%)	*0*	6.0	-23.9	-18.4	-22.2	-17.1	**-42.7**	-38.5
Changed debris properties	-	z_0	k	z_0, k	h_d	h_d, z_0	h_d, k	h_d, z_0, k

6.1.4 Tasman and Djankuat

At Tasman, most models are clustered together (Fig. 6i), but a cold bias in surface temperature (with median RMSE 3.8°C, Table S27) corresponds to a melt overestimation of about 58% (Table 3, Fig. 5). At this site, debris thermal conductivity is very high (k = 1.8 W m^{-1} K^{-1}) compared to literature values (Table S1). It was taken from Rohl (2008), who in turn used a

value from McSaveney et al. (1975) describing a pure mixture of rock and water without pore space. It is therefore inaccurate for the conductivity of a porous debris layer, and likely responsible for the melt overestimation and cold bias. Even though most models overestimate melt, an ablation stake at the same location measured higher total melt than that observed at the ultrasonic depth gauge (Fig. S2i), suggesting the melt overestimation may be lower than reported. Finally, we cannot exclude that at Tasman, which is the warmest and wettest of our sites (Fig. 2, Table S3), processes not included in the models, such as those related to the water content in the debris, might be playing a role.

Finally, at Djankuat, even the more empirical models, ~~in~~ both ~~their~~ calibrated and uncalibrated ~~versions~~, fail to match the observed melt, with the exception of the calibrated KM1. Djankuat is the site with the second thickest debris cover (61 cm, Table 1) and the one most similar to the European sites in terms of conditions (temperature and relative humidity, Fig. 2). For this site, debris properties, and conductivity in particular, clearly play a key role in explaining the model failure. Debris conductivity is extremely high (2.8 W m^{-1} K^{-1}), the highest of all sites (Table S1), and was taken from Bozhinskiy et al. (1986). Upon scrutiny, we realised that this is the conductivity of the rock material itself, and not that of a porous debris layer, which would be strongly reduced, reinforcing our conclusion that site-specific properties (representing the actual debris layer) are needed.

6.1.5 Knowledge of debris properties remains a key gap

An interesting site in comparison to all other sites is Changri Nup, where the overall performance is relatively high. At this site, the provided debris properties had been optimised to match observed melt with an energy-balance model not participating in the experiment (Table S1, Lejeune et al., 2013; Giese et al., 2020), thus explaining the high performance of most energy balance models (Fig. 5). The Changri Nup case exemplifies a relatively common strategy to determine debris properties by optimization, which seems a valid alternative when there are no reliable estimates from direct field measurements (Melo-Velasco et al., 2025). It seems particularly useful relative to literature values that may not be relevant (e.g. Tasman), or when direct methods provide divergent estimates with high uncertainty (e.g. Lirung). From all the cases considered in this intercomparison, and from the variety of approaches adopted to determine debris properties, it is apparent that debris properties are not well constrained at most sites, estimates from literature are often not appropriate, and even more importantly, that published methods to determine conductivity in the field (Conway and Rasmussen, 2000; Nicholson and Benn, 2012; Reid et al., 2012) may not agree, as exemplified by the case of Lirung, and confirmed by a separate study (Melo-Velasco et al., 2025). In addition, even if we are able to constrain debris properties at an individual automatic weather station, their values are affected by differences in porosity and pore water content across the debris-covered areas of a glacier, and therefore are also likely to vary considerably in time and space. Neither aspect has seen much investigation to date. Limited knowledge of the variability of debris properties also led to applying literature-derived uncertainty estimates to the debris properties (~10% for surface roughness and thermal conductivity), which are likely too narrow, leading to a resulting uncertainty in modelled melt smaller than the actual model error (Fig. 9). A lack of debris property data is also relevant for the $GRO17_B$ model, for which there are no data available of debris porosity and grain size, and which thus necessitated the calibration of the latent heat flux component. Knowledge of debris properties emerges thus from this intercomparison experiment as a key knowledge gap that the community should address, by both making a large community effort to compile and scrutinise existing estimates and datasets (e.g. Fontrodona-Bach et al., 2025), and by developing and thoroughly testing field methods to determine these crucial parameters.

6.2. Model performance, strengths and limitations

We have considered four groups of models ranging from empirical temperature-index models to physically-based energy balance models, with two models of intermediate complexity (Fig. 3). Our results indicate that model performance varies

greatly, reflecting structural choices, distinct parameterisations of fluxes and levels of complexity in the representation of processes (Fig. 3 and Table 4). Energy-balance models offer a variety of model structures, flux calculations, temporal resolution, numerical solutions and vertical discretisation of the debris domain, and consequently produce a large range of model performance. Temperature index models perform well when calibrated, with a performance similar - and superior in some cases - to that of the energy balance models, and poorly when uncalibrated, although it was not possible to assess their performance over an independent evaluation period.

The discussion below is guided by the ranking of models in Table 4. We note however that an objective and unambiguous ranking is difficult to obtain, and this evaluation is valid for one melt season and the sites available to this intercomparison. Model choices respond to data requirements and specific research questions, and therefore our discussion does not disqualify models from being used for a particular research question or a given data availability.

Energy-balance models

Large spread and variability among energy balance models' performance

There are clear differences among the energy-balance models, with a large variability and spread, as well as a general difficulty to model melt at some sites. Models differ substantially in their calculations of both latent and sensible turbulent heat fluxes (Fig. 11 and 12). The two models with the highest degree of complexity according to our definition (see Sect. 4), DEB_{CF} and ROU15, were the highest ranked models based on their melt performance at the end of the modelling period (Table 4). However, the third most complex model, $GRO17_B$, clearly overestimates melt, especially at sites with thick debris. The $GRO17_B$ model, based on Evatt et al. (2015), includes the heat convection within the porous debris layer. Including this process adds complexity and uncertainty due to two extra parameters - the wind speed attenuation constant and friction velocity (Fig. 3) - which are required for the calculation of the evaporative (latent) heat flux. To obtain realistic estimates of these parameters, they were calibrated against the bulk method of Nicholson and Benn (2006) for the calculation of turbulent heat fluxes (see Sect. S2.1, Table S6), but not to match melt observations as this was a requirement of this intercomparison experiment. The inclusion of this process is the main difference between $GRO17_B$ (including the latent heat flux within the debris) and $GRO17_A$ (not including the latent heat flux within the debris), and nevertheless showed no significant effect for experiment one of this intercomparison, with thickness prescribed based on actual measurements. This is likely because convection is most relevant for thinner debris than at the sites considered in this experiment (Evatt et al., 2015). Its effect for thin debris is evident at Piramide, Tapado and Tasman for the model runs of experiment 3 (Fig. 10). The large melt overestimation at sites with thick debris is likely due to other model features, including the assumption of a linear temperature gradient in the debris layer. The same assumption is also made by the THRED and A-Melt models (Table S18), which are also ranked lower than the other energy balance models (Table 4), and have lower performances against surface temperature observations (Table S27, Fig. S5), suggesting that the assumption of a linear temperature gradient may limit model accuracy, especially for thick debris and short time intervals (e.g. hourly).

Disentangling differences between the best performing models

Among the energy balance models, the models with the highest degree of complexity (DEB_{CF}, ROU15 and d2EB) show the best performance, excluding the problematic sites of Lirung, Tasman and Djankuat (see Sect. 6.2). These are also the models with the highest consistency of performance against the two independent datasets. We thus further explore differences between these three best ranking models, as well as the DEB_{PG}, which is similar to DEB_{CF} but has nevertheless a lower performance. The main differences between models are: i) the debris layer is discretized in a similar way in DEB_{CF} and ROU15, but differently in d2EB and DEB_{PG} (Table 2); ii) a snowmelt module is included in DEB_{CF} and ROU15, but not in the other two models; iii) the rain heat flux is included in DEB_{CF} and ROU15 but not in d2EB and DEB_{PG}; iv) the latent heat fluxes are

activated in an equivalent in manner in DEB_{CF} and ROU15, which both assume debris is saturated when it rains, but differently in d2EB, which uses a parameterisation based on surface temperature (Steiner et al., 2018) and in DEB_{PG}, which assumes debris is saturated when the relative humidity of the air is 100% (Table S16); and finally v) stability corrections are included in DEB_{CF} and DEB_{PG}, but not in ROU15 and d2EB (Table S15).

To evaluate the importance of each of these aspects, we explore the relative importance of these assumptions and flux implementations using the DEB_{CF} model, and compare the new simulations with the standard DEB_{CF} run. The heat flux due to rain is very small and neglecting it does not affect melt in any considerable manner at any of the sites (Table 6). Occasional snowmelt is present only at three sites, Arolla, Changri Nup and Pirámide, and important only at Changri Nup, where snow is present on the ground 8% of the time and snowmelt accounts for 47% of the ice- and snowmelt calculated with DEB_{CF} (Fig. S13). In Arolla, snow and snowmelt are negligible, and in Pirámide snowmelt is very small (Fig. S13). The manner in which the debris layer is discretised does not seem to play a key role. Table 6 shows the effect of using two alternative approaches to discretise the debris: i) the approach by DEB_{PG} and the one employed by Reid and Brock (2010). Both approaches lead to more numerous, thinner layers the thicker the debris, but neither approach has an effect on melt except for Tapado (5% difference), suggesting that the debris discretisation is slightly important only at thick debris (> 61 cm) sites.

Table 6. Results of changes in layer discretisation, LE calculations and assumption of precipitation flux obtained from running DEB_{CF} and compared to the standard run. Values in italics in the header are the debris thickness in cm.

% difference from DEB-model standard run	ARO	CN	DJA	LIR	MIA	PIR	SDF	TAP	TAS
h_d (cm)	*6*	*10*	*61*	*30*	*22*	*18*	*11*	*80*	*30*
1 cm layers	-0.02	0.00	0.78	0.52	0.37	0.03	0.03	5.35	0.41
DEB_{PG} layers (3 cm)	-0.11	-0.07	0.28	0.01	0.07	-0.01	-0.33	4.84	0.27
No precipitation flux	-0.06	-0.18	-0.04	0.23	0.00	-0.02	-0.01	0.00	-0.03
RH_s = 100% when RH_a = 100%	1.10	10.14	21.64	1.49	2.79	10.79	1.14	0.00	1.87

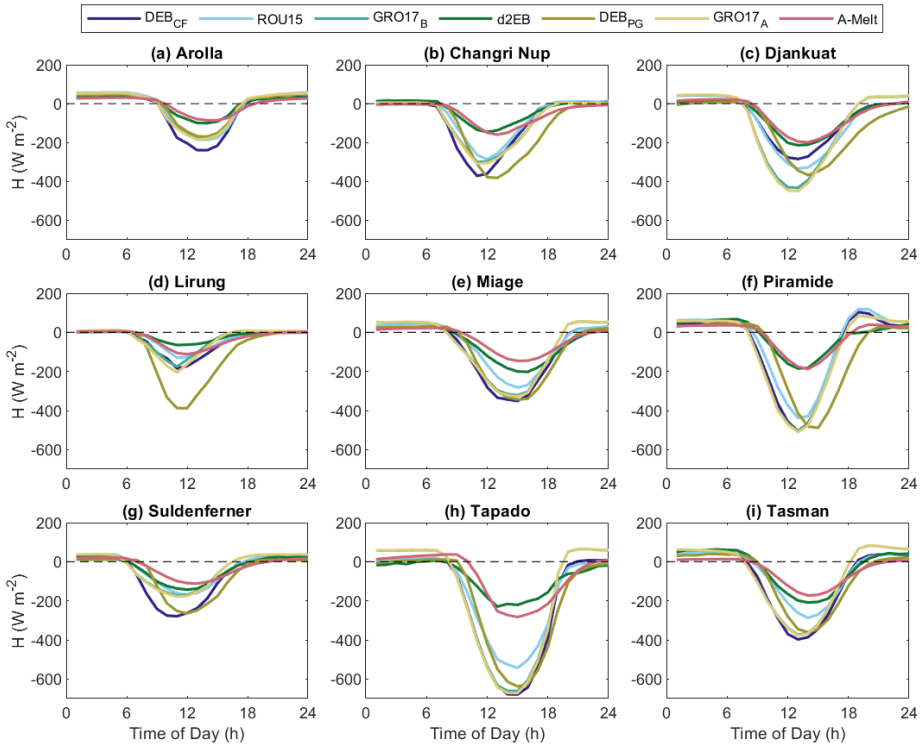


Figure 11. Average sub-daily cycle of the turbulent sensible heat energy flux (H) calculated by the energy balance models at all sites. Averages calculated over the entire simulation period are shown in Tables S28-S30. Note that fluxes are negative when going away from the surface.

The treatment of the turbulent fluxes seems instead a key element of distinction (Fig. 11 and 12). The use - or lack of - stability corrections can be a cause of large differences, but it was impossible in this experiment to quantify differences in total turbulent fluxes or total melt attributable to the use of stability corrections as models differed in too many other assumptions to be able to isolate this effect. We notice here the large differences in the turbulent sensible heat flux (Fig. 11) and call for an experiment that can systematically identify their causes (see Section 6.3).

Key differences are evident also when considering the latent heat flux. The two models that use the same theoretical approach, DEB_{CF} and DEB_{PG}, differ in the way the latent heat fluxes are activated and debris surface moisture is calculated. Since no observations of debris surface moisture were available, the two models used common (e.g. Rounce et al., 2015) but distinct assumptions. DEB_{CF} assumed that water was available to evaporate at all timesteps during rain events (i.e. relative humidity of the debris surface is 100% during rain events), an assumption also made by ROU15. DEB_{PG} instead assumed that there was water available to evaporate when the relative humidity of the air was 100%. As the latter situation barely occurred at our sites, the latent heat flux calculated by DEB_{PG} is almost always zero (Fig. 12). The assumption about the relative humidity of the debris surface seems far more important than the inclusion of stability corrections (Table 6, Fig. 12). The ROU15 and DEB_{CF} models, which differ in the manner the turbulent fluxes are calculated (Table S15 and S16) but use the same approach to determine the debris relative humidity (Table S16), produce similar turbulent fluxes at most sites, except for Djankuat (Fig. 12), which emerges from this intercomparison as a rather anomalous and problematic site (see Sect. 6.1).The d2EB model includes an empirical parameterisation to determine the relative humidity of the surface from the vapour pressure of the air (Steiner et al., 2018, and Sect. S2.1), which requires two empirical parameters originally estimated from 10 days of measurements on Lirung glacier in autumn, and used at all other sites. The latent heat flux simulations with this model are distinct from all others, and from DEB_{CF} and ROU15 in particular (Fig. 12).

We have no way to evaluate which of those formulations is the most correct besides doing so indirectly through assessment of melt and surface temperature accuracy. The fact that DEB_{CF} has one of the highest performances, and that some of the other characteristics of this model seem to have little influence (Table 6), seems to suggest that the approach to calculate the turbulent fluxes adopted in that model might be the most suitable for the whole range of sites based on the measured and assumed debris properties. Neglecting latent heat fluxes will likely lead to overestimation of the sensible heat fluxes (Collier et al., 2014; Nicholson and Stiperski, 2020), so it seems crucial to validate the flux calculations directly, and not indirectly as in this intercomparison experiment (see Sect. 6.3). Measurements of latent heat fluxes are becoming available (Steiner et al., 2018; Nicholson and Stiperski 2020), creating the opportunity for such an effort.

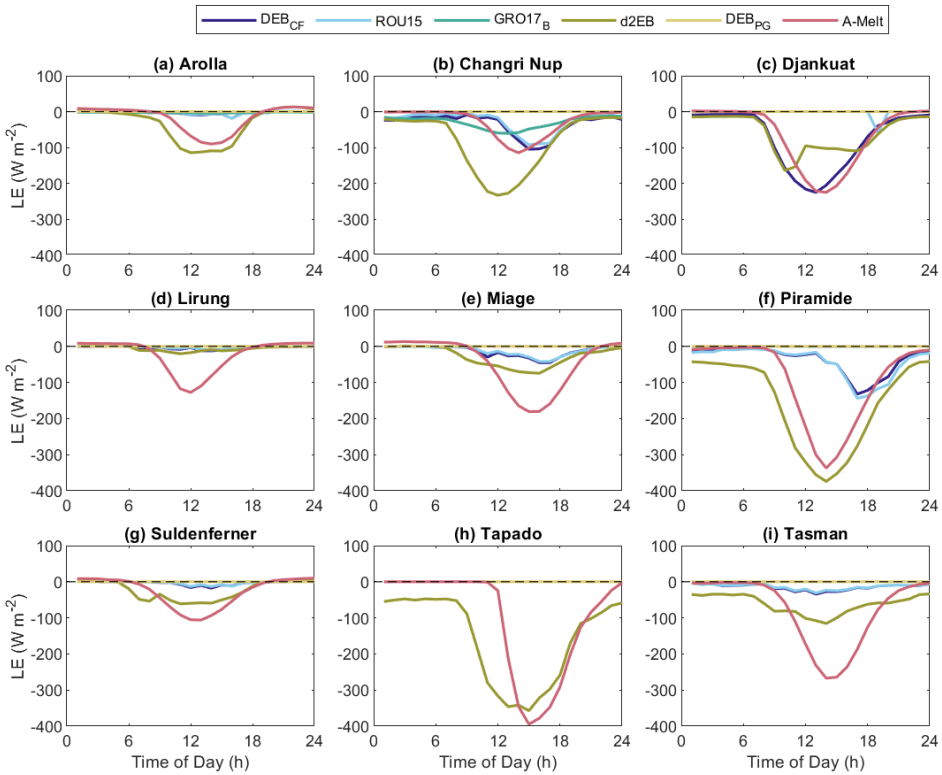


Figure 12. Average sub-daily cycle of the turbulent latent energy flux (LE) calculated by the energy balance models at all sites. Averages calculated over the entire simulation period are shown in Table S28-S30. Lines are not visible when the flux is close to zero and models overlap each other.

Temperature index models

Temperature index models offer computational efficiency. In this experiment, they ran on average 10^3-10^4 times faster than energy balance models (Table S26), a substantial advantage when used for spatially-distributed or long-term simulations. In addition, they require only a few, commonly available input data. We show that they perform well in reproducing daily melt rates when calibrated against melt observations from the same time period, but also note that no independent validation was possible. The Hyper-fit model shows that empirical models can perform reasonably well with parameters determined from independent sites and time periods on the same glacier. Since they rely on air temperature as input, they cannot reproduce sub-daily sub-debris melt rates accurately and the strong sub-daily variability that characterises melt rates under thin to moderate debris (Fig. S9). This is something that has been suggested already for debris-covered ice melt (Carenzo et al., 2016) and is well known for debris-free ice melt (e.g. Gabbi et al., 2014).

Our results show that calibration greatly improves the performance of temperature index models. The simpler the model, the more important its calibration, as most of the variability in melt rates and the processes controlling them is represented by the few parameters. The KM1 model performs considerably better than ~~its uncalibrated version~~ KP1, a very similar model which is uncalibrated (Fig. 7, Table 3). The Hyper-fit model has a stronger performance when calibrated than uncalibrated (Fig. 7, Table 3). The uncalibrated version uses parameters derived for the same glacier as in this intercomparison but for time periods distinct from those of the intercomparison. This differs from the other uncalibrated models which use a single parameter set applied to each study site. Nonetheless, the uncalibrated version of the Hyper-fit model reveals that parameters determined outside of the study period and from other locations on the same glacier can produce consistent high model skill (Fig. 7, Table 3), at least within the experimental setup of this model intercomparison. This means that for glaciers where Hyper-fit model parameters (DDF_{ice} and h*) have been constrained in the past, sub-debris melt may be modelled reasonably well (Table 3). This ability of the Hyper-fit model is possible because of the stability of the characteristic debris thickness, h* in space and time across individual glaciers.

The calibration strategy chosen by each modeller has a strong impact on model performance. How empirical models were calibrated was left to modellers as part of the experimental design. These models were calibrated both against the cumulative daily melt, sub-period melt and total melt, the latter being the metric we use for their evaluation (see Sect. 4.3). It is clear that if the error metric chosen to evaluate model output is different from what the model was tuned to then the performance would be different, and possibly reduced. Use of a different calibration metric might result also in distinct parameter values, something to remember when considering parameter transferability. As an example, we recalibrated the KM1 model based on daily melt, rather than total melt, and this resulted in lower performance against total melt at many sites (Fig. S12), but an increased ability to reproduce the variability during the period of simulations (e.g. Changri Nup, Lirung), suggesting that at some sites relatively accurate calculation of total melt might result from compensation of errors in the daily melt rates.

Models of intermediate complexity

T~~he t~~wo models of intermediate complexity, SEB and ETI, perform similarly to that of the best energy balance models (Table 4 and Fig. 7). Both models calculate net shortwave radiation flux using identical inputs, with the DETI having three empirical parameters and the SEB using two empirical parameters and debris conductivity to calculate heat conduction through the debris (Fig. 3). A strength of both models is their ability to reproduce the sub-daily cycle of melt rates, similarly to energy balance models (Fig. S9), an ability that derives from inclusion of the shortwave radiation balance. Crucially, both models require calibration against variables that exhibit sub-daily variability. However, the uncalibrated $DETI_m$, using parameters from Carenzo et al. (2016) for Miage glacier, performs comparably to its calibrated version across sites (Fig. 7, Table 3). This suggests ~~suggests~~ that the parameter values derived from Carenzo et al. (2016) are fairly transferable across sites and that calibration may not significantly enhance performance if the energy balance model to calibrate against is suboptimal.

6.3. Future perspectives

Cooperation between a large number of scientists has allowed this intercomparison to provide a robust assessment of different approaches to modelling sub-debris melt at glaciers across the world. It has also highlighted model and data limitations and provided insight into what data should be collected to enable a future model intercomparison focused on the physical processes that control melt in different climatic settings.

A new vision for old measurements

While it is clear that the growing debris-covered glaciers community is making a large effort to collect data at a variety of sites (Fontrodona-Bach et al., 2025), it is still young in terms of data collection and standardisation. Meteorological data quality has been a major issue as no data provider was able to establish the measurement accuracy beyond the sensor accuracy. It seems important to design observations where both the accuracy of the measurements and the actual meaning of a measurement are clearly defined. Despite recent advances in deriving debris properties (Rounce et al., 2015; Quincey et al., 2017; Miles et al., 2017, McCarthy et al., 2017; Nicholson and Mertes, 2017; Nicholson et al., 2018, Bisset et al., 2022; Messmer and Groos, 2024), debris properties remain a very major gap in knowledge, as discussed in Sect. 6.1. Comparable, standardised field and laboratory measurements should be designed to determine debris properties, and debris thermal conductivity in particular, and their variability in space, time and with depth, under distinct meteorological and surface conditions.

A discrepancy that should be considered is that often the three sets of data needed for model simulations (input meteorological data, debris properties and validation data) are measured at different locations. Considering the high heterogeneity of debris thickness and properties even within short distances (e.g. Nicholson and Benn, 2012), it is important to account for these spatial discrepancies. For instance, the point where automatic weather station sensors 'look at' is often metres distant from the ablation stake or ultrasonic depth gauge measuring melt. Melt records are often measured at only one site because of the effort required

and can be non-representative of the area around automatic weather stations. This has been exemplified by the case of Lirung, where differences in debris thickness between the location of the AWS and ultrasonic depth gauge affected the comparison of modelled and observed melt rates.

Recommendations for novel measurements

Despite the strong need for standardisation and improvements in the quality of existing measurements, more routine measurements and estimates of their uncertainty, it seems also crucial that new datasets are collected that allow both internal validation of existing models and model developments. Measurements of internal variables allowing for the testing of state variables and intermediate fluxes simulated by energy balance models (e.g. debris temperature profiles, water content, single fluxes estimates) should be collected. What has been called "internal validation" is since a decade or longer widely recognised and accepted in hydrological research, and increasingly also in glaciological mass balance and runoff models (Huss et al. 2008, Ragettli et al., 2015). It should be adopted widely also by the debris-covered glaciers research community.

It seems also crucial to collect accurate datasets at sub-daily resolution to evaluate the performance of all models, and energy balance models in particular, that can work at high temporal resolution, as part of this strategy of internal validation and to avoid equifinality problems (Gabbi et al., 2014). Ultrasonic depth gauge (UDG) records do in theory offer sub-daily records, but these are noisy and there is little agreement as to how those datasets should be treated and processed, nor about what is their actual accuracy at hourly or higher resolution time intervals. Supplementary information (e.g photos of the station/UDG, measurements from the surface to UDG at start and end of the data set) are needed to confirm the accuracy of UDG data. A systematic collection of additional independent validation datasets such as within-debris temperature and measurements of debris moisture seem important (Sakai et al., 2004) to allow validation of internal processes and thus testing of individual model components, and not only bulk simulations. Community compilation of manuals of best practice guidance for observations may be a useful way forward to increase the systematic collection of relevant data.

We have shown that there are still major knowledge gaps, and models are in need of further testing and developments, even in a field relatively well established such as melt modelling at the point scale. At this scale, the uncertainties associated with the spatial variability of meteorological data and debris properties are removed, and still uncertainties are large. It seems particularly important to design acquisition of observations to test the model component proposed by Evatt et al. (2015) to account for air flow and heat convection into a porous debris layer and to allow evaluation of the schemes designed to estimate the turbulent heat fluxes (with e.g. eddy-covariance measurements as in Collier et al., 2014). Similarly, water flow through the debris matrix should be accounted for (Fyffe et al. 2020), as a saturated layer thickens and thins with daily melt cycles and in response to warm and cold weather, causing temporal variations in both debris conductivity and the evaporative flux.

A key aspect of investigation that is left open is a thorough understanding of the relative importance of processes. This should be systematically assessed for both distinct debris thicknesses (and thus debris properties) and climate conditions. We were able to show that the conditions for latent heat flux activation are more important at some sites (Changri Nup, Djanukuat and Pirámide, with errors of 10 to 25% in total melt, Table 6) than at others. Those three sites however are remarkably distinct, and include: i) a high elevation, cold, humid (80% average relative humidity) site with relatively thin debris (10 cm), Changri Nup; ii) a relatively warm site with average relative humidity of 60% and thick debris (61 cm); and iii) a relatively cold, very dry (30% average relative humidity) Andean site with 18 cm debris. It is thus almost impossible to identify which characteristics make these three sites sensitive to changes in the parameterisation of debris surface moisture content. A devoted field experiment, where each of those factors (debris thickness, debris moisture and water flow, climatic controls) are controlled, should be designed to disentangle the driving factors.

An aspect of model performance that has emerged from this intercomparison as worth further investigation is the dependence of model skills as a function of debris thickness. With increasing debris thickness, processes such as heat storage and conductivity might become more relevant, while others such as the air flow within the debris and associated convective heat exchange with the underlying ice (Evatt et al., 2015) might lose importance. Both thick and thin debris should be investigated at sites if possible. It is also clear from our analysis that at some sites, and Lirung in particular, thin debris is where models diverge most, and we are not able to constrain an Østrem curve with a model ensemble. Areas of thin debris, usually located in the upper region of the debris-covered area, is where maximum rates of sub-debris ablation are predicted (Fyffe et al., 2020) and where determining ablation rates with confidence is critical to pin down the interplay between melt and ice flux in determining the glacier profile (e.g. Nicholson and Benn, 2012). This consideration should guide data acquisition to explain the model divergence for thin debris across sites.

Suggestions for future model developments

The models that joined this intercomparison experiment reflect the state-of-the-art of modelling skills and structures available to the community. There is still large potential for model improvements, however. No model in this intercomparison includes freezing/refreezing and the cold content of the frozen porous debris layer. Although this may not be relevant during the ablation season, it may be important at high elevation sites and during shoulder seasons (Giese et al., 2020). On Lirung Glacier, for instance, Steiner et al. (2021) demonstrated that during up to a week in spring all energy available was used to defrost the debris. A second crucial limitation of all models is that none of them considers water content in the debris, which has been shown to impact modelled melt volumes (Collier et al. 2014). A key aspect related to this is the lack of observations of debris water content, either for model development or validation. Refreezing and water content in the debris should therefore be included in future model developments (e.g. Winter-Billington et al. 2022).

Snow has been neglected from this intercomparison, which as a first step and to allow the inclusion of the largest possible number of models, has focused on the ablation season. For calculation of mass balance and long-term glacier changes, however, it is important to run models for an entire year and then for multiple years. This would require inclusion of snow accumulation and melt above the debris layer.

Finally, this intercomparison has evidenced the need to assess how important the heat convection into a dry porous layer is (Wicky and Hauck, 2020), when it occurs, and at which sites it is dominant. This would allow evaluating the relative importance of the model developments suggested by Evatt et al. (2015).

Limitations and recommendations for Phase II

There are a number of limitations to our study, which could be addressed in a phase II of this experiment. First, we have evaluated model skills against low temporal resolution melt observations, daily at best, which cannot reproduce the diurnal cycle of melt, and from sensors that might have considerable errors. The results are striking, as the highest ranked models are calibrated temperature-index models (Hyper-fit, KM1 and DDF_{debris}, Table 4) that cannot reproduce the daily cycle because of intrinsic structural limitations (Fig. S9). While at the seasonal scale the sub-daily variability may not matter, model intercomparisons should ideally be carried out at a variety of scales, which each correspond to a distinct model ability and model purpose (e.g. mass balance versus runoff simulations). It also points to the fact that we do not have in most cases measurements of high resolution available, and improved melt measurements at daily resolution (rather than weekly/seasonal stake measurements) would be beneficial.

Second, this study has made a large effort to gather existing data sets across the globe so as to include in the comparison as many sites with distinct debris and climate characteristics as possible. However, all our conclusions rely on data from only one melt season at each of the nine sites. Although some uncalibrated temperature-index models performed well using literature

parameters, this limitation did not allow testing the model robustness and parameter transferability in time over multiple melt seasons. To fully assess and equally compare the performance and transferability of empirical temperature-index models and energy balance models, and their robustness in space and time in particular, we suggest that a follow-up effort should test the model performance of calibrated models over an independent validation period. In general, model simulations should be tested over multiple years to assess their robustness and how the errors we identified over one melt season propagate (or not) over the temporal scales needed for mass balance change calculations.

Third, it was apparent from this intercomparison that the broad variety of choices in model structures, coding languages, debris discretisation, parameterisations, temporal resolutions and assumptions masked at times actual differences in representation of processes. This hindered an unambiguous identification of strengths and limitations, as too many differences in model choices did not allow to identify the actual reasons for distinct model performances. A lesson learned from our experiment is thus that a consistent modelling framework including multiple parameterisations and modelling assumptions is needed to identify the importance of individual processes and assumptions. We would thus advocate for such a modelling framework that includes the suites of models tested in this experiment, as well as the processes we identified as missing above.

7. Conclusions

We have compared 15~~4~~ models of different complexity to simulate glacier ice melt under debris (Fig. 3), including energy balance models, temperature index models and models of intermediate complexity. The models were run at 9 sites across the globe, for one melt season of variable duration, and validated against sub-debris ice melt and surface temperature observations. Our main conclusions are as follows:

- Energy balance models and empirical temperature index models perform in a distinct manner and serve distinct purposes. ~~In general,~~ Temperature index models perform ~~very well (median performance)~~ best in this experiment, and better than energy balance models, when calibrated, and poorly (worse than energy balance models) when applied in their uncalibrated form. However, ~~the Hyper-fit model and~~ the DETI model shows one of the best performances with parameters from one site applied to all other sites. The Hyper-fit model shows that site-specific ~~literature~~ estimated parameters ~~determined outside of the study period~~ (i.e. parameters from the same site but outside of the study period) can produce ~~viable~~ satisfactory melt estimates.

- Energy balance models show a range of performance and model skills~~.~~, with an overall overestimation of melt and relatively large errors (20.6% median, 44.8% MAE) when all sites are considered, and a slight underestimation of melt and much smaller errors (-6.3% median, 18.7% MAE) when the three sites with poorly constrained debris properties are not considered. A clear finding from this work is that the energy balance models that perform best are those with the highest degree of complexity at the debris-atmosphere interface. We were not able to demonstrate the added value of additional complexity within the debris, because of lack of data representative of processes within the debris layer. The use of simplifying assumptions (and of a linear temperature profile within the debris in particular) within the model that included convection in the debris made it difficult to disentangle the importance of this process.

- A striking result of our intercomparison is that models perform well at the three Alpine sites of Arolla, Miage and Suldenferner, and consistently poorly at Djankuat, Lirung and Tasman. While a variety of reasons contribute to explain this inter-site variability in model skill, we have been able to clearly identify variable knowledge of debris properties, and thermal conductivity in particular, as a key cause.

- Debris properties are a major control on melt simulations and model performance. We thus need estimates of debris properties (especially surface roughness and thermal conductivity) of high accuracy. Crucially, we currently are not

able to constrain them using measured values at most sites. The uncertainty ranges of debris properties typically used in the literature are insufficient for most models to encompass observed melt values. Efforts should be devoted to measure debris properties in the field and establish their variability in time, space and with depth.

- Despite tremendous advances in recent years, we showed that sub-debris melt models still need to be improved, and important model developments are needed. Most are not able to cope with debris-snow interactions, moisture in the debris is not accounted for and refreezing rarely represented.

We suggest that a follow-up to this intercomparison should focus on systematically improving a single model framework (written in the same language, with the same assumptions and main schemes) to understand the effects of individual processes such as debris moisture related processes (phase change, percolation, conduction properties change) and convection within the debris. This would allow to both: i) determine the most appropriate representations and parameterisations of each process; and ii) understand where the balance lies between increasing the number of parameters and achieving substantial improvements in model skills, thus also avoiding equifinality problems and error compensations that might be related to a too high number of parameters. Adding more complex process representation bears the risk of transferring uncertainty into parameters that might be impossible to measure. It also seems imperative to establish a hierarchy of processes in terms of their importance, and to focus on advanced, robust representations of those. A way forward is thus to develop a flexible, modular model that would allow testing distinct calculations and parameterisations within the same model structure, allowing comparisons of single model components.

Our main conclusions are that models need to be improved and knowledge of debris properties substantially advanced. Data collection is often regarded as functional more to model validation than development. We suggest that data collection and model development should be closely coupled, iteratively informing each other - something that only a truly community, cooperative effort might afford. In addition to the model development needs identified above, we need to clearly identify community priorities and methodologies for making debris measurements.

Data Availability. All data were stored on the Zenodo community on debris-covered glaciers (zenodo.org/communities/iacswgondcgs) set up by the working group on debris-covered glaciers from the International Association of Cryospheric Sciences. The data are publicly accessible at:

- https://zenodo.org/records/3047649 (Arolla)
- https://zenodo.org/records/3048780 (Changri Nup)
- https://zenodo.org/records/3049871 (Djankuat) (Rets et al. 2019)
- https://zenodo.org/records/3050327 (Lirung)
- https://zenodo.org/records/3050557 (Miage)
- https://zenodo.org/records/3056072 (Piramide)
- https://zenodo.org/records/3056524 (Suldenferner)
- https://zenodo.org/records/3362402 (Tapado)
- https://zenodo.org/records/3354105 (Tasman)

Model simulations were submitted individually and are also available on Zenodo at https://doi.org/10.5281/zenodo.15754455.

Author contributions. FP, AFB, DR, CF are the core authors. The rest of the authors are in alphabetical order. FP and DR had the initial project idea, developed the experimental set up and recruited modellers and data providers. AFB and FP coordinated

the project. FP wrote most of the manuscript and gave constant input to the data analysis and manuscript preparation. AFB managed the experiment data, did most data analyses, and contributed to writing the manuscript. CF contributed to the data analysis. LA, AFB, KF, CF, PG, ARG, WI, CM, MM, ER, DR, AS, JS and AWB provided model simulations. AA, BB, WI, SF, SM, JM, EM, HP, TS, JS, and PW provided observational data. AFB, FP and ER prepared and processed observational data. All authors commented on the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 772751, RAVEN, "Rapid mass losses of debris covered glaciers in High Mountain Asia". It was also supported by the SNSF RENOIR project "Resolving the thickness of debris on Earth's glaciers and its rate of change (RENOIR)", project number 204322.

David Rounce received support from NASA-ROSES program grants NNX17AB27G and 80NSSC17K0566.

Walter Immerzeel and Jakob Steiner acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 676819).

Ben Brock acknowledges support from the EU/FP7 ACQWA (Assessing Climate impacts on the Quantity and quality of WAter) project, NERC grant NE/C514282/1, the British Council-Italian Ministry of University and Research Partnership programme and the Carnegie Trust for the Universities of Scotland.

The authors acknowledge the International Association of Cryospheric Sciences (IACS) for supporting the creation of the Debris-Covered Glaciers Working Group (DCG-WG) which enabled this model intercomparison experiment.

The authors thank Martin Heynen for producing figures 3 and 4.

The authors thank Duncan Quincey and an anonymous reviewer for their constructive feedback and comments.

References

Anderson, R. S.: Near-surface thermal profiles in alpine bedrock: Implications for the frost weathering of rock, Arctic and Alpine Research, 30(4), 362–372, 1998.

Anderson, L. S. and Anderson, R. S.: Modeling debris-covered glaciers: Response to steady debris deposition, The Cryosphere, 10(3), 1105–1124, doi:10.5194/tc-10-1105-2016, 2016.

Anderson, L. S. and Anderson, R. S.: Debris thickness patterns on debris-covered glaciers, Geomorphology, 311, 1–12, doi:10.1016/j.geomorph.2018.03.014, 2018.

Anderson, L. S., Armstrong, W. H., Anderson, R. S., and Buri, P.: Debris cover and the thinning of Kennicott Glacier, Alaska: In situ measurements, automated ice cliff delineation and distributed melt estimates, The Cryosphere, 15(1), 265–282, doi:10.5194/tc-15-265-2021, 2021.

Anslow, F. S., Hostetler, S., Bidlake, W. R., and Clark, P. U.: Distributed energy balance modeling of South Cascade Glacier, Washington and assessment of model uncertainty, Journal of Geophysical Research: Earth Surface, 113(F2), 2007JF000850, doi:10.1029/2007JF000850, 2008.

Anslow, F. S., Hostetler, S., Bidlake, W. R., and Clark, P. U.: Distributed energy balance modeling of South Cascade Glacier, Washington and assessment of model uncertainty, Journal of Geophysical Research: Earth Surface, 113(F2), 2007JF000850, doi:10.1029/2007JF000850, 2008.

Benn, D. I., Bolch, T., Hands, K., Gulley, J., Luckman, A., Nicholson, L. I., Quincey, D., Thompson, S., Toumi, R., and Wiseman, S.: Response of debris-covered glaciers in the Mount Everest region to recent warming, and implications for outburst flood hazards, Earth-Science Reviews, 114(1–2), 156–174, doi:10.1016/j.earscirev.2012.03.008, 2012.

Bhambri, R., Bolch, T., and Chaujar, R. K.: Mapping of debris-covered glaciers in the Garhwal Himalayas using ASTER DEMs and thermal data, International Journal of Remote Sensing, 32(23), 8095–8119, doi:10.1080/01431161.2010.532821, 2011.

Bisset, R. R., Nienow, P. W., Goldberg, D. N., Wigmore, O., Loayza-Muro, R. A., Wadham, J. L., Macdonald, M. L., and Bingham, R. G.: Using thermal UAV imagery to model distributed debris thicknesses and sub-debris melt rates on debris-covered glaciers, Journal of Glaciology, 69(276), 981–996, doi:10.1017/jog.2022.116, 2023.

Bonekamp, P. N. J., Van Heerwaarden, C. C., Steiner, J. F., and Immerzeel, W. W.: Using 3D turbulence-resolving simulations to understand the impact of surface properties on the energy balance of a debris-covered glacier, The Cryosphere, 14(5), 1611–1632, doi:10.5194/tc-14-1611-2020, 2020.

Bozhinskiy, A. N., Krass, M. S., and Popovnin, V. V.: Role of Debris Cover in the Thermal Physics of Glaciers, Journal of Glaciology, 32(111), 255–266, doi:10.3189/S0022143000015598, 1986.

Brock, B. W., Mihalcea, C., Kirkbride, M. P., Diolaiuti, G., Cutler, M. E. J., and Smiraglia, C.: Meteorology and surface energy fluxes in the 2005–2007 ablation seasons at the Miage debris-covered glacier, Mont Blanc Massif, Italian Alps, Journal of Geophysical Research: Atmospheres, 115(D9), 2009JD013224, doi:10.1029/2009JD013224, 2010.

Brook, M., Hagg, W., and Winkler, S.: Debris cover and surface melt at a temperate maritime alpine glacier: Franz Josef Glacier, New Zealand, New Zealand Journal of Geology and Geophysics, 56(1), 27–38, doi:10.1080/00288306.2012.736391, 2013.

Brun, F., Buri, P., Miles, E. S., Wagnon, P., Steiner, J., Berthier, E., Ragettli, S., Kraaijenbrink, P., Immerzeel, W. W., and Pellicciotti, F.: Quantifying volume loss from ice cliffs on debris-covered glaciers using high-resolution terrestrial and aerial photogrammetry, Journal of Glaciology, 62(234), 684–695, doi:10.1017/jog.2016.54, 2016.

Brun, F., Wagnon, P., Berthier, E., Jomelli, V., Maharjan, S. B., Shrestha, F., and Kraaijenbrink, P. D. A.: Heterogeneous Influence of Glacier Morphology on the Mass Balance Variability in High Mountain Asia, Journal of Geophysical Research: Earth Surface, 124(6), 1331–1345, doi:10.1029/2018JF004838, 2019.

Brun, F., Wagnon, P., Berthier, E., Shea, J. M., Immerzeel, W. W., Kraaijenbrink, P. D. A., Vincent, C., Reverchon, C., Shrestha, D., and Arnaud, Y.: Ice cliff contribution to the tongue-wide ablation of Changri Nup Glacier, Nepal, central Himalaya, The Cryosphere, 12(11), 3439–3457, doi:10.5194/tc-12-3439-2018, 2018.

Buri, P., Miles, E. S., Steiner, J. F., Immerzeel, W. W., Wagnon, P., and Pellicciotti, F.: A physically based 3-D model of ice cliff evolution over debris-covered glaciers, Journal of Geophysical Research: Earth Surface, 121(12), 2471–2493, doi:10.1002/2016JF004039, 2016.

Buri, P. and Pellicciotti, F.: Aspect controls the survival of ice cliffs on debris-covered glaciers, Proceedings of the National Academy of Sciences, 115(17), 4369–4374, doi:10.1073/pnas.1713892115, 2018.

Buri, P., Pellicciotti, F., Steiner, J. F., Miles, E. S., and Immerzeel, W. W.: A grid-based model of backwasting of supraglacial ice cliffs on debris-covered glaciers, Annals of Glaciology, 57(71), 199–211, doi:10.3189/2016AoG71A059, 2016.

Carenzo, M., Pellicciotti, F., Mabillard, J., Reid, T., and Brock, B. W.: An enhanced temperature index model for debris-covered glaciers accounting for thickness effect, Advances in Water Resources, 94, 457–469, doi:10.1016/j.advwatres.2016.05.001, 2016.

Collier, E., Nicholson, L. I., Brock, B. W., Maussion, F., Essery, R., and Bush, A. B. G.: Representing moisture fluxes and phase changes in glacier debris cover using a reservoir approach, The Cryosphere, 8(4), 1429–1444, doi:10.5194/tc-8-1429-2014, 2014.

Conway, H. and Rasmussen, L. A.: Summer temperature profiles within supraglacial debris on Khumbu Glacier, Nepal, IAHS PUBLICATION, 89–98, 2000.

Deline, P.: Change in surface debris cover on Mont Blanc massif glaciers after the'Little Ice Age' termination, The Holocene, 15(2), 302–309, doi:10.1191/0959683605hl809rr, 2005.

Elagina, N., Rets, E., Korneva, I., Toropov, P., and Lavrentiev, I.: Simulation of mass balance and glacial runoff of Mount Elbrus from 1984 to 2022, Hydrological Sciences Journal, doi:10.1080/02626667.2025.2516080, 2025.

Evatt, G. W., Abrahams, I. D., Heil, M., Mayer, C., Kingslake, J., Mitchell, S. L., Fowler, A. C., and Clark, C. D.: Glacial melt under a porous debris layer, Journal of Glaciology, 61(229), 825–836, doi:10.3189/2015JoG14J235, 2015.

Farinotti, D., Brinkerhoff, D. J., Clarke, G. K. C., Fürst, J. J., Frey, H., Gantayat, P., Gillet-Chaulet, F., Girard, C., Huss, M., Leclercq, P. W., Linsbauer, A., Machguth, H., Martin, C., Maussion, F., Morlighem, M., Mosbeux, C., Pandit, A., Portmann, A., Rabatel, A., ... and Andreassen, L. M.: How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment, The Cryosphere, 11(2), 949–970, doi:10.5194/tc-11-949-2017, 2017.

Fontrodona-Bach, A., Groeneveld, L., Miles, E., McCarthy, M., Shaw, T., Melo Velasco, V., and Pellicciotti, F.: DebDab: A database of supraglacial debris thickness and physical properties, Earth Syst. Sci. Data Discuss. [preprint], doi:10.5194/essd-2024-559, accepted for publication, 2025.

Fujita, K. and Ageta, Y.: Effect of summer accumulation on glacier mass balance on the Tibetan Plateau revealed by mass-balance model, Journal of Glaciology, 46(153), 244–252, doi:10.3189/172756500781832945, 2000.

Fujita, K. and Sakai, A.: Modelling runoff from a Himalayan debris-covered glacier, Hydrology and Earth System Sciences, 18(7), 2679–2694, doi:10.5194/hess-18-2679-2014, 2014.

Fyffe, C. L., Reid, T. D., Brock, B. W., Kirkbride, M. P., Diolaiuti, G., Smiraglia, C., and Diotri, F.: A distributed energy-balance melt model of an alpine debris-covered glacier, Journal of Glaciology, 60(221), 587–602, doi:10.3189/2014JoG13J148, 2014.

Fyffe, C. L., Brock, B. W., Kirkbride, M. P., Mair, D. W. F., Arnold, N. S., Smiraglia, C., Diolaiuti, G., and Diotri, F.: Do debris-covered glaciers demonstrate distinctive hydrological behaviour compared to clean glaciers?, Journal of Hydrology, 570, 584–597, doi:10.1016/j.jhydrol.2018.12.069, 2019.

Fyffe, C. L., Woodget, A. S., Kirkbride, M. P., Deline, P., Westoby, M. J., and Brock, B. W.: Processes at the margins of supraglacial debris cover: Quantifying dirty ice ablation and debris redistribution, Earth Surface Processes and Landforms, 45(10), 2272–2290, doi:10.1002/esp.4879, 2020.

Gabbi, J., Carenzo, M., Pellicciotti, F., Bauder, A., and Funk, M.: A comparison of empirical and physically based glacier surface melt models for long-term simulations of glacier response, Journal of Glaciology, 60(224), 1140–1154, doi:10.3189/2014JoG14J011, 2014.

Gabbi, J., Huss, M., Bauder, A., Cao, F., and Schwikowski, M.: The impact of Saharan dust and black carbon on albedo and long-term mass balance of an Alpine glacier, The Cryosphere, 9(4), 1385–1400, doi:10.5194/tc-9-1385-2015, 2015.

Giese, A., Boone, A., Wagnon, P., and Hawley, R.: Incorporating moisture content in surface energy balance modeling of a debris-covered glacier, The Cryosphere, 14(5), 1555–1577, doi:10.5194/tc-14-1555-2020, 2020.

Groos, A. R., Mayer, C., Smiraglia, C., Diolaiuti, G. and Lambrecht, A.: A first attempt to model region-wide glacier surface mass balances in the Karakoram: Findings and future challenges, Geografia Fisica e Dinamica Quaternaria, 40, 137–159, doi:10.4461/GFDQ.2017.40.10, 2017.

Hagg, W., Mayer, C., Lambrecht, A., and Helm, A.: Sub-debris melt rates on southern inylchek glacier, central tian shan, Geografiska Annaler: Series A, Physical Geography, 90(1), 55–63, doi:10.1111/j.1468-0459.2008.00333.x, 2008.

Han, H., Wang, J., Wei, J., and Liu, S.: Backwasting rate on debris-covered Koxkar glacier, Tuomuer mountain, China, Journal of Glaciology, 56(196), 287–296, doi:10.3189/002214310791968430, 2010.

Herreid, S. and Pellicciotti, F.: The state of rock debris covering Earth's glaciers, Nature Geoscience, 13(9), 621–627, doi:10.1038/s41561-020-0615-0, 2020.

Huss, M., Farinotti, D., Bauder, A., and Funk, M.: Modelling runoff from highly glacierized alpine drainage basins in a changing climate, Hydrological Processes, 22(19), 3888–3902, doi:10.1002/hyp.7055, 2008.

Immerzeel, W. W., Kraaijenbrink, P. D. A., Shea, J. M., Shrestha, A. B., Pellicciotti, F., Bierkens, M. F. P., and De Jong, S. M.: High-resolution monitoring of Himalayan glacier dynamics using unmanned aerial vehicles, Remote Sensing of Environment, 150, 93–103, doi:10.1016/j.rse.2014.04.025, 2014.

Immerzeel, W. W., Pellicciotti, F., and Bierkens, M. F. P.: Rising river flows throughout the twenty-first century in two Himalayan glacierized watersheds, Nature Geoscience, 6(9), 742–745, doi:10.1038/ngeo1896, 2013.

Immerzeel, W. W., Van Beek, L. P. H., Konz, M., Shrestha, A. B., and Bierkens, M. F. P.: Hydrological response to climate change in a glacierized catchment in the Himalayas, Climatic Change, 110(3–4), 721–736, doi:10.1007/s10584-011-0143-4, 2012.

Juen, M., Mayer, C., Lambrecht, A., Han, H., and Liu, S.: Impact of varying debris cover thickness on ablation: A case study for Koxkar Glacier in the Tien Shan, The Cryosphere, 8(2), 377–386, doi:10.5194/tc-8-377-2014, 2014.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., and Arnaud, Y.: Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, Nature, 488(7412), 495–498, doi:10.1038/nature11324, 2012.

Kayastha, R. B., Takeuchi, Y., Nakawo, M., and Ageta, Y.: Practical prediction of ice melting beneath various thickness of debris cover on Khumbu Glacier, Nepal, using a positive degree-day factor, 264, 71–81, 2000.

Khan, M. I.: Ablation on Barpu Glacier, Karakoram Himalaya, Pakistan: A Study of Melt Processes on a Faceted, Debris-Covered Ice Surface [Master's Thesis], Wilfrid Laurier University, 1989.

Kirkbride, M. P. and Deline, P.: The formation of supraglacial debris covers by primary dispersal from transverse englacial debris bands, Earth Surface Processes and Landforms, 38(15), 1779–1792, doi:10.1002/esp.3416, 2013.

Kirkbride, M. P. and Dugmore, A. J.: Glaciological response to distal tephra fallout from the 1947 eruption of Hekla, south Iceland, Journal of Glaciology, 49(166), 420–428, doi:10.3189/172756503781830575, 2003.

Kirkbride, M. P., Sherriff, S. C., Rowan, A. V., Egholm, D. L., Quincey, D. J., Miles, E., Hubbard, B., and Miles, K.: Provenance and transport of supraglacial debris revealed by variations in debris geochemistry on Khumbu Glacier, Nepal Himalaya, Earth Surface Processes and Landforms, 48(14), 2737–2753, doi:10.1002/esp.5657, 2023.

Konovalov, V.: Computations of melting under moraine as a part of regional modelling of glacier runoff, Debris-Covered Glaciers; IAHS Publication, 264, 109–118, 2000.

Kraaijenbrink, P. D. A., Bierkens, M. F. P., Lutz, A. F., and Immerzeel, W. W.: Impact of a global temperature rise of 1.5 degrees Celsius on Asia's glaciers, Nature, 549(7671), 257–260, doi:10.1038/nature23878, 2017.

Kraaijenbrink, P. D. A., Shea, J. M., Litt, M., Steiner, J. F., Treichler, D., Koch, I., and Immerzeel, W. W.: Mapping Surface Temperatures on a Debris-Covered Glacier With an Unmanned Aerial Vehicle, Frontiers in Earth Science, 6, 64, doi:10.3389/feart.2018.00064, 2018.

Kraaijenbrink, P. D. A., Shea, J. M., Pellicciotti, F., Jong, S. M. D., and Immerzeel, W. W.: Object-based analysis of unmanned aerial vehicle imagery to map and characterise surface features on a debris-covered glacier, Remote Sensing of Environment, 186, 581–595, doi:10.1016/j.rse.2016.09.013, 2016.

Kuzmin, P. P.: The Process of Snow Cover Melting, Hydrometeoizdat, Leningrad, USSR, pp. 235–294, 1961.

Lejeune, Y., Bertrand, J.-M., Wagnon, P., and Morin, S.: A physically based model of the year-round surface energy and mass balance of debris-covered glaciers, Journal of Glaciology, 59(214), 327–344, doi:10.3189/2013JoG12J149, 2013.

Lukas, S., Nicholson, L. I., Ross, F. H., and Humlum, O.: Formation, Meltout Processes and Landscape Alteration of High-Arctic Ice-Cored Moraines—Examples From Nordenskiold Land, Central Spitsbergen, Polar Geography, 29(3), 157–187, doi:10.1080/789610198, 2005.

MacDougall, A. H. and Flowers, G. E.: Spatial and Temporal Transferability of a Distributed Energy-Balance Glacier Melt Model, Journal of Climate, 24(5), 1480–1498, doi:10.1175/2010JCLI3821.1, 2011.

Machguth, H., Purves, R. S., Oerlemans, J., Hoelzle, M., and Paul, F.: Exploring uncertainty in glacier mass balance modelling with Monte Carlo simulation, The Cryosphere, 2(2), 191–204, doi:10.5194/tc-2-191-2008, 2008.

Margirier, A., Brondex, J., Rowan, A. V., Schmidt, C., Pedersen, V. K., Lehmann, B., Anderson, L. S., Veness, R., Watson, C. S., Swift, D., and King, G. E.: Tracking Sediment Transport Through Miage Glacier, Italy, Using a Lagrangian Approach With Luminescence Rock Surface Burial Dating of Englacial Clasts, Journal of Geophysical Research: Earth Surface, 130(3), e2024JF007773, doi:10.1029/2024JF007773, 2025.

Mattson, L. E.: Energy exchanges and ablation rates on the debris-covered Rakhiot Glacier, Pakistan, Zeitschrift Fur Gletscherkunde Und Glazialgeolgie, 25, 17–32, 1991.

Mattson, L. E.: Ablation on debris covered glaciers: An example from the Rakhiot Glacier, Punjab, Himalaya, Snow and Glacier Hydrology, 1993.

Mattson, L. E.: The influence of a debris cover on the mid-summer discharge of Dome Glacier, Canadian Rocky Mountains, 264, 25–33, 2000.

McCarthy, M., Pritchard, H., Willis, I., and King, E.: Ground-penetrating radar measurements of debris thickness on Lirung Glacier, Nepal, Journal of Glaciology, 63(239), 543–555, doi:10.1017/jog.2017.18, 2017.

McCarthy, M.: A simplified energy-balance model for ice melt below debris (Version 1.0.0), GitHub, 2025.

McSaveney, M. J.: The Sherman Glacier rock avalanche of 1964: Its emplacement and subsequent effects on the glacier beneath it, The Ohio State University, 1975.

Melo-Velasco, V., Miles, E., McCarthy, M., Shaw, T. E., Fyffe, C., Fontrodona-Bach, A., and Pellicciotti, F.: Method dependence in thermal conductivity and aerodynamic roughness length estimates on a debris-covered glacier, Journal of Geophysical Research: Earth Surface, 130, e2025JF008360, doi:10.1029/2025JF008360, 2025.

Messmer, J. and Groos, A. R.: A low-cost and open-source approach for supraglacial debris thickness mapping using UAV-based infrared thermography, The Cryosphere, 18(2), 719–746, doi:10.5194/tc-18-719-2024, 2024.

Mihalcea, C., Mayer, C., Diolaiuti, G., Lambrecht, A., Smiraglia, C., and Tartari, G.: Ice ablation and meteorological conditions on the debris-covered area of Baltoro glacier, Karakoram, Pakistan, Annals of Glaciology, 43, 292–300, doi:10.3189/172756406781812104, 2006.

Miles, E. S., Pellicciotti, F., Willis, I. C., Steiner, J. F., Buri, P., and Arnold, N. S.: Refined energy-balance modelling of a supraglacial pond, Langtang Khola, Nepal, Annals of Glaciology, 57(71), 29–40, doi:10.3189/2016AoG71A421, 2016.

Miles, E. S., Steiner, J. F., and Brun, F.: Highly variable aerodynamic roughness length (z0) for a hummocky debris-covered glacier, Journal of Geophysical Research: Atmospheres, 122(16), 8447–8466, doi:10.1002/2017JD026510, 2017.

Miles, E. S., Willis, I., Buri, P., Steiner, J. F., Arnold, N. S., and Pellicciotti, F.: Surface Pond Energy Absorption Across Four Himalayan Glaciers Accounts for 1/8 of Total Catchment Ice Loss, Geophysical Research Letters, 45(19), doi:10.1029/2018GL079678, 2018.

Miles, K. E., Hubbard, B., Irvine-Fynn, T. D. L., Miles, E. S., Quincey, D. J., and Rowan, A. V.: Hydrology of debris-covered glaciers in High Mountain Asia, Earth-Science Reviews, 207, 103212, doi:10.1016/j.earscirev.2020.103212, 2020.

Mölg, N., Bolch, T., Walter, A., and Vieli, A.: Unravelling the evolution of Zmuttgletscher and its debris cover since the end of the Little Ice Age, The Cryosphere, 13(7), 1889–1909, doi:10.5194/tc-13-1889-2019, 2019.

Nakawo, M. and Young, G. J.: Estimate of Glacier Ablation under a Debris Layer from Surface Temperature and Meteorological Variables, Journal of Glaciology, 28(98), 29–34, doi:10.3189/S002214300001176X, 1982.

Nicholson, L. and Benn, D. I.: Calculating ice melt beneath a debris layer using meteorological data, Journal of Glaciology, 52(178), 463–470, doi:10.3189/172756506781828584, 2006.

Nicholson, L. and Benn, D. I.: Properties of natural supraglacial debris in relation to modelling sub-debris ice ablation, Earth Surface Processes and Landforms, 38(5), 490–501, doi:10.1002/esp.3299, 2013.

Nicholson, L. I., McCarthy, M., Pritchard, H. D., and Willis, I.: Supraglacial debris thickness variability: Impact on ablation and relation to terrain properties, The Cryosphere, 12(12), 3719–3734, doi:10.5194/tc-12-3719-2018, 2018.

Nicholson, L. and Stiperski, I.: Comparison of turbulent structures and energy fluxes over exposed and debris-covered glacier ice, Journal of Glaciology, 66(258), 543–555, doi:10.1017/jog.2020.23, 2020.

Østrem, G.: Ice Melting under a Thin Layer of Moraine, and the Existence of Ice Cores in Moraine Ridges, Geografiska Annaler, 41(4), 228–230, doi:10.1080/20014422.1959.11907953, 1959.

Østrem, G.: Problems of Dating Ice-Cored Moraines, Geografiska Annaler: Series A, Physical Geography, 47(1), 1–38, doi:10.1080/04353676.1965.11879710, 1965.

Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M., and Corripio, J.: An enhanced temperature-index glacier melt model including the shortwave radiation balance: Development and testing for Haut Glacier d'Arolla, Switzerland, Journal of Glaciology, 51(175), 573–587, doi:10.3189/172756505781829124, 2005.

Pellicciotti, F., Helbing, J., Rivera, A., Favier, V., Corripio, J., Araos, J., Sicart, J., and Carenzo, M.: A study of the energy balance and melt regime on Juncal Norte Glacier, semi-arid Andes of central Chile, using melt models of different complexity, Hydrological Processes, 22(19), 3980–3997, doi:10.1002/hyp.7085, 2008.

Popovnin, V. V. and Rozova, A. V.: Influence of Sub-Debris Thawing on Ablation and Runoff of the Djankuat Glacier in the Caucasus, Hydrology Research, 33(1), 75–94, doi:10.2166/nh.2002.0005, 2002.

Potter, E. R., Orr, A., Willis, I. C., Bannister, D., and Wagnon, P.: Meteorological impacts of a novel debris-covered glacier category in a regional climate model across a Himalayan catchment, Atmospheric Science Letters, 22(3), e1018, doi:10.1002/asl.1018, 2021.

Quincey, D., Smith, M., Rounce, D., Ross, A., King, O., and Watson, C.: Evaluating morphological estimates of the aerodynamic roughness of debris covered glacier ice, Earth Surface Processes and Landforms, 42(15), 2541–2553, doi:10.1002/esp.4198, 2017.

Ragettli, S., Pellicciotti, F., Immerzeel, W. W., Miles, E. S., Petersen, L., Heynen, M., Shea, J. M., Stumm, D., Joshi, S., and Shrestha, A.: Unraveling the hydrology of a Himalayan catchment through integration of high resolution in situ data and remote sensing with an advanced simulation model, Advances in Water Resources, 78, 94–111, doi:10.1016/j.advwatres.2015.01.013, 2015.

Reid, T. D. and Brock, B. W.: An energy-balance model for debris-covered glaciers including heat conduction through the debris layer, Journal of Glaciology, 56(199), 903–916, doi:10.3189/002214310794457218, 2010.

Reid, T. D. and Brock, B. W.: Assessing ice-cliff backwasting and its contribution to total ablation of debris-covered Miage glacier, Mont Blanc massif, Italy, Journal of Glaciology, 60(219), 3–13, doi:10.3189/2014JoG13J045, 2014.

Reid, T. D., Carenzo, M., Pellicciotti, F., and Brock, B. W.: Including debris cover effects in a distributed model of glacier ablation, Journal of Geophysical Research: Atmospheres, 117(D18), 2012JD017795, doi:10.1029/2012JD017795, 2012.

Rets, E. and Kireeva, M.: Hazardous hydrological processes in mountainous areas under the impact of recent climate change: Case study of Terek River basin, 340, 126–134, 2010.

Rets, E. P., Popovnin, V. V., Toropov, P. A., Smirnov, A. M., Tokarev, I. V., Chizhova, J. N., Budantseva, N. A., Vasil'chuk, Y. K., Kireeva, M. B., Ekaykin, A. A., Veres, A. N., Aleynikov, A. A., Frolova, N. L., Tsyplenkov, A. S., Poliukhov, A. A., Chalov, S. R., Aleshina, M. A., and Kornilova, E. D.: Djankuat glacier station in the North Caucasus, Russia: A database of glaciological, hydrological, and meteorological observations and stable isotope sampling results during 2007–2017, Earth System Science Data, 11(3), 1463–1481, doi:10.5194/essd-11-1463-2019, 2019.

Reznichenko, N., Davies, T., Shulmeister, J., and McSaveney, M.: Effects of debris on ice-surface melting rates: An experimental study, Journal of Glaciology, 56(197), 384–394, doi:10.3189/002214310792447725, 2010.

Robertson, E. C.: Thermal properties of rocks, 1988.

Röhl, K.: Characteristics and evolution of supraglacial ponds on debris-covered Tasman Glacier, New Zealand, Journal of Glaciology, 54(188), 867–880, doi:10.3189/002214308787779861, 2008.

Rounce, D. R., King, O., McCarthy, M., Shean, D. E., and Salerno, F.: Quantifying Debris Thickness of Debris-Covered Glaciers in the Everest Region of Nepal Through Inversion of a Subdebris Melt Model, Journal of Geophysical Research: Earth Surface, 123(5), 1094–1115, doi:10.1029/2017JF004395, 2018.

Rounce, D. R. and McKinney, D. C.: Debris thickness of glaciers in the Everest area (Nepal Himalaya) derived from satellite imagery using a nonlinear energy balance model, The Cryosphere, 8(4), 1317–1329, doi:10.5194/tc-8-1317-2014, 2014.

Rounce, D. R., Quincey, D. J., and McKinney, D. C.: Debris-covered glacier energy balance model for Imja–Lhotse Shar Glacier in the Everest region of Nepal, The Cryosphere, 9(6), 2295–2310, doi:10.5194/tc-9-2295-2015, 2015.

Rowan, A. V., Egholm, D. L., Quincey, D. J., and Glasser, N. F.: Modelling the feedbacks between mass balance, ice flow and debris transport to predict the response to climate change of debris-covered glaciers in the Himalaya, Earth and Planetary Science Letters, 430, 427–438, doi:10.1016/j.epsl.2015.09.004, 2015.

Rye, C. J., Arnold, N. S., Willis, I. C., and Kohler, J.: Modeling the surface mass balance of a high Arctic glacier using the ERA-40 reanalysis, Journal of Geophysical Research: Earth Surface, 115(F2), 2009JF001364, doi:10.1029/2009JF001364, 2010.

Sakai, A., Fujita, K., and Kubota, J.: Evaporation and percolation effect on melting at debris-covered Lirung Glacier, Nepal Himalayas, 1996, Bulletin of Glaciological Research, 21, 9–16, 2004.

Sakai, A., Nakawo, M., and Fujita, K.: Distribution Characteristics and Energy Balance of Ice Cliffs on Debris-covered Glaciers, Nepal Himalaya, Arctic, Antarctic, and Alpine Research, 34(1), 12–19, doi:10.1080/15230430.2002.12003463, 2002.

Sakai, A., Takeuchi, N., Fujita, K., and Nakawo, M.: Role of supraglacial ponds in the ablation process of a debris-covered glacier in the Nepal Himalayas, 264, 119–130, 2000.

Salerno, F., Thakuri, S., Tartari, G., Nuimura, T., Sunako, S., Sakai, A., and Fujita, K.: Debris-covered glacier anomaly? Morphological factors controlling changes in the mass balance, surface area, terminus position, and snow line altitude of Himalayan glaciers, Earth and Planetary Science Letters, 471, 19–31, doi:10.1016/j.epsl.2017.04.039, 2017.

Schauwecker, S., Rohrer, M., Huggel, C., Kulkarni, A., Ramanathan, Al., Salzmann, N., Stoffel, M., and Brock, B.: Remotely sensed debris thickness mapping of Bara Shigri Glacier, Indian Himalaya, Journal of Glaciology, 61(228), 675–688, doi:10.3189/2015JoG14J102, 2015.

Scherler, D. and Egholm, D. L.: Production and Transport of Supraglacial Debris: Insights From Cosmogenic 10 Be and Numerical Modeling, Chhota Shigri Glacier, Indian Himalaya, Journal of Geophysical Research: Earth Surface, 125(10), e2020JF005586, doi:10.1029/2020JF005586, 2020.

Scherler, D., Wulf, H., and Gorelick, N.: Global Assessment of Supraglacial Debris-Cover Extents, Geophysical Research Letters, 45(21), doi:10.1029/2018GL080158, 2018.

Shaw, T. E., Brock, B. W., Fyffe, C. L., Pellicciotti, F., Rutter, N., and Diotri, F.: Air temperature distribution and energy-balance modelling of a debris-covered glacier, Journal of Glaciology, 62(231), 185–198, doi:10.1017/jog.2016.31, 2016.

Shea, J. M., Immerzeel, W. W., Wagnon, P., Vincent, C., and Bajracharya, S.: Modelling glacier change in the Everest region, Nepal Himalaya, The Cryosphere, 9(3), 1105–1128, doi:10.5194/tc-9-1105-2015, 2015.

Steiner, J. F., Buri, P., Miles, E. S., Ragettli, S., and Pellicciotti, F.: Supraglacial ice cliffs and ponds on debris-covered glaciers: Spatio-temporal distribution and characteristics, Journal of Glaciology, 65(252), 617–632, doi:10.1017/jog.2019.40, 2019.

Steiner, J. F., Kraaijenbrink, P. D. A., and Immerzeel, W. W.: Distributed Melt on a Debris-Covered Glacier: Field Observations and Melt Modeling on the Lirung Glacier in the Himalaya, Frontiers in Earth Science, 9, 678375, doi:10.3389/feart.2021.678375, 2021.

Steiner, J. F., Litt, M., Stigter, E. E., Shea, J., Bierkens, M. F. P., and Immerzeel, W. W.: The Importance of Turbulent Fluxes in the Surface Energy Balance of a Debris-Covered Glacier in the Himalayas, Frontiers in Earth Science, 6, 144, doi:10.3389/feart.2018.00144, 2018.

Steiner, J. F. and Pellicciotti, F.: Variability of air temperature over a debris-covered glacier in the Nepalese Himalaya, Annals of Glaciology, 57(71), 295–307, doi:10.3189/2016AoG71A066, 2016.

Stokes, C. R., Popovnin, V., Aleynikov, A., Gurney, S. D., and Shahgedanova, M.: Recent glacier retreat in the Caucasus Mountains, Russia, and associated increase in supraglacial debris cover and supra-/proglacial lake development, Annals of Glaciology, 46, 195–203, doi:10.3189/172756407782871468, 2007.

Thakuri, S., Salerno, F., Smiraglia, C., Bolch, T., D'Agata, C., Viviano, G., and Tartari, G.: Tracing glacier changes since the 1960s on the south slope of Mt. Everest (central Southern Himalaya) using optical satellite imagery, The Cryosphere, 8(4), 1297–1315, doi:10.5194/tc-8-1297-2014, 2014.

Thompson, S., Benn, D. I., Mertes, J., and Luckman, A.: Stagnation and mass loss on a Himalayan debris-covered glacier: Processes, patterns and rates, Journal of Glaciology, 62(233), 467–485, doi:10.1017/jog.2016.37, 2016.

Tielidze, L. G., Bolch, T., Wheate, R. D., Kutuzov, S. S., Lavrentiev, I. I., and Zemp, M.: Supra-glacial debris cover changes in the Greater Caucasus from 1986 to 2014, The Cryosphere, 14(2), 585–598, doi:10.5194/tc-14-585-2020, 2020.

Tielidze, L. G., Bolch, T., Wheate, R. D., Kutuzov, S. S., Lavrentiev, I. I., and Zemp, M.: Supra-glacial debris cover changes in the Greater Caucasus from 1986 to 2014, The Cryosphere, 14(2), 585–598, doi:10.5194/tc-14-585-2020, 2020.

Van Woerkom, T., Steiner, J. F., Kraaijenbrink, P. D. A., Miles, E. S., and Immerzeel, W. W.: Sediment supply from lateral moraines to a debris-covered glacier in the Himalaya, Earth Surface Dynamics, 7(2), 411–427, doi:10.5194/esurf-7-411-2019, 2019.

Vincent, C., Wagnon, P., Shea, J. M., Immerzeel, W. W., Kraaijenbrink, P., Shrestha, D., Soruco, A., Arnaud, Y., Brun, F., Berthier, E., and Sherpa, S. F.: Reduced melt on debris-covered glaciers: Investigations from Changri NupGlacier, Nepal, The Cryosphere, 10(4), 1845–1858, doi:10.5194/tc-10-1845-2016, 2016.

Wang, L., Li, Z., and Wang, F.: Spatial distribution of the debris layer on glaciers of the Tuomuer Peak, western Tian Shan, Journal of Earth Science, 22(4), 528–538, doi:10.1007/s12583-011-0205-6, 2011.

Watson, C. S., Quincey, D. J., Carrivick, J. L., Smith, M. W., Rowan, A. V., and Richardson, R.: Heterogeneous water storage and thermal regime of supraglacial ponds on debris-covered glaciers, Earth Surface Processes and Landforms, 43(1), 229–241, doi:10.1002/esp.4236, 2018.

Wei, Y., Tandong, Y., Baiqing, X., and Hang, Z.: Influence of supraglacial debris on summer ablation and mass balance in the 24k glacier, southeast tibetan plateau, Geografiska Annaler: Series A, Physical Geography, 92(3), 353–360, doi:10.1111/j.1468-0459.2010.00400.x, 2010.

Westoby, M. J., Rounce, D. R., Shaw, T. E., Fyffe, C. L., Moore, P. L., Stewart, R. L., and Brock, B. W.: Geomorphological evolution of a debris-covered glacier surface, Earth Surface Processes and Landforms, 45(14), 3431–3448, doi:10.1002/esp.4973, 2020.

Wicky, J. and Hauck, C.: Air Convection in the Active Layer of Rock Glaciers, Front. Earth Sci., 8, 335, doi:10.3389/feart.2020.00335, 2020.

Winter-Billington, A., Moore, R. D., and Dadic, R.: Evaluating the transferability of empirical models of debris-covered glacier melt, J. Glaciol., 66, 978–995, doi:10.1017/jog.2020.57, 2020.

Winter-Billington, A., Dadić, R., Moore, R. D., Flerchinger, G., Wagnon, P., and Banerjee, A.: Modelling Debris-Covered Glacier Ablation Using the Simultaneous Heat and Water Transport Model. Part 1: Model Development and Application to North Changri Nup, Front. Earth Sci., 10, 796877, doi:10.3389/feart.2022.796877, 2022.

Wirbel, A., Jarosch, A. H., and Nicholson, L.: Modelling debris transport within glaciers by advection in a full-Stokes ice flow model, The Cryosphere, 12, 189–204, doi:10.5194/tc-12-189-2018, 2018.

Xie, F., Liu, S., Wu, K., Zhu, Y., Gao, Y., Qi, M., Duan, S., Saifullah, M., and Tahir, A. A.: Upward Expansion of Supra-Glacial Debris Cover in the Hunza Valley, Karakoram, During 1990 ~ 2019, Front. Earth Sci., 8, 308, doi:10.3389/feart.2020.00308, 2020.

Yang, W., Yao, T., Zhu, M., and Wang, Y.: Comparison of the meteorology and surface energy fluxes of debris-free and debris-covered glaciers in the southeastern Tibetan Plateau, J. Glaciol., 63, 1090–1104, doi:10.1017/jog.2017.77, 2017.

Zhang, Y., Fujita, K., Liu, S., Liu, Q., and Nuimura, T.: Distribution of debris thickness and its effect on ice melt at Hailuogou glacier, southeastern Tibetan Plateau, using in situ surveys and ASTER imagery, J. Glaciol., 57, 1147–1157, doi:10.3189/002214311798843331, 2011.

Supplement of

DCG-MIP: The Debris-Covered Glacier melt Model Intercomparison exPeriment

Francesca Pellicciotti, Adrià Fontrodona-Bach et al.

Correspondence to: Francesca Pellicciotti (francesca.pellicciotti@ista.ac.at) and Adrià Fontrodona-Bach (adria.fontrodona-bach@ista.ac.at)

1. Study sites and data

Table S1. Debris properties at the sites. Source indicates if the property was directly measured in the field, derived from field measurements, optimised, or assumed (i.e. taken from the literature).

Site	Surface roughness, z_0 (m)			Thermal conductivity, k (W m^{-1}K^{-1})		
	Value	Uncertainty	Source	Value	Uncertainty	Source
Arolla	0.0160	0.0010	Assumed - Brock et al. (2010)	0.96	0.10	Assumed - Brock et al. (2010)
Changri Nup	0.0500	0.005–0.5	Optimised - Lejeune et al. (2013)	0.70	0.10	Optimised - Lejeune et al. (2013)
Djankuat	0.0160	0.0008	Assumed - Bozhinsky (1986)	2.80	0.10	Assumed - Bozhinsky (1986)
Lirung	0.0350	0.0150	Derived - Miles et al. (2017)	1.55	0.30	Derived - Unpublished
Miage	0.0160	0.0010	Derived - Brock et al. (2010)	1.04	0.07	Derived - Brock et al. (2010)
Piramide	0.0190	0.0019	Derived - McPhee et al. (2019)	0.92	0.09	Assumed - Robertson (1988)
Suldenferner	0.0036	0.0004	Optimised - McCarthy (2018)	0.70	0.07	Derived - McCarthy (2018)
Tapado	0.0200	0.0020	Assumed - McPhee et al. (2019b)	0.94	0.09	Assumed - Brock et al. (2010)
Tasman	0.0160	0.0010	Assumed - Brock et al. (2010)	1.80	0.20	Derived - Röhl (2008)
Site	Porosity, ϕ (-)			Emissivity, ε (-)		
	Value	Uncertainty	Source	Value	Uncertainty	Source
Arolla	0.33	10%	Assumed - Brock et al. (2010)	0.96	1%	Assumed - Carenzo et al. (2016)
Changri Nup	0.37	8%	Measured - Wagnon (2019)	0.95	1%	Optimised - Lejeune et al. (2013)
Djankuat	0.42	12%	Assumed - Bozhinsky (1986)	0.90	5%	Assumed - Bozhinsky (1986)
Lirung	0.29	10%	Measured - Steiner et al. (2021)	0.94	2%	Assumed - Brock et al. (2010)
Miage	0.33	10%	Measured - Brock et al. (2010)	0.94	1%	Derived - Brock et al. (2010)
Piramide	0.33	10%	Assumed - Brock et al. (2010)	0.93	1%	Assumed - Brock et al. (2010)
Suldenferner	0.33	10%	Assumed - Brock et al. (2010)	0.94	1%	Assumed - Brock et al. (2010)
Tapado	0.33	10%	Assumed - Brock et al. (2010)	0.94	1%	Assumed - Brock et al. (2010)
Tasman	0.33	10%	Assumed - Brock et al. (2010)	0.94	2%	Assumed - Brock et al. (2010)

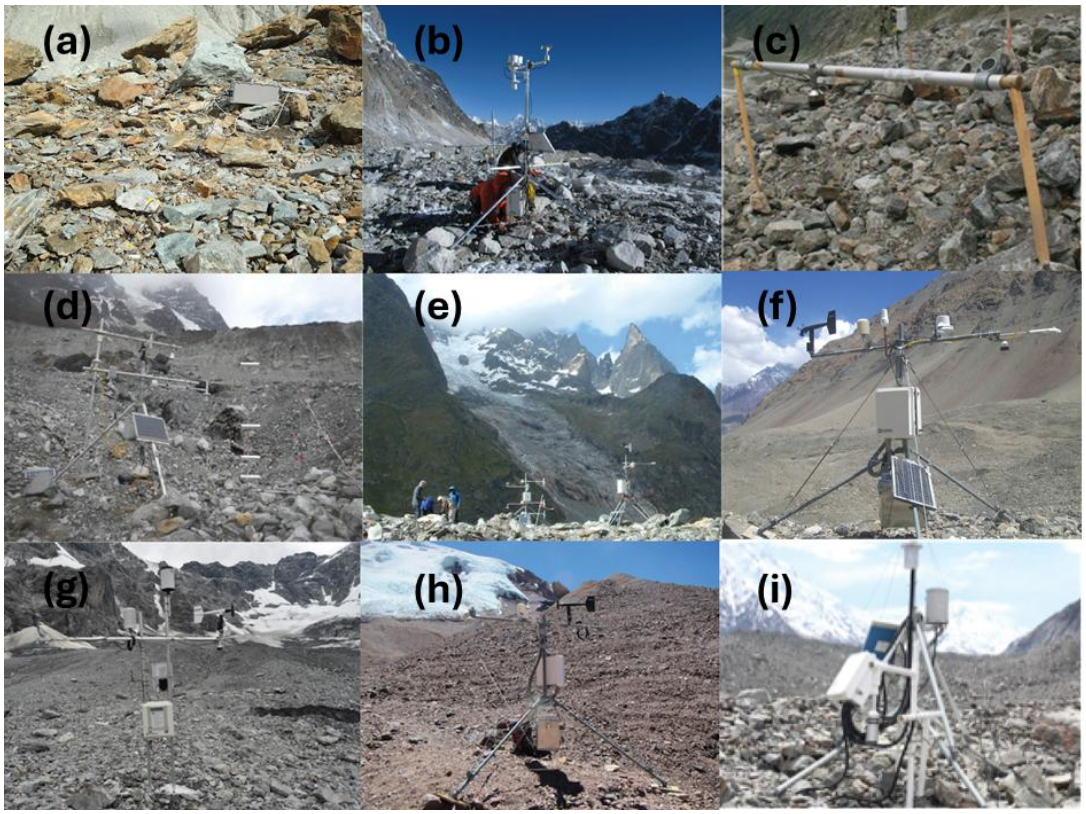


Figure S1. Automatic weather station location at the nine study sites. (a) Arolla; (b) Changri Nup; (c) Djankuat; (d) Lirung; (e) Miage; (f) Piramide; (g) Suldenferner; (h) Tapado; (i) Tasman.

Table S2. Estimates of validation data uncertainty given by data providers for measurements at the UDG, ablation stakes, and for the longwave radiation measurements used to derive surface temperature. Most estimates are the sensor's accuracy, not the actual measurement uncertainty.

Site	UDG / Stakes Uncertainty	Longwave Radiation Uncertainty
Arolla	± 5%	± 10%
Changri Nup	± 5 cm	± 3%
Djankuat	± 1 cm	± 15 W m^{-2}
Lirung	± 5%	± 5%
Miage	± 2 cm (stakes)	± 10%
Pirámide (stakes)	± 2 cm (stakes)	± 5%
Suldenferner	± 1 cm or 0.4%	± 10%
Tapado	± 5%	± 10%
Tasman	± 5%	± 10%

Table S3. Summary of meteorological forcing variables at each site. Mean temperature (T), relative humidity (RH), wind speed (FF), incoming shortwave and longwave radiation ($S\downarrow$, $L\downarrow$), outgoing shortwave radiation ($S\uparrow$), albedo (α, calculated as $S\uparrow$ divided by $S\downarrow$) and sum of precipitation (PP) over the period used for modelling. All means are daily.

Site	T (°C)	RH (%)	FF (m s^{-1})	$S\downarrow$ (W m^{-2})	$S\uparrow$ (W m^{-2})	α (-)	$L\downarrow$ (W m^{-2})	P (mm)
Arolla	5.7	72.8	3.3	190.7	29.8	0.16	294.5	215.2
Changri Nup	-1.4	75.5	1.9	213.5	51.0	0.24	279.7	257.7
Djankuat	7.3	66.7	2.3	267.6	36.6	0.14	289.7	337.3
Lirung	7.1	83.1	0.4	206.6	29.2	0.14	316.3	520.9
Miage	10.8	66.7	2.4	248.1	31.5	0.13	304.7	310.8
Pirámide	5.8	34.5	4.0	259.3	38.9	0.15	255.7	106.2
Suldenferner	6.7	75.1	2.4	199.2	36.7	0.18	301.0	234.6
Tapado	2.3	29.1	2.6	357.6	41.3	0.12	223.2	0.0
Tasman	12.9	57.0	2.8	240.8	21.6	0.09	328.5	1529.8

2. Model descriptions

This section of the supplement provides a small summary of all participating models in the debris-covered glacier melt model intercomparison experiment. An extended summary of all models, using a standard template that was filled in by each modeller, can be found at https://doi.org/10.5281/zenodo.15754455.

As in the main text, the models are sorted and described from more complex to less complex, according to our definition of complexity stated in Section 4 in the main text, and Fig. 3.

2.1. Energy balance models

An overview table of all energy balance models and how each model calculates each flux is presented in Table SI4 below the descriptions.

DEB_{CF}

Most models in this intercomparison experiment are at least partly based on the DEB model (debris energy-balance model) by Reid and Brock (2010), which we therefore describe as first and in more detail than the others. According to our definition of model complexity, the configuration of DEB_{CF} is the most complex of all participating models.

The model was developed first as a point-scale model by Reid and Brock (2010) and then was restructured into a distributed form in Fyffe *et al.* (2014). It calculates the energy balance at the surface of the debris, using the iterative Newton-Raphson method to derive the debris surface temperature. The model used for this experiment is the version used in Fyffe *et al.* (2014).

It is very similar to that in Reid and Brock (2010) but with three changes: 1) a snow component was added to simulate snow melt when snow covers debris; 2) thresholds were added to the Richardson number used to calculate the turbulent heat flux (see below); and 3) the debris is split into 10 layers for the calculation of heat conduction through the debris (so each layer is the debris thickness divided by 10). Changes to parameters resulted from the guidelines of this intercomparison experiment.

At the surface, the model calculates net shortwave radiation (*S*), net longwave radiation (*L*), sensible heat flux (*H*), latent heat flux (*LE*), heat flux due to precipitation (*P*) and the conductive heat flux below the surface (*G*).

$$M = S + L + H + LE + P + G \tag{S1}$$

where *M* is the energy available for ablation of debris-covered ice.

When there were occasional snow falls a simple energy balance snowmelt model was applied, with the general form:

$$a = \frac{\Delta t}{\rho_w L_f}[S + L + H + LE + P] \tag{S2}$$

where here *a* is snow ablation, Δt is the model time step, ρ_w is the water density and L_f is the latent heat of fusion of water. Conduction of heat or penetration of shortwave radiation into the snow was not modelled. The snow surface was assumed to remain at 0°C and the surface relative humidity was assumed to be 100%. Emissivity and surface roughness lengths are given literature values for snow (see Table S7). The calculation of the fluxes was the same as for the debris melt model.

For this experiment, *S* was calculated from the given incoming and outgoing shortwave radiation, rather than prescribing a snow/debris albedo, and if outgoing shortwave radiation was greater than incoming shortwave radiation it was made equal to the incoming shortwave radiation. L was calculated following standard methods outlined in Reid and Brock (2010), i.e. following the Stefan-Boltzmann law.

The turbulent sensible and latent heat fluxes were calculated using the bulk aerodynamic method, using the Richardson number to calculate the stability of the surface layer. For the MMI this was as outlined in Reid and Brock (2010), but with the thresholds on the Richardson number added in Fyffe et al. (2014) so that if $R_b > 0.2$ (turbulence ceased, airflow laminar) or $R_b < -1$ (free convection, conditions very unstable) then $R_b = 0$. For this experiment, if the instrument heights of wind speed and air temperature were different, then the height of the air temperature instrument was used for the sensible heat flux, similarly if the measurement heights of wind speed and relative humidity were different then the height of the relative humidity instrument was used to calculate the latent heat flux. Since no surface relative humidity data were provided it was assumed that the surface relative humidity equaled 100% whenever precipitation occurred. The latent heat flux was therefore only calculated at the timesteps with precipitation. The heat flux due to precipitation was calculated as given in Reid and Brock (2010).

The conductive heat flux, *G*, at the surface of the debris (note not the same as G_i) is approximated using T_s and the temperature of the debris at the first calculation layer, $T_d(1)$:

$$G = K\left(\frac{dT_d}{dz}\right)_{surface} \approx K\frac{T_d(1) - T_s}{h_l} \tag{S3}$$

where h_l is the calculation layer thickness (m), *K* is the debris thermal conductivity (W m^{-1} K^{-1}) and *z* is depth in the debris. Each layer within the debris is the debris thickness divided by 10 (this is the method used in Fyffe *et al.*, 2014). In Reid and Brock (2010) the debris calculation layer thickness was 1 cm unless the cover was thinner than 5 cm, in which case the number of layers was fixed at 5. Either method can easily be implemented in the DEB model. The conductive heat flux which reaches the glacier ice, G_i, depends on the temperature gradient at the base of the debris:

$$G_i(t) = K\left(\frac{dT_d}{dz}\right)_{base} \approx K\frac{T_d(N-1) - T_f}{h_l} \tag{S4}$$

where T_f is the temperature at the ice debris interface, which equals 0 °C. Then the melt beneath the debris (*a*) is calculated from:

$$a = \frac{G_i \Delta t}{\rho_w L_f} \tag{S5}$$

The debris surface temperature, T_s is required for G, and the outgoing longwave radiation, sensible heat flux, latent heat flux and the heat flux due to precipitation, but it is not known. To solve this problem the model solves the energy balance numerically; so that it varies the debris surface temperature until the sum of the heat fluxes at the debris surface is zero, using the iterative Newton-Raphson method:

$$T_s(n+1) = T_s(n) - \frac{F(T_s(n))}{F'(T_s(n))} \quad \text{(S6)}$$

where $F'(T_s)$ is the derivative of the total surface flux with respect to T_s (the debris surface temperature, °C), and is calculated numerically by the central difference method. For each model time step $T_s(n = 0)$ must have an initial value, so for the first time step it is equal to the air temperature, and for further timesteps it is the value from the previous time step. This equation is repeated until $|T_s(n + 1) - T_s(n)| < 0.01$. To calculate the conduction of heat through the debris to the ice/debris interface the following equation is used:

$$\rho_d C_{pd} \frac{\partial T_d\,(z_d,t)}{\partial t} = \frac{\partial}{\partial z_d}\left(K \frac{\partial T_d\,(z_d,t)}{\partial z_d}\right) \quad \text{(S7)}$$

which uses the partial derivatives of debris temperature, T_d, with respect to time, t and depth in the debris, z_d (m), where ρ_d is debris density (kg m^{-3}), C_{pd} is the debris specific heat (J kg^{-1} K^{-1}) and K is the debris thermal conductivity (W m^{-1} K^{-1}).

Table S4. Values of physical and empirical parameters used by the DEB model. *From Robinson and Coruh (1988), cited in Brock et al. (2010, p10). **From Lide (2004), cited in Brock et al. (2010, p10). ***From Brock et al. (2006, p288). ****Same as ice emissivity in Reid and Brock (2010). Note that the ice density was changed from 915 kg m^{-3} to 900 kg m^{-3} following the DCG-MIP guidelines.

Parameter	Debris Density* (kg m^{-3})	Density of Ice (kg m^{-3})	Debris Specific Heat Capacity** (J kg^{-1} K^{-1})	Snow Aerodynamic Roughness*** (m)	Snow Emissivity****
All sites	1496	900	948	0.00231	0.97

ROU15

This model is based on Rounce et al. (2015) and calculates the same fluxes at the debris surface as the DEB_{CF} model: S, L, H, LE, P and G. Despite the similarities with DEB_{CF}, It is considered slightly simpler than DEB_{CF} as it assumes neutral conditions in the calculation of the turbulent heat fluxes.

The model has a snowpack module to calculate snow accumulation and melt based on meteorological forcing. However, snow depth was provided as part of the input data, and therefore when there is snow on the surface, the model assumes the net radiation flux, latent heat flux, sensible heat flux, and precipitation heat flux are zero, and only accounts for the ground heat flux and the conductive heat flux at the snow/debris surface interface.

The calculation of the turbulent heat fluxes differs from DEB_{CF}. The sensible heat flux is estimated according to Nicholson and Benn (2006), assuming neutral conditions without stability corrections. The latent heat flux is also estimated according to Nicholson and Benn (2006), with the debris surface assumed to be saturated when it is raining. The precipitation heat flux and the heat conduction through the debris are estimated in the same way as the model DEB_{CF}, both following Reid and Brock (2010), and with the debris layer discretized in 10 layers of equal thickness.

Table S5. Values of physical and empirical parameters used by the ROU15 model.

Parameter	Debris Density* (kg m^{-3})	Density of Ice (kg m^{-3})	Debris Specific Heat Capacity** (J kg^{-1} K^{-1})
All sites	2700	900	750

$GRO17_B$

The theoretical model of glacial melt under a porous debris layer developed by Evatt et al. (2015) provides the basis for the energy balance model GRO17 (Groos et al., 2017; Groos & Mayer, 2017). A main difference compared to earlier models (e.g. Nicholson and Benn, 2006; Reid and Brook, 2010) is the incorporation of the turbulent latent heat flux within the porous debris layer. Due to the treatment of the debris layer as a porous medium and the consideration of evaporation at the debris-ice

interface, the model is able to capture the shape of Østrem curves derived from in-situ measurements (Evatt et al., 2015). Two different model versions, GRO17A and GRO17B, were applied in the model intercomparison experiment: GRO17B assumes a permeable debris layer and accounts for the turbulent latent heat flux within the debris, while GRO17A assumes an impermeable debris layer and accounts only for heat conduction to transfer energy through the debris layer.

For GRO17B, the energy balance at the debris-atmosphere interface is given by

$$Q_D = Q_S + Q_L + Q_{SH} \tag{S8}$$

and at the debris-ice interface by

$$Q_D = Q_M + Q_V \tag{S9}$$

where Q_D is the ground heat flux (W m^{-2}), Q_S is the shortwave energy flux (W m^{-2}), Q_L is the longwave energy flux (W m^{-2}), Q_{SH} is the sensible heat flux (W m^{-2}), Q_M is the latent heat of melting (W m^{-2}) and Q_V is the evaporative (or turbulent latent) heat flux (W m^{-2}) (for a detailed description and derivation of the individual fluxes see equations 1, 2, 5, 6, 29 and 31 in Evatt et al., 2015).

The model could theoretically manage the snow cover on top of the debris using a simple temperature index approach, but for simplicity and consistency, sub-debris ice melt was set to zero if the height of the snow cover was >0.0 m.

In contrast to the other participating energy balance models, the GRO17B model would require additional parameters like friction velocity (u_*), debris porosity (phi) and grain size to describe the geometry of the porous debris layer. These properties determine the wind speed attenuation parameter (γ) that controls the decay of the airflow through the debris layer (see equations 23 and 52 in Evatt et al., 2015). However, information on grain size was not available and porosity was not measured at most sites (see Table S1). Therefore, the wind speed attenuation constant (γ) was calibrated by minimising the difference (RMSE) between Q_{SH} modelled by Eq. 31 from Evatt et al. (2015) and Q_{SH} modelled by Eq. 7 from Nicholson and Benn (2006). Similarly, the friction velocity (u_*), required for the calculation of the sensible heat flux is parameterised as $u_* = u_m / \alpha$, where u_m is the measured wind speed and α is a scaling factor calibrated by minimising the difference (RMSE) between Q_{SH} modelled by Eq. 32 from Evatt et al. (2015) and Q_{SH} modelled by Eq. 7 from Nicholson and Benn (2006).

The downward heat flux through the debris is modelled by Fourier's law, assuming a linear temperature profile derived from the quasi steady-state heat equation (see Evatt et al., 2015). For the calculation of the internal temperature profile, the debris layer was discretised into n layers of 1 cm thickness.

Solving the heat conduction through the debris in a quasi steady state makes this model simpler than DEBCF and ROU15, despite the added complexity of this model with regards to solving the latent heat flux within the porous debris layer. GRO17 is the only model which takes into account a fraction of debris embedded in the ice when calculating ice melt under debris from the conductive heat flux reaching the ice.

Table S6. Summary of parameter values for GRO17$_B$ sites, including friction velocity (u^*), wind speed attenuation constant (γ), ice density (ρ_i), volume fraction of debris (θ_d), and measurement height. u_m is the mean wind speed at the site.

Site	u^* (m/s)	γ	ρ_i (kg/m^3)	θ_d	Height (m)
Arolla	$u_m/12.1$	67	900	0.01	2
Changri Nup	$u_m/9.2$	20	900	0.01	2
Djankuat	$u_m/12.0$	82	900	0.01	2
Lirung	$u_m/9.8$	47	900	0.01	2
Miage	$u_m/12.0$	67	900	0.01	2
Piramide	$u_m/11.6$	49	900	0.01	2
Suldenferner	$u_m/15.7$	308	900	0.01	2
Tapado	$u_m/11.5$	52	900	0.01	2
Tasman	$u_m/12.1$	52	900	0.01	2

d2EB

This model is based on Steiner et al. (2018, 2021) and calculates the following fluxes: *S*, *L*, *H*, *LE* and *G*.

In general, the model is simply Reid and Brock (2010) with the only adaptation that turbulent fluxes are calculated based on Steiner et al. (2018, 2021) and Nicholson and Benn (2006). This was done according to Steiner et al. as the parametrization with the Richardson number fails on debris due to the strong heating. The model does not deal with snow.

The sensible heat flux is calculated following Equation 1 and 4 in Steiner et al. (2018) and assuming neutral stability over the debris cover as in Nicholson and Benn (2006). The latent heat flux is calculated following Equations 2 and 4 in Steiner et al. (2018). The model requires specific humidity but is calculated from standard formulas for relative humidity and saturation vapour pressure as a function of temperature.

Heat conduction through the debris is calculated as in Reid and Brock (2010) and the number of debris layers is a function of the thickness (N = thickness [m] / 0.01).

Table S7. Values of physical and empirical parameters used by the d2EB model.

Parameter	Debris Density (kg m^{-3})	Density of Ice (kg m^{-3})	Specific Heat Capacity of Debris (J kg^{-1} K^{-1})
All sites	2650	900	890

DEB_{PG}

This model is the same as DEB_{CF} described above, except for the following differences:

- DEB_{PG} does not deal with snow, so it does not calculate snow accumulation and melt on top of the debris.
- DEB_{PG} does not calculate the precipitation heat flux and therefore does not need precipitation as input data.
- DEB_{PG} calculates the latent heat flux when the relative humidity of the surface is 100% and assumes that is the case when the relative humidity of the air is 100%.
- DEB_{PG} has a different debris layer discretisation than DEB_{CF}. For debris thickness less than 6 cm, d/3 layers assumed. For debris thickness more than 6 cm, each calculation layer was 2-2.05 cm thick.

Table S8. Values of physical and empirical parameters used by the DEBGA model. *From Robinson and Coruh (1988), cited in Brock et al. (2010, p10).

Parameter	Debris Density* (kg m^{-3})	Density of Ice (kg m^{-3})	Debris Specific Heat Capacity* (J kg^{-1} K^{-1})
All sites	1496	900	948

$GRO17_A$

This model is the same as GRO17B described above, except that GRO17A assumes a permeable debris layer and therefore has the following two differences in model configuration:

- GRO17A disregards the evaporation heat flux (Q_V).
- GRO17A expresses the sensible heat flux analogue to previous studies (e.g. Nicholson and Benn, 2006; as in equation 32 in Evatt et al., 2015).

Table S9. Summary of parameter values for GRO17A sites, including friction velocity (u^*), ice density (ρ_i), volume fraction of debris (Φ), and measurement height. u_m is the mean wind speed at the site.

Site	u^* (m/s)	ρ_i (kg/m^3)	Φ	Height (m)
Arolla	$u_m/12.1$	900	0.01	2
Changri Nup	$u_m/9.2$	900	0.01	2
Djankuat	$u_m/12.0$	900	0.01	2
Lirung	$u_m/9.8$	900	0.01	2
Miage	$u_m/12.0$	900	0.01	2
Piramide	$u_m/11.6$	900	0.01	2
Suldenferner	$u_m/15.7$	900	0.01	2
Tapado	$u_m/11.5$	900	0.01	2
Tasman	$u_m/12.1$	900	0.01	2

A-Melt

The A-Melt model is a distributed energy-balance model developed for alpine regions of snow and ice melting and meltwater runoff formation (Rets and Kireeva, 2010; Elagina et al., 2025). For this study a simplified 1D version of the model was developed.

The heat balance of the surface of snow or debris is defined at every time step as:

$$\omega = S_{\text{down}} - S_{\text{out}} + E_{\text{lrd}} - E_{\text{lu}} \pm LE \pm H \pm Q_m \pm Q_{\text{act}} \quad \text{(S10)}$$

where ω is the net energy flux on the surface, W m^{-2}; S_{down} is the downward shortwave radiation flux, W m^{-2}; Sout is an upward shortwave radiation flux, W m^{-2}; E_{lru} is outgoing long-wave radiation, W m^{-2}; E_{lrd} is the counter radiation of the atmosphere, W m^{-2}; LE is the turbulent latent - heat flux density, W m^{-2}; H is the turbulent sensible - heat flux density, W m^{-2}, Q_m is the molecular thermal conductivity through the debris cover, W m^{-2}, Q_{act} is the energy change due to processes of water infiltration into the firm and the transfer of thermal conductivity deep into the active layer of the glacier, W m^{-2}

The model calculates snow melt as long as snow is on the ground, then reverts to melt of ice under debris. It is supposed that there is snow on top of the debris, if the snow depth (H_{snow}) according to the input data is >=0, excluding cases when H_{snow} is thin (less than 2 cm), and the surface albedo is less than 0.5 according to measurements.

The A-Melt model is the only model in this intercomparison experiment that does not assume a constant ice temperature of 0ºC. The temperature in the surface ice layer can decrease below zero if the energy balance is negative, and warm back to 0ºC. The melting can occur only if the temperature of the surface ice layer has reached 0ºC.
The temperature of the ice surface layer is calculated as:

$$T_{\text{ice}}^{t} = T_{\text{ice}}^{t-1} + \frac{Q_m + Q_{\text{act}}}{c_{\text{ice}}\rho_{\text{ice}}\Delta h_{\text{ice}}}\Delta t, \text{ K} \quad \text{(S11)}$$

where T_{ice}^{t} is the temperature of the ice surface layer on this time step, T_{ice}^{t-1} is the temperature of the ice surface layer on the previous time step, Q_m is the molecular thermal conductivity through the debris cover, W m^{-2}, Q_{act} is the molecular thermal conductivity in the active layer, W m^{-2}, c_{ice} is the ice heat capacity, ρ_{ice} – ice density (set to 0.917 kg m^{-3}), Δh_{ice} – thickness of ice surface layer (set with the parameter ΔT_{prof}), Δt is the time step.
Q_{act} is calculated as:

$$Q_{act} = \lambda_{ice}\frac{T_{act} - T_{ice}}{h_{act}} \quad \text{(S12)}$$

where h_{act} is active layer thickness, m; T_{ice} is the temperature of the surface layer of ice, underlying debris; T_{act} is active layer bottom temperature, K, set as a constant value, λ_{ice} - the thermal conductivity of the ice, calculated according to ice density (Sturm et al., 1997), W m^{-1} K^{-1}.

The model uses a simplified bulk aerodynamic approach to calculate turbulent sensible and latent heat fluxes, using a formula developed by Kuzmin (1961) specifically for turbulent heat and moisture transfer over snow and ice surfaces. This approach uses constant bulk coefficients.

Maximum air humidity at the temperature of the ice surface, e_0, was calculated as the partial pressure of water vapor of the saturated air at the temperature :

$$Q_m = \lambda_m \frac{T_m - T_{ice}}{h_m} \quad \text{(S14)}$$

where T_m is debris surface temperature, T_{ice} is the temperature of the surface layer of ice, underlying debris. We assume the linear distribution of the air temperature in pores of debris, as the model has one layer of the debris.

The ice, covered with debris, is melting under the influence of the heat flux conducted through the debris by means of molecular thermal conductivity Q_m:

$$Q_m = \lambda_m \frac{T_m - T_{ice}}{h_m} \quad \text{(S14)}$$

where λ_m is the thermal conductivity of the debris cover, W m^{-1} K^{-1}; h_m is the debris cover thickness, m; T_{ice} is the temperature of the surface layer of ice, underlying debris; T_m is debris surface temperature, K, calculated according to its heat balance.

Table S10. Values of physical and empirical parameters used by the A-Melt model. Constant per site parameters: ice density = 917 kg m^{-3}, $\rho_{OldSnow} = 0.45$, $\Delta T_{prof} = 100$ cm, $\Delta T = 60$ min, $\alpha_{snow} = 0.5$. h_m = debris thickness (cm); dh_m = surface layer of debris (cm); $\rho_{OldSnow}$ = mean density of old snow (kg m^{-3}); ΔT_{prof} = layer thickness of temperature profile of snow (cm); T_{act} = glacier active layer bottom temperature (°C); h_{ActIce} = glacier active layer depth (cm); specific heat of debris (C_{deb}) – set according to lithology and literature (Robertson, E, 1988); ρ_{deb} = debris density based on lithology (source: https://pubs.usgs.gov/of/1988/0441/report.pdf); α_{snow} = snow albedo.

Parameter	ARO	CN	DJA	LIR	MIA	PIR	SDF	TAP	TAS
ε_{snow} (-)	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
h_m (cm)	6	10	61	30	22	18	11	80	30
dh_m (cm)	6	10	10	10	10	10	10	10	10
T_{act} (°C)	0	-8	0	0	0	-1	-1	-2	0
h_{ActIce} (cm)	10	2	10	10	10	10	10	10	10
C_{deb} (J kg^{-1} K^{-1})	900	800	800	500	800	800	800	800	900
ρ_{deb} (kg m^{-3})	2670	2700	2750	2700	2700	2200	2700	2700	2600

THRED

The thermal resistance based energy balance model for debris-covered glaciers (THRED) was developed for calculating runoff from Himalayan debris-covered glaciers, in which the spatial distribution of the thermal properties of the debris mantle is estimated from remotely sensed multi-temporal data (Fujita and Sakai, 2014).

This is the only energy balance model in this intercomparison experiment that is run at **daily** resolution. The model calculates the following fluxes at the debris surface:

$$G_d = \frac{T_s}{R_T} = (1-\alpha)R_s + R_L - \varepsilon\sigma(T_s + 273.15)^4 + H_s + H_L \quad \text{(S15)}$$

where G_d is the conductive heat through debris, T_S the surface temperature; R_T the thermal resistance; α the surface albedo, R_S the downward solar radiation, R_L the downward long-wave radiation, ε the emissivity of debris surface, σ the Stefan-Boltzmann constant; H_S the sensible heat, H_L the latent heat.

Thermal resistance, a proxy of debris thickness, is defined as the debris thickness divided by the thermal conductivity of the debris layer as:

$$R_T = \frac{h}{\lambda} \quad \text{(S16)}$$

where R_T, h, and λ are the thermal resistance (m^2 K W^{-1}), thickness of debris layer (m), and thermal conductivity of debris (m^{-1} K^{-1} W), respectively. In the original study, the thermal resistance was calculated from surface temperature and albedo, which

are obtained from satellite data, and reanalysis meteorological data by assuming a linear temperature profile within the debris layer as:

$$G_d = \frac{T_s - T_i}{R_T} \tag{S17}$$

where G_d, T_s, and T_i are the conductive heat flux through the debris layer (W m^{-2}), surface temperature (ºC), and temperature at the interface of debris and ice (assumed to be 0 ºC), respectively. In the experiment, thermal resistance is given as the boundary conditions (thickness and thermal conductivity). In the original study, satellite based albedo was used to calculate net shortwave radiation while the SWout is a given input variable in the experiment. Downward longwave radiation is also a given input variable though it was estimated from temperature, relative humidity and solar radiation in the original study.

The turbulent heat fluxes are calculated based on simplified bulk approaches following Fujita and Sakai (2014). A constant bulk coefficient for the turbulent flux on the debris surface (C_d, dimensionless) was used in the original study. For this experiment the model was modified to participate in the sensitivity experiment (see Section 2 in manuscript), and the bulk coefficient was calculated as:

$$C_d = \left[\frac{\kappa}{\ln\left(\frac{z_t}{z_0}\right)}\right]^2 \tag{S18}$$

where κ, z_t, and z_0 are von Karman's constant (0.4, dimensionless), measurement height (m), and surface roughness length (m), respectively.

Table S11. Values of physical and empirical parameters used by THRED model.

Parameter	Value
Density of ice (kg m^{-3})	880

2.2. Simplified energy balance model

MCC19

The simplified energy balance model (SEB) (MCC19, McCarthy, 2025), is a one-dimensional, simplified energy-balance model for ice melt below debris.

The model computes explicitly the shortwave radiation flux and the conductive heat flux at the atmosphere-debris interface. Other energy fluxes are included implicitly in the free parameter terms of the simplified debris-surface energy balance. When the debris is snow-covered, the model sets the surface temperature of the debris to 0ºC. Conductive heat fluxes are calculated at intervals within the debris. The model does not deal explicitly with moisture within the debris. At the debris-ice interface, only a conductive heat flux is calculated. The temperature of the ice was fixed at 0ºC.

The model solves heat conduction through a debris layer as such:

$$\rho_d c_d \frac{\partial T_d}{\partial t} = \frac{\partial}{\partial z}\left(k_d \frac{\partial T_d}{\partial z}\right) \tag{S19}$$

where ρ_d is debris density (assumed 1496 kg m^{-3}), c_d is debris specific heat capacity (assumed 948 J kg^{-1}), k_d is debris thermal conductivity, t is time, z is depth within the debris, and T_d is debris temperature.

The boundary condition at the debris surface is the simplified debris-surface energy balance:

$$S\downarrow(t) - S\uparrow(t) + c_1[T_a(t) - T_d(0,t)] + c_2 - k_d \frac{\partial T_d(0,t)}{\partial z} = 0 \tag{S20}$$

where $S\downarrow$ is incoming shortwave radiation, $S\uparrow$ is outgoing shortwave radiation, T_a is air temperature, and c_1 and c_2 are free parameters to be calibrated. This equation has a similar form to the simplified (debris-free) energy balance of Oerlemans (2001). Surface temperature is determined iteratively for each model timestep using Newton's method, following Reid and Brock (2010).

The boundary condition at the ice surface is the temperature of melting ice, T_i (273.15 K):

$$T_d(h,t) = T_i \quad \text{(S21)}$$

where h is debris thickness.

Melt rate, M, is computed as follows:

$$M(t) = \frac{\Delta t}{\rho_i L_f} k_d \frac{\partial T_d(h,t)}{\partial z} \quad \text{(S22)}$$

where Δt is the model timestep, ρ_i is ice density (assumed 915 kg m^{-3}), and L_f is the latent heat of fusion of water (334000 J kg^{-1}).

For this intercomparison experiment, the debris was discretised into 0.01 m thick layers (or ten thinner layers if debris thickness was < 0.1 m), following Reid and Brock (2010). All the layers were considered to have the same physical properties.

No spin-up period was used. A linear temperature gradient through the debris was used as the initial condition, again following Reid and Brock (2010), where debris surface temperature was the first recorded air temperature value of each time series and the ice surface temperature was the temperature of melting ice, as follows:

$$T_d(z,0) = [T_a(l) - T_i]z \quad \text{(S23)}$$

Additional input parameters include debris density (assumed 1496 kg m^{-3}) and debris specific heat capacity (assumed 948 J kg^{-1}). Two free parameters, c_1 and c_2 are determined by calibration for each site using the observed surface temperature and surface lowering data provided. This was done using a multiparameter multiobjective optimisation approach, following Rye et al (2010):

$$f_{agg}(c_1,c_2) = [f_1^{norm}(c_1,c_2) + f_2^{norm}(c_1,c_2)] \quad \text{(S24)}$$

where f_{agg} is the aggregate objective function that was minimised, and f_1^{norm} and f_2^{norm} are the individual normalised objective functions f_{agg} is composed of.

The two individual objective functions $f_1(c_1, c_2)$ and $f_2(c_1, c_2)$ used in the optimisation procedure, prior to normalisation, were as follows:

$$f_1(c_1,c_2) = \frac{1}{n}\sum_{d=1}^{n}\left|\frac{m_{obs}^d - m_{mod}^d}{d-1}\right| \quad \text{(S25)}$$

$$f_2(c_1,c_2) = \frac{\sum_{t=1}^{n}(T_d^t(0)_{mod} - T_d^t(0)_{obs})^2}{\sum_{t=1}^{n}\left(T_d^t(0)_{obs} - \overline{T_d(0)_{obs}}\right)^2} \quad \text{(S26)}$$

where n is the number of observations, m_{mod}^d and m_{obs}^d are modelled and observed cumulative melt at day d, and $T_d^t(0)_{mod}$ and $T_d^t(0)_{obs}$ are the modelled and observed surface temperatures for timestep t. The first of these objective functions is the mean absolute error of cumulative melt per day. The second is 1-the Nash-Sutcliffe Efficiency coefficient of the surface temperature record. Both were normalised prior to minimisation according to:

$$f_n^{norm}(c_1,c_2) = \frac{f_n(c_1,c_2) - f_n^{min}}{f_n^{max} - f_n^{min}} \quad \text{(S27)}$$

where f_n^{min} and f_n^{max} are minimum and maximum expected objective function values. These were set to 0 and 0.0056 m i.e. d^{-1}, respectively, for the melt function, and to 0 and 0.5, respectively for the surface temperature function.

Table S12. Optimal free parameter values determined for the nine sites for model MCC19.

Site	c_1 (W m^{-2} K^{-1})	c_2 (W m^{-2})
Arolla	64.8	88.1
Changri-Nup	56.2	214
Djankuat	34.7	8.63
Lirung	54.6	96.6
Miage	45.8	-33.7
Piramide	48.0	-64.8
Suldenferner	40.2	78.3
Tapado	32.6	-24.0
Tasman	61.0	-34.3

Overview tables for EB and SEB

Table S13. Overview of the net shortwave radiation (S_{net}) calculation methods by each model.

Model	S_{net} Calculation
DEB_{CF}	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
ROU15	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
GRO17_{B}	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
d2EB	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
DEB_{PG}	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
GRO17_{A}	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
A-Melt	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
THRED	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations
MCC19	Both $S_{\downarrow}$ and $S_{\uparrow}$ from observations

Table S14. Overview of the net longwave radiation (L_{net}) calculation methods by each model.

Model	L_{net} Calculation
DEB_{CF}	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
ROU15	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
$GRO17_{B}$	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
d2EB	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
DEB_{PG}	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
$GRO17_{A}$	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
A-Melt	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
THRED	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s
MCC19	$L_{\downarrow}$ from measurement, $L_{\uparrow}$ calculated from internally computed T_s

Table S15. Overview of the sensible heat flux (SH) calculation methods by each model.

Model	SH Calculation
DEB_{CF}	Bulk aerodynamic method using Richardson number (Ri), as in Reid and Brock (2010). Stability functions calculated as a function of Ri.
ROU15	Simplified bulk approach (Nicholson and Benn, 2006). No stability corrections, assumption of neutral conditions.
$GRO17_{B}$	Following Evatt et al. (2015) with a non-linear wind-speed profile within the porous debris layer.
d2EB	Simplified bulk approach following Steiner et al. (2018). No stability corrections, assumption of neutral conditions.
DEB_{PG}	Bulk aerodynamic method using Richardson number (Ri), as in Reid and Brock (2010). Stability functions calculated as a function of Ri.
$GRO17_{A}$	Simplified approach as in Nicholson and Benn (2006). No stability corrections.
A-Melt	Simplified bulk approach, with constant coefficients (Kuzmin, 1961): $H = (\alpha_1 + \beta_1 u^2)(T_2 - T_0)$.
THRED	Simplified bulk approach following Fujita and Sakai (2014). No stability corrections, assumption of neutral conditions.
MCC19	Simplified bulk approach following Steiner et al. (2018). No stability corrections, assumption of neutral conditions.

Table S16. Overview of the latent heat flux (LE) calculation methods by each model.

Model	LE Calculation
DEB_{CF}	Same as for SH, as in Reid and Brock (2010). RH_s was assumed to be 100% (saturated surface) during rain.
ROU15	Simplified bulk approach (Nicholson and Benn, 2006). No stability corrections, assumption of neutral conditions. Debris is assumed saturated when it rains; otherwise, the flux is zero.
$GRO17_{B}$	Evaporative heat flux within the debris.
d2EB	Simplified bulk approach following Steiner et al. (2018). Debris surface moisture is calculated based on a parameterisation developed in Steiner et al. (2018).
DEB_{PG}	Same as for SH, as in Reid and Brock (2010). Calculated assuming that $RH_s = 100\%$ when air RH is 100%.
$GRO17_{A}$	No latent heat flux included.
A-Melt	Simplified bulk approach, with constant coefficients (Kuzmin, 1961): $LE = (\alpha_2 + \beta_2 u^2)(e_2 - e_0)$.
THRED	Simplified bulk approach following Fujita and Sakai (2014), including dependence on debris surface wetness. No stability corrections, assumption of neutral conditions.
MCC19	Simplified bulk approach following Steiner et al. (2018). No stability corrections, assumption of neutral conditions.

Table S17. Overview of the heat flux provided by liquid precipitation (P) calculation methods by each model.

Model	P Calculation
DEB_{CF}	Yes, as in Reid and Brock (2010).
ROU15	Yes, as in Reid and Brock (2010).
$GRO17_B$	No
d2EB	No
DEB_{PG}	No
$GRO17_A$	No
A-Melt	No
THRED	No
MCC19	No

Table S18. Overview of the ground heat flux (G) calculation methods by each model.

Model	G Calculation
DEB_{CF}	Fourier's law and iterative Newton-Raphson method to vary debris surface temperature and solve the energy balance, as in Reid and Brock (2010).
ROU15	Fourier's law and iterative Newton-Raphson method to vary debris surface temperature and solve the energy balance, as in Reid and Brock (2010).
$GRO17_B$	Fourier's law: simplified approach assuming a linear temperature gradient between T_s and T_i.
d2EB	Fourier's law and iterative Newton-Raphson method to vary debris surface temperature and solve the energy balance, as in Reid and Brock (2010).
DEB_{PG}	Fourier's law and iterative Newton-Raphson method to vary debris surface temperature and solve the energy balance.
$GRO17_A$	Fourier's law: simplified approach assuming a linear temperature gradient between T_s and T_i.
A-Melt	Simplified approach, one layer, linear temperature gradient between debris surface and ice surface.
THRED	Simplified approach, one layer, linear temperature gradient between debris surface and ice surface.
MCC19	No

Table S19. Overview of snow conditions and snowmelt calculation methods by each model.

Model	Snow Conditions and Snowmelt Calculation
DEB_{CF}	Yes. The model checks the snow depth record and calculates snowmelt as long as snow is on the ground, then reverts to the melt of ice under debris.
ROU15	When there is snow on the surface, the model assumes fluxes are zero except for the ground heat flux and the conductive heat flux at the snow/debris surface interface.
$GRO17_B$	No. Sub-debris ice melt was set to zero if the height of the snow cover was >0.0 m.
d2EB	No.
DEB_{PG}	No.
$GRO17_A$	No. Sub-debris ice melt was set to zero if the height of the snow cover was >0.0 m.
A-Melt	Yes. The model calculates snowmelt as long as snow is on the ground, then reverts to the melt of ice under debris. Snow on top of the debris is accounted for if snow depth (H_{snow}) is ≥ 0, except for thin snow layers (<2 cm) with surface albedo < 0.5.
THRED	Yes. Described in Fujita and Sakai (2014).
MCC19	No.

Table S20. Overview of debris properties required by each model. The parameters are as follows: h_d = debris thickness, k = thermal conductivity, z_0 = aerodynamic surface roughness length, ε = debris surface emissivity, c_d = debris specific heat capacity, ρ_d = debris density, ϕ = debris porosity, u^* = friction velocity, γ = wind speed attenuation constant, and Φ = fraction of debris embedded in ice.

Model	Parameters
DEB_{CF}	h_d, k, z_0, ε, c_d, ρ_d
ROU15	h_d, k, z_0, ε, c_d, ρ_d
$GRO17_B$	h_d, k, z_0, ε, u^* , γ, Φ
d2EB	h_d, k, z_0, ε, ϕ, ρ_d, c_d
DEB_{PG}	h_d, k, z_0, ε, c_d, ρ_d
$GRO17_A$	h_d, z_0, k, ε, Φ, u^*
A-Melt	h_d, k, ε, ϕ, c_d, ρ_d
THRED	h_d, k, z_0, ε
MCC19	h_d, k, c_d, ρ_d

2.3. Enhanced temperature index models

The KO2 model is an enhanced temperature index model and is considered in this category in this study. However in this Supplement the KO2 model is described after KM1 and KP1 in the temperature-index models section because it has several similarities with those.

$DETI_m$

The debris enhanced temperature index model (DETI) was developed by Carenzo et al. (2016) and includes a temperature factor (TF), a shortwave radiation factor (SRF), and a lag factor (lag) to account for the time it takes to transfer energy through the debris. Each of these three parameters need to be calibrated as a function of debris thickness, as in equations 6, 7 and 8 in Carenzo et al. (2016).

A slight modification with respect to the original model has been implemented for this intercomparison experiment, and therefore we introduce the *m* subindex in the DETI acronym ($DETI_m$). Here, the term *i-lag*, accounting for the time lag of the energy transfer between the debris surface and the ice surface, is also applied to the temperature threshold condition. That is, melt is set to zero if the temperature was below T_T at the time step *i-lag*. Conversely, melt occurs if the temperature at the time step *i-lag* was above T_T (see Equation 1 below). This accounts for the delay in energy transfer through the debris layer also to the temperature threshold. For instance, if temperature drops to zero after a few hours of high temperatures, melt will only stop after the lag from the last time step with temperature above zero has reached the ice. Conversely, after a cold period, melt will not start immediately after temperature crosses T_T, but will start after that lag time has passed.

$$M = \begin{cases} TF \cdot T(i - lag_T) + SRF \cdot (1 - \alpha) \cdot I(i - lag_I) & T(i - lag) > T_T \\ 0 & T(i - lag) \leq T_T \end{cases} \quad \text{(S28)}$$

The DETI model is calibrated against hourly melt rate simulations from the DEB_{CF} model, as in the original paper by Carenzo et al. (2016) where the model was developed. For each site, the *lag*, *TF* and *SRF* parameters are parameterised as a function of debris thickness. Therefore, the DEB_{CF} runs for the debris thicknesses used in Experiment 3 of the model intercomparison (1, 2, 4, 6, 8, 10, 12, 15, 20, 30, 50, 100 cm) were used. For each thickness, a range of parameters was used to run the model and the highest Nash and Sutcliffe Efficiency between DETI and DEB_{CF} was sought to obtain the optimal parameters. With the optimal parameters for each thickness, equations 2, 3 and 4 are derived, and the parameters lag_1, lag_2, TF_1, TF_2, SRF_1 and SRF_2 obtained from them. We assumed that the optimal curve fitting functions are the same as those in Carenzo et al. 2016, i.e. linear for the lag parameter, power law for TF and exponential for SRF. The optimal parameters obtained are in Table S21.

$$lag = lag_T = lag_I = lag_1 \cdot d + lag_2 \quad \text{(S29)}$$

$$TF = TF_1 \cdot d^{TF_2} \quad \text{(S30)}$$

$$SRF = SRF_1 \cdot e^{SRF_2 \cdot d} \quad \text{(S31)}$$

Table S21. $DETI_m$ calibrated and uncalibrated model parameters.

Site	lag_1	lag_2	TF_1	TF_2	SRF_1	SRF_2
AROLLA	21.93	-1.23	0.019	-0.526	0.0081	-13.8533
CHANGRI NUP	24.56	-1.45	0.017	-0.780	0.0039	-13.9848
DJANKUAT	10.99	-0.55	0.044	-0.212	0.0095	-6.1140
LIRUNG	15.34	-0.67	0.069	-0.127	0.0114	-10.3353
MIAGE	19.97	-1.00	0.021	-0.441	0.0093	-13.1247
PIRAMIDE	19.97	-1.00	0.011	-0.611	0.0073	-13.5738
SULDENFERNER	25.44	-1.05	0.017	-0.368	0.0083	-13.8533
TAPADO	21.93	-1.23	0.015	-0.667	0.0061	-12.0041
TASMAN	12.98	-0.65	0.032	-0.475	0.0095	-8.8402
Uncalibrated (Carenzo et al. 2016)	21.54	-1.19	0.016	-0.621	0.0079	-11.2100

2.4. Temperature index models

KM1 and KP1

These are temperature-index models configured as in Winter-Billington et al. (2020). ~~with KP1 being the uncalibrated version and KM1 the calibrated version of~~ The two models share the same model structure with the same fixed-effects parameters, but different random-effects parameters. The models were fitted using data from 27 debris-covered glaciers around the world (Winter-Billington et al., 2020). They estimate the underlying relation between melt, air temperature, and debris thickness (fixed effects) while accounting for the statistical variation of melt factors among glaciers and among glacier-years (random effects). Total ice melt over a period of time is predicted in these models as the product of modelled melt factors and daily mean positive degree days. The main model equations are as follows:

$$\mathrm{PDD} = \delta d \sum_{i=1}^{n_d} (T_i - T_b)\,\delta(T_i) \qquad \text{with} \qquad \delta(T_i) = \begin{cases} 1, & \text{if } T_i \geq T_b, \\ 0, & \text{if } T_i < T_b. \end{cases} \quad \text{(S32)}$$

where subscript i is one day in the observation period of n_d days, T_i is mean air temperature on day i, T_b is a threshold air temperature below which ablation ceases and $\delta(T_i)$ is a binary variable that sets ablation on day i to zero when $T_i \leq T_b$, and

$$melt = k \times D_i \quad \text{(S33)}$$

where k is a predicted melt factor (mm day^{-1} °C^{-1}) and D_i is daily mean PDD (PDD/n_d).

~~where PDD is calculated as the sum of positive degree days above a temperature threshold T_0, and~~ The value of k is the melt factor (mm day^{-1} C^{-1}) dependent **on debris thickness** (h) and fitted parameters according to equation:

$$k = 10^{(b_0 + b_{0i} + b_{0ij}) + (b_1 + b_{1i} + b_{1ij}) \times h} \quad \text{(S34)}$$

where g is glacier identity and y is the nested glacier-year. The difference between KM1 and KP1 is the data used in model fitting, which resulted in different fitted coefficient values (see Winter-Billington et al. (2020) for details). The fixed- effects coefficients for models KM1 and KP1 are in Table S22. The random-effects coefficients for models KM1 and KP1 are in Tables S23 to S26 (see Winter-Billington et al. (2020) for more details).

~~For both versions, the parameters in (2) are original values fitted using data collated from the literature (independent data – not the same data used in this intercomparison experiment) (Table S21).~~

The only model component that is calibrated is T_0 for the calculation of ~~PD~~Di as input to model KM1. For the calibrated runs (~~KP1~~model KM1), the fitted values of T_0 are shown in Table S22. For uncalibrated runs (~~KM1~~model KP1), $T_b = 0$.

With the time series of Ti that were provided for each site, time series of Di were calculated using ~~Values of PDD were calculated for each glacier using a range of~~ T_0 = [-10, 9]. The model was run using every set of ~~PD~~Di. The value of T_0 that

resulted in the smallest RMSE ~~difference~~ between cumulative observed and predicted melt was selected (Table S22). The model parameters ~~in (1)~~ were fitted assuming $T_0 = 0$, so the calibrated values of T_0 are not physically meaningful.

Table S22. Summary of fixed-effects coefficients per site, for models KM1 and KP1, as well as the calibrated values of T_b for each site for KM1.

Site	KM1			KP1	
	b_0	b_1	T_b	b_0	b_1
AROLLA	0.4937	-1.4460	-2	0.6215	-1.4572
CHANGRI NUP	0.2691	-1.4614	-8	0.6215	-1.4572
DJANKUAT	0.5981	-1.4179	0	0.6215	-1.4572
LIRUNG	0.5583	-1.4179	-1	0.6259	-1.4034
MIAGE	0.4422	-1.4461	-3	0.6223	-1.4505
PIRAMIDE	0.5479	-1.4405	-1	0.6215	-1.4572
SULDENFERNER	0.4465	-1.4506	-3	0.6215	-1.4572
TAPADO	0.5981	-1.4179	0	0.6215	-1.4572
TASMAN	0.2150	-1.4630	-10	0.6215	-1.4572

Table S23. Random effects coefficients for model KM1. Glacier-specific coefficients.

Glacier	β_0	β_1
Baltoro glacier	0.0357340606696045	-0.183784796122387
Barpu Glacier	0.0142430361478071	-0.0732537392184386
Batal glacier	-0.1139397559226	0.586006601426085
Chorabari Glacier	-0.0936873740159294	0.481846023280708
Dokriani Glacier	-0.159065988224529	0.818096543753797
Eliot Glacier	0.260196953533077	-1.33822591232525
Franz Josef Glacier	0.000503675392434745	-0.00259046652936516
Ghiacciaio del Belvedere	0.0559567421848158	-0.287792610949187
Hailuogou Glacier	0.0804089622140197	-0.413553477827503
Koxkar Glacier	-0.147602975966774	0.759140816554963
Larsbreen	0.0316679375651566	-0.162872218535068
Lirung Glacier	0.0340603702227083	-0.175176819106456
Miage Glacier	0.0048722633088835	-0.0250586613974906
Qingbingtan Glacier No. 72	-0.0567097700710563	0.291665539382937
Rakhiot Glacier	0.0591447276272265	-0.304188830014309
Southern Inylchek Glacier	0.0452024202172116	-0.232481775248828
Venerocolo Glacier	-0.0464798277217959	0.239051640743079
Vernagtferner	-0.00450545716029045	0.0231721421328688

Table S24. Random effects coefficients for model KM1. Glacier-year-specific coefficients.

Glacier:year	β_0	β_1
Baltoro glacier:2004	0.121258503701998	0.122278149313058
Barpu Glacier:1987	0.0156266115392138	0.0157580135059646
Batal glacier:2014	-0.0759629069769631	-0.0766016683201904
Chorabari Glacier:2003-2010	-0.300552483575998	-0.303079786910736
Chorabari Glacier:2010	0.0298298993709972	0.0300807347767082
Dokriani Glacier:2010	0.0241620317599601	0.0243652068683952
Dokriani Glacier:2011	-0.0240436636760613	-0.0242458434440167
Dokriani Glacier:2012	-0.00353733119957876	-0.00356707611743936
Dokriani Glacier:2013	0.104982877014882	0.105865663182587
Eliot Glacier:1987- 1989	-0.0377894787305967	-0.0381072451136174
Franz Josef Glacier:2012	-0.00259775140087978	-0.0026195955250745
Ghiacciaio del Belvedere:2002	0.108843577109364	0.10975882735824
Hailuogou Glacier:2009	0.106738217036862	0.107635763610597
Koxkar Glacier:2004	0.0613835685357198	0.0618997342835874
Larsbreen:2003	-0.0340250041984198	-0.0343111156474153
Lirung Glacier:1995	0.0243306475161885	0.0245352404906747
Lirung Glacier:2013	0.0551957186630032	0.0556598516562845
Lirung Glacier:2014	-0.396290376691568	-0.399622726431689
Miage Glacier:2005	0.0535715883768126	0.0540220642881777
Miage Glacier:2010	0.0324151758431857	0.03268775047318
Miage Glacier:2011	0.0244809107344371	0.0246867672510775
Qingbingtan Glacier No. 72:2008	0.0758207738285469	0.0764583399942274
Rakhiot Glacier:1986	0.00615363159401982	0.00620537661193907
Southern Inylchek Glacier:2005	0.0177847873556592	0.0179343371176734
Venerocolo Glacier:2007	-0.0737453147321626	-0.0743654286557821
Vernagtferner:2010	0.0859657912014107	0.0866886653836225

Table S25. Random effects coefficients for model KP1. Glacier-specific coefficients.

Glacier	β_0	β_1
24k glacier	-0.0513122597179868	0.238585166994192
Baltoro glacier	-0.000598337394081189	0.00278207250801877
Barpu Glacier	-0.00662116445614081	0.0307862416534858
Batal glacier	-0.0810552911548811	0.376880501506569
Chorabari Glacier	-0.141233390473522	0.656688912873378
Djankuat Glacier	0.00892381079289292	-0.0414927913903704
Dokriani Glacier	-0.167670640338444	0.779613448034292
Eliot Glacier	0.271686662329542	-1.26325381221257
Franz Josef Glacier	-0.0130934137054091	0.0608800764690168
Ghiacciaio del Belvedere	0.0464673851539856	-0.216058090360439
Hailuogou Glacier	0.0339672459261235	-0.157936545499207
Ice-cored Fox Glacier Moraine	0.0579433215891738	-0.269417428379309
Khumbu Glacier	0.0747926300215065	-0.34776118264308
Koxkar Glacier	-0.127554834824709	0.593088118411241
Larsbreen	0.0298702576801357	-0.138886894787769
Lirung Glacier	0.0165455990074614	-0.0769316051156168
Maliy Aktru	0.0352657909355486	-0.163974353610288
Miage Glacier	0.00935199836185992	-0.0434837230548166
Qingbingtan Glacier No. 72	-0.03405061833338	0.158324199830108
Rakhiot Glacier	0.0281443510389046	-0.130861995348968
Southern Inylchek Glacier	0.0804509502631034	-0.37407051470461
Summit Crater Glacier	0.0457376248258707	-0.212664944342198
Svínafellsjökull	-0.0513805148502741	0.238902530957979
Venerocolo Glacier	-0.0305396433717916	0.141999318561444
Vernagtferner	-0.0142434356348657	0.0662273010035504
Zopkhito Glacier	-0.0197940836708316	0.0920359926469375

Table S26. Random effects coefficients for model KP1. Glacier-year-specific coefficients.

Glacier:Year	β_0	β_1
24k glacier:2008	0.0240883566892971	0.0582565436236891
Baltoro glacier:2004	0.0159771345852822	0.0386399392019384
Barpu Glacier:1987	-0.0427784001953234	-0.103457523868276
Batal glacier:2014	0.0553787493979968	0.13393086841674
Chorabari Glacier:2003-2010	-0.234272539312624	-0.566576980834992
Chorabari Glacier:2010	0.0179004850394911	0.0432914706901386
Djankuat Glacier:2007	0.00589414175292368	0.0142547011646491
Dokriani Glacier:2010	0.0216836054138076	0.0524407671723665
Dokriani Glacier:2011	-0.0106504367408129	-0.0257575741095776
Dokriani Glacier:2012	0.0114889875766503	0.0277855693762846
Dokriani Glacier:2013	0.095923494245830 6	0.231986403188386
Eliot Glacier:1987- 1989	-0.0207367232494906	-0.050150777746208
Franz Josef Glacier:2012	-0.060636420035475	-0.146646294495902
Ghiacciaio del Belvedere:2002	0.0528924858250661	0.127917958355475
Hailuogou Glacier:2009	0.015525174784138	0.0375468959441054
Ice-cored Fox Glacier Moraine:2007	-0.00624997167821862	-0.0151152588952132
Khumbu Glacier:1999	0.327258636495271	0.791459429097824
Koxkar Glacier:2004	0.254123631601645	0.614585902274912
Koxkar Glacier:2010	-0.161925960996869	-0.391610226147693
Larsbreen:2003	0.019435035098718	0.0470027069368109
Lirung Glacier:1995	0.0849223550501107	0.20538067189122
Lirung Glacier:2013	0.0391842662233663	0.0947652820010778
Lirung Glacier:2014	-0.344971135240858	-0.834296263887609
Maliy Aktru:2007	-0.17752924922947	-0.429346035747331
Miage Glacier:2005	0.0264769316595675	0.0640331984511162
Miage Glacier:2010	-0.0431587960706609	-0.104377493179246
Miage Glacier:2011	-0.0228850941217367	-0.0553465104005023
Qingbingtan Glacier No. 72:2008	0.11709403870841	0.283186356880931
Rakhiot Glacier:1986	-0.0965425089034547	-0.233483460661867
Southern Inylchek Glacier:2005	-0.024007803110899	-0.0580617286301002
Summit Crater Glacier:2008	0.196198999245799	0.474497937153413
Svínafellsjökull:2013	-0.148115619288772	-0.358210572341782
Venerocolo Glacier:2007	-0.0581675440027374	-0.140675435372641
Vernagtferner:2010	0.0717589733940422	0.173545660164465
Zopkhito Glacier:2009	-0.000577280610070789	-0.0013961256667476

KO2 (ETI)

This model is configured as KM1/KP1 with a modified calculation of the melt factor k, now dependent on both debris thickness (h) and **mean net shortwave radiation** (SW_{net}):

$$k = 10^{b_0 + b_1 \times h + b_2 \times SW_{net}} \quad \text{(S35)}$$

~~The parameters were not recalibrated but taken from the originally fitted model in Winter-Billington et al. (2020).~~

In Winter-Billington et al. (2020), the models were fit using the data that were also used in this intercomparison project, from Lirung Glacier and Miage Glacier. Therefore, the models were refit using the data in Winter-Billington et al. (2020) with the data from Lirung and Miage Glaciers excluded. The refitted fixed-effect coefficient values for KO2 are in Table S27, and random effects coefficients in Table S28.

Table S27. Summary of fixed-effects coefficients for the KO2 model.

Parameter	b_0	b_1	b_2
All sites	-0.6783	-1.4442	0.0058

Table S28. Random-effects coefficients for model KO2. Glacier (top) and glacier-year (bottom) specific coefficients.

Glacier	β_0	β_1
24k glacier	0.00407425627772242	-0.00028907189192505
Baltoro glacier	-0.0986121343590905	0.00699661519261567
Djankuat Glacier	-0.0124186768013377	0.000881115701927656
Franz Josef Glacier	0.0806347709649952	-0.00572110585442618
Ice-cored Fox Glacier Moraine	-0.144299356452944	0.010238162505367
Khumbu Glacier	0.302254699982923	-0.0214452290458508
Koxkar Glacier	-0.189638418599908	0.0134550077939038
Maliy Aktru	-0.159261366675735	0.0112997299167774
Southern Inylchek Glacier	0.0276884323677957	-0.00196451798498316
Summit Crater Glacier	0.0465263205303233	-0.00330108218456408
Svínafellsjökull	0.0884292163233635	-0.00627412841763036
Zopkhito Glacier	0.0546222564419513	-0.003875495731216
Glacier:Year	β_0	β_1
24k glacier:2008	$1.74986897826604 \times 10^{-6}$	$1.30962443856709 \times 10^{-7}$
Baltoro glacier:2004	$-4.50859186903996 \times 10^{-5}$	$3.37428811445672 \times 10^{-6}$
Djankuat Glacier:2007	$-5.63357348437098 \times 10^{-6}$	$4.21623881743798 \times 10^{-7}$
Franz Josef Glacier:2012	$3.67970669243553 \times 10^{-5}$	$-2.75393979265104 \times 10^{-6}$
Ice-cored Fox Glacier Moraine:2007	$-6.57671733778732 \times 10^{-5}$	$4.92209980180854 \times 10^{-6}$
Khumbu Glacier:1999	$1.3790846587808 \times 10^{-4}$	$-1.03212468729056 \times 10^{-5}$
Koxkar Glacier:2010	$-8.66353455218419 \times 10^{-5}$	$6.48390063189397 \times 10^{-6}$
Maliy Aktru:2007	$-7.26105408702559 \times 10^{-5}$	$5.43426622234825 \times 10^{-6}$
Southern Inylchek Glacier:2005	$1.27372621257994 \times 10^{-5}$	$-9.53273071730848 \times 10^{-7}$
Summit Crater Glacier:2008	$2.12535979209276 \times 10^{-5}$	$-1.59064659071256 \times 10^{-6}$
Svínafellsjökull:2013	$4.03685491116543 \times 10^{-5}$	$-3.02123411082499 \times 10^{-6}$
Zopkhito Glacier:2009	$2.49177410056761 \times 10^{-5}$	$-1.86487576957076 \times 10^{-6}$

Hyper-fit

The Hyper-fit model is a temperature-index melt model, developed by Anderson and Anderson (2016). The model takes a bare-ice melt rate and decreases that melt rate due to the insulating effects of debris cover. The shape of the debris thickness-melt relationship output by the model is hyperbolic and the rate of decline of that curve is defined by a characteristic debris thickness scale h_*.

In the model, the bare-ice melt rate $\dot{b}_{ice}$ is estimated using a degree-day factor for bare-ice DDF_{ice}:

$$\dot{b}_{ice} = DDF_{ice} \times T^{+} \tag{S36}$$

where and T^{+} is the positive degree-days defined as the mean hourly air temperature at 2m above the surface (when above 0° C) . Despite using an hourly time step the degree day nomenclature is still used as is convention (e.g., Hock, 2003).

At each hourly time step, the bare ice melt rate is then reduced based on:

$$\dot{b}_{debris} = \dot{b}_{ice} \left(\frac{h_*}{h_* + h_{debris}} \right) \tag{S37}$$

where $\dot{b}_{debris}$ is the sub-debris melt rate, h_{debris} is the local debris thickness, and h_* is the characteristic debris thickness scale. If ice is assumed to be at the freezing point, h_* can be estimated from physical inputs and parameters following:

$$h_* = \frac{kR}{(1-\phi)} \tag{S38}$$

where k and φ are the thermal conductivity and porosity of the debris cover and R is the thermal resistance of the debris layer. Here we define R as:

$$R = \frac{T_s^-}{L\rho_{ice}\dot{b}_{ice}} \tag{S39}$$

where L and ρ_{ice} the latent heat of fusion and density of ice, $\bar{T_s}$ the average debris surface temperature over the period used to estimate h_* and $\dot{b}_{ice}$ is the bare-ice melt rate over the period used to estimate h_*. We consider the case in which this model has two parameters: DDF_{ice} and h_*.

This model assumes that heat is transferred through debris by conduction. Sub-debris melt should therefore vary inversely with debris thickness (excluding the possibility for melt enhancement for debris less than ~ 3 cm). In other words the debris thickness-melt relationship (or Østrem's curve) for a given site should be hyperbolic as conduction is governed by the temperature gradient within the debris (e.g., Nicholson & Benn, 2006). In this formulation, sub-debris melt rates approach bare-ice melt rates and debris thins ($h_{debris} << h_*$) and asymptotes toward zero melt as debris thickens ($h_{debris} >> h_*$).

There are two input variables needed for this model: 1) 2 m air temperature, which can be from on glacier or off glacier sources; and 2) local debris thickness. If desired the model can also use thermal conductivity, thermal resistivity, and debris porosity as inputs, but in practice estimating h_* from empirical data appears to be effective. The model has two parameters, the bare-ice degree-day factor DDF_{ice} and the characteristic debris thickness h_*.

Uncalibrated (estimated) parameters

For uncalibrated simulations, values for the two parameters were derived from previous publications, independent from the data provided in this intercomparison project and therefore regarded as estimated instead of uncalibrated.

h_* values can be estimated if debris thickness-melt data (or Østrem's curves) are available. With debris thickness-melt data from each site, independent from the data provided in this intercomparison project, the modeller optimised both h_* and $\dot{b}_{ice}$ to obtain a best fit curve using Scipy's curve_fit package (see Table SI XXX). If no Østrem curve was available the global mean value of 0.066 m was applied (Anderson and Anderson, 2016). In the case of the SDF site debris thickness-melt data were available but the data was too noisy to produce a viable estimate of h_*. For SDF debris less than 3 cm thick melt values ranged between 8.7 and 2.2 cm/day. The noise precluded the estimate of a unique, viable h_* value. For TAS debris thickness-melt values were available but varied between 10 m/yr and 0.5 m/yr between 18 and 28 cm debris thickness, so the global mean h_* value of 0.066 m was applied.

For the uncalibrated (estimated) h_* values the mean was 0.085±0.034 m (1σ) and ranged between 0.055 to 0.16 m. These numbers are similar to those presented in the global compilation of h_* values from Anderson and Anderson (2016) the h_* mean was 0.066±0.029 m (1σ) and ranged between 0.03 and 0.13 m.

The DDF_{ice} parameters applied at each site were derived from previous publications from each site, with two exceptions. In the case of PIR, a DDF_{ice} was applied from adjacent glaciers 2-5 km away (Bello and Yeso Glaciers; Table 1). In the case of CN, a DDF_{ice} was applied from Khumbu Glacier, 5 km away (Table 1). If DDF_{ice} values were provided in a publication those values were used. Otherwise the values of DDF_{ice} were calculated by the modeller using melt and air temperature data provided in each publication. For the uncalibrated DDF_{ice} parameters the mean was 0.45±0.27 mm/hr/K (1σ) and ranged between 0.023 to 1.1 mm/hr/K.

Calibrated parameters

Of the two hyper-fit model parameters one was calibrated. The characteristic debris thickness h_* was calibrated with the cumulative melt for each hour within the full measurement period at each site. The root mean squared error was then minimised for viable parameter choices. For sites that do not provide mean hourly melt rates (i.e., melt was measured using ablation stakes) the cumulative melt was interpolated to each hour using the mean melt rate provided by the ablation stakes. For the h_* values calibrated here the mean was 0.066±0.013 m (1σ) and ranged between 0.049 and 0.089 m. This excludes the two outliers, likely caused by low DDF_{ice} values from the literature (Table 1). These numbers are similar to those presented in the global compilation from Anderson and Anderson (2016) the h_* mean was 0.066±0.029 m (1σ) and ranged between 0.03 and 0.13 m.

The second parameter, the DDF_{ice} was not calibrated during the simulations in which h_* was calibrated. Instead at each site a value from the literature was applied (see below). The DDF_{ice} in principle could have been calibrated with the data provided for each site but no calibration was performed here.

Table S24. Hyper-fit literature (lit.) and calibrated (cal.) parameter values for each site.

Glacier	DDF_{ice} [mm hr^{-1} K^{-1}]	Reference for DDF_{ice}	h* (lit.) [m]	Reference for h* (lit.)	h* (cal.) [m]
ARO	0.5^{+}	Pellicciotti et al., 2005	0.075^{i}	Reid et al., 2012	0.056
CN	0.7^{+}	Kayastha et al., 2000	0.078^{i}	Lejeune et al., 2013	$>0.5^{\#}$
DJA	0.31^{+}	Lambrecht et al., 2011	0.16^{+}	Bozhinskiy et al., 1986	$0.45^{\#\#}$
LIR	0.33^{+}	Chand and Kayastha, 2018	0.068^{i}	Chand and Kayastha, 2018	0.066
MIA	$0.23^{\dagger}$	Brock et al., 2010	0.13^{i}	Brock et al., 2010	0.089
PIR	$0.3^{\dagger,\circ}$	Ayala et al., 2017	0.055^{i}	Ayala et al., 2016	0.049
SDF	0.36^{+}	M. McCarthy, personal comm.	$0.066^{\star}$	Global mean	0.06
TAP	$1.1^{\dagger}$	Ayala et al., 2017	$0.066^{\star}$	Global mean	0.061
TAS	$0.25^{\dagger}$	Kirkbride, 1995	$0.066^{\star}$	Global mean	0.083

$^{+}$ Reported value from literature or personal communication. † Calculated based on available bare-ice melt and air temperature data from cited paper. ^{i} Based on available debris thickness-melt data from cited paper. $^{\circ}$ For the PIR case, a DDF_{ice} was applied from Bello and Yeso Glaciers 2-5 km to the northwest. * No debris thickness-melt data was available from the literature so the global mean based on Anderson and Anderson (2016) was applied. For SDF and TAS debris-thickness-melt data were available but the data was too noisy to produce a viable h* value. $^{\#}$ The calibrated h* for CN never reaches a viable minimum for the CN case. Ultimately, a higher DDF_{ice} value would reduce the calibrated h* into a reasonable range. The model appears to be especially sensitive to input parameters for sites with low air temperatures but with high relative shortwave radiation input, as should be the case for all temperature index models solely dependent on air temperature. Note that in simulations for CN a calibrated value of 50 cm is used for h*, while a larger value could be used, the results do not change much even with a 200 cm value. $^{\#\#}$ The calibrated h* for DJA is a strong outlier amongst the population of calibrated h* values and the global compilation of h* values. Note that from the literature on DJA, Lambrecht et al. (2011) mentions a value for h* of 10 cm, their debris thickness-melt data produce an h* of 8.4 cm, but a back of the envelope calculation based on provided debris properties produces a much higher h* value. Using equation 3 with $k = 2.8$ W/m/K, $R = 10.2$ K m^2/W (the mean reported value from Lambrecht et al. (2011) for DJA), and $\phi = 0.42$, then $h* = 49$ cm. For extreme values of R from DJA, $h*$ can range from 34 to 75 cm. The modeler chose the reported value of 16 cm from Bozhinskiy et al. (1986). Ultimately, a higher DDF_{ice} value would reduce the calibrated h* into a viable range. This may also point to where the hyper-fit model assumptions break down.

DDF_{debris}

The simplest form of temperature index model with a single parameter is applied. Sub-debris melt b_{debris} is calculated using a degree-day factor DDF_{debris}:

$$\sum_{i=1}^{n} b_{debris} = DDF_{debris} \times \sum_{i=1}^{n} T^{+} \Delta t \tag{S40}$$

where and T^{+} is the positive degree-days defined as the mean hourly air temperature at 2 m above the surface (when above 0° C) and Δt is one hour. Despite using an hourly timestep the degree day nomenclature is still used as is convention (e.g., Hock, 2003).

The model accounts for the effect of debris by reducing (or increasing) the degree-day factor for bare ice to represent the melt suppressing (enhancing) effects of debris. As debris thickness is not used by the model, this model is not run in uncalibrated form, as this would not make sense.

The sub-debris degree-day factor was calibrated with the cumulative melt for each hour within the full measurement period at each site. The root mean squared error was then minimized to find the best-fit sub-debris degree-day factor. For sites that do not provide mean hourly melt rates (i.e., melt was measured at periods longer than 1 hour from ablation stakes) the cumulative melt was interpolated to each hour using the mean melt rate provided from the ablation stake measurements.

Table S25. Calibrated DDF_{debris} parameter values for each site.

Glacier	DDF_{debris} [mm hr^{-1} $°C^{-1}$]	RMSE [cm]
ARO	0.24	11
CN	0.69	16
DJA	0.13	3.8
LIR	0.06	3.8
MIA	0.066	5.3
PIR	0.064	6.4
SDF	0.125	2.1
TAP	0.078	3
TAS	0.055	1.4

Table S26. Model execution time for one single standard run over Miage, excluding the time to load and export the data, and programming language in which the model is written.

Model Execution time	Execution time (s)	Processor used	Model code language
$GRO17_A$	0.05	i7 processor	R
$GRO17_A$	0.05	i7 processor	R
DEB_{CF}	~1.8	i5 processor	Matlab
ROU15	~2	i7 processor	Python
A-Melt	~0.1	i5 processor	Fortran 90
THRED	~1	Xeon Bronze	Fortran 90
d2EB	~20	i7 processor	R
DEB_{PG}	~4	i5 processor	Matlab
MCC19	~4	i7 processor	Matlab
$DETI_m$	~0.001	i7 processor	Matlab
KM1/KP1	~0.003	DDR3	R
KO2	~0.001	DDR3	R
Hyper-fit	~0.00003	i7 processor	Python
DDF_{debris}	~0.000008	i7 processor	Python

3. Results

3.1. Performance of model ensemble at sites

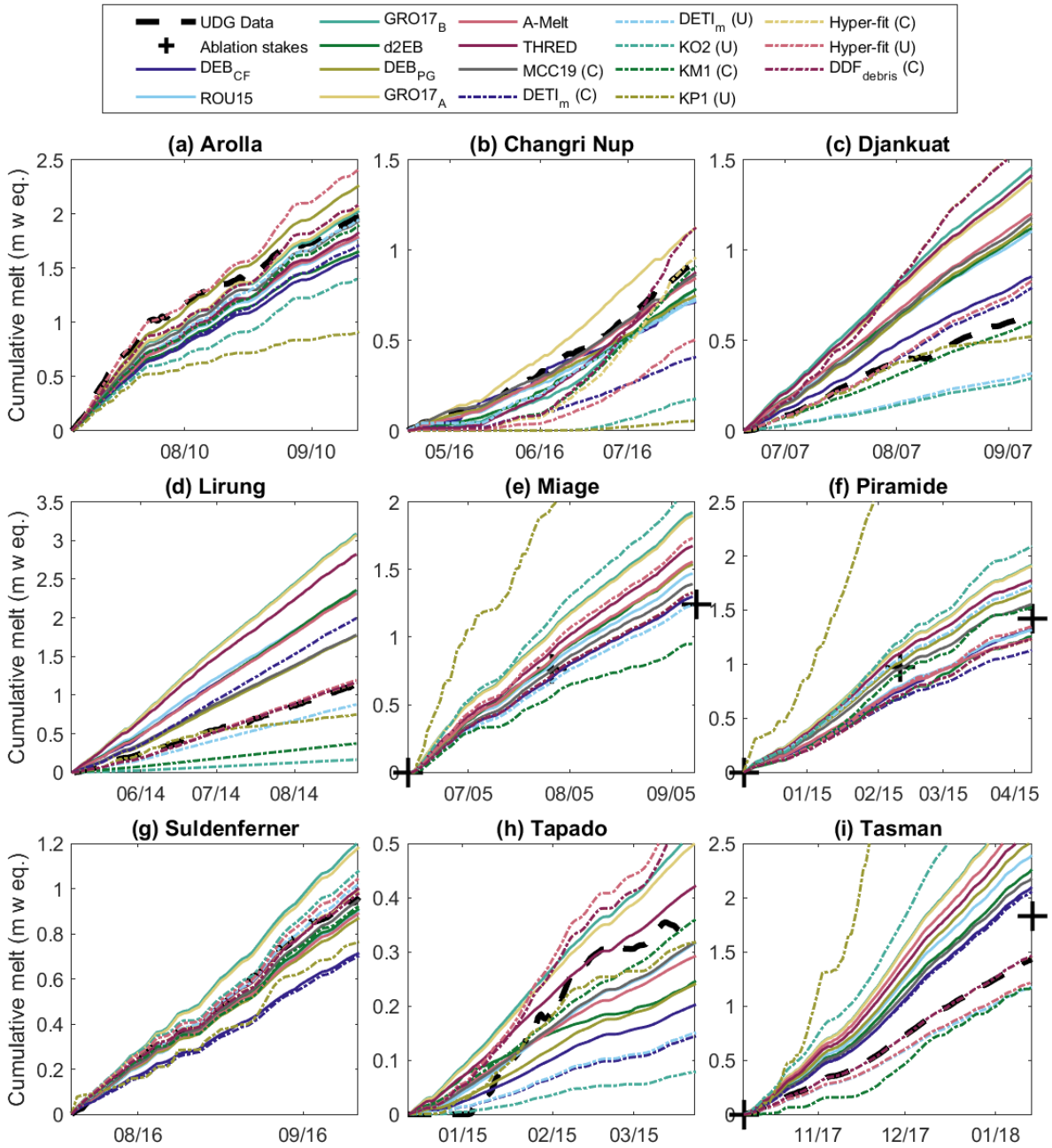


Figure S2. Comparison of cumulative melt simulations across all models with observed melt data at each site.

Table S27. Root mean square error (°C) of modelled surface temperature of debris across models and across sites at the hourly (h) and daily (d) scale. The last two columns correspond to the median and interquartile range (IQR) across sites per model, and the last row corresponds to the median across models per site.

Model	Resolution	ARO	CNU	DJA	LIR	MIA	PIR	SDF	TAP	TAS	Median (%)	IQR (%)
DEB_{CF}	h	2.3	2.8	4.8	7.7	2.5	4.5	2.7	3.0	3.3	3.0	1.9
DEB_{CF}	d	1.4	1.1	2.8	5.9	1.5	3.2	1.3	1.1	3.1	1.5	1.9
ROU15	h	3.0	3.8	3.9	6.8	3.8	4.5	5.5	5.1	2.9	3.9	1.6
ROU15	d	2.0	2.1	2.1	5.7	2.9	2.7	3.4	3.3	1.5	2.7	1.3
$GRO17_B$	h	3.7	5.6	7.6	15.6	6.6	5.9	8.4	11.2	7.1	7.1	3.2
$GRO17_B$	d	2.2	3.1	2.1	6.5	2.8	1.1	4.2	3.8	1.4	2.8	2.0
d2EB	h	2.9	4.6	4.8	7.0	5.6	6.8	5.5	9.7	4.4	5.5	2.3
d2EB	d	1.7	3.1	3.2	5.7	4.2	4.7	3.6	6.1	2.6	3.6	1.9
DEB_{PG}	h	1.5	2.6	3.3	2.0	2.0	3.6	3.3	2.8	3.2	2.8	1.3
DEB_{PG}	d	0.7	0.9	1.5	2.0	1.1	2.5	0.9	0.8	2.8	1.1	1.3
$GRO17_A$	h	3.8	5.9	8.4	15.9	7.0	6.1	8.6	12.1	7.2	7.2	3.5
$GRO17_A$	d	2.2	3.2	2.3	7.1	3.0	1.0	4.3	4.3	1.4	3.0	2.3
A-Melt	h	3.3	5.3	5.3	7.2	4.6	4.1	4.0	4.9	4.4	4.6	1.2
A-Melt	d	2.1	4.2	3.0	5.3	3.7	3.2	3.3	2.3	2.5	3.2	1.4
THRED	h	-	-	-	-	-	-	-	-	-	-	-
THRED	d	2.3	4.0	4.6	8.4	4.7	1.2	4.7	6.6	1.3	4.6	3.1
MCC19	h	3.2	4.4	3.4	2.0	2.7	2.8	4.2	3.7	2.7	3.2	1.1
MCC19	d	3.0	3.8	2.1	2.4	2.3	0.8	4.0	3.4	2.4	2.4	1.2
Median (%)	h	3.1	4.5	4.8	7.1	4.2	4.5	4.9	5.0	3.8	Overall	4.5
Median (%)	d	2.1	3.1	2.3	5.7	2.9	2.5	3.6	3.4	2.4	Overall	2.8

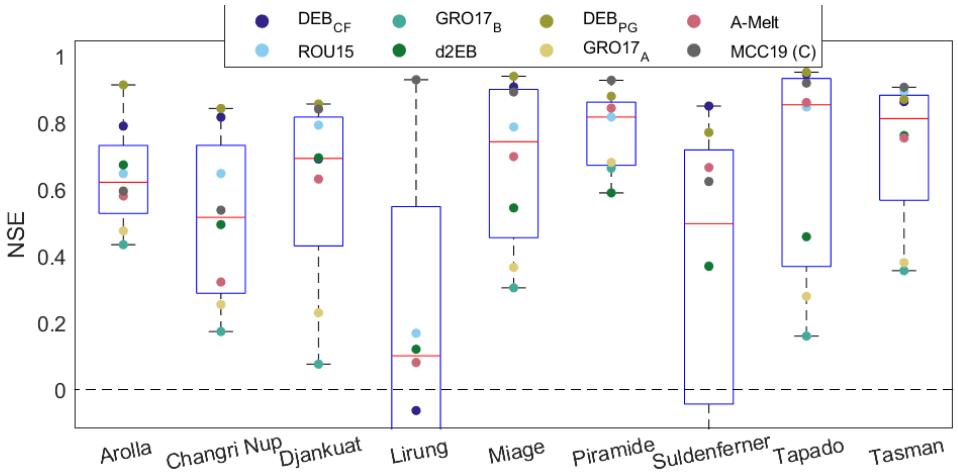


Figure S3. Validation of the energy balance model simulations against surface temperature across sites. Statistical boxes for each site show the distribution of the Nash Sutcliffe Efficiency of surface temperature for each model. Note that the THRED model is run at daily resolution and therefore not validated with NSE as discussed in the text. $GRO17_A$ and $GRO17_B$ are negative at Lirung (-4.1 and -3.4, respectively) and Suldenferner (-0.55 and -0.46, respectively) but are not displayed for better visibility of the plot.

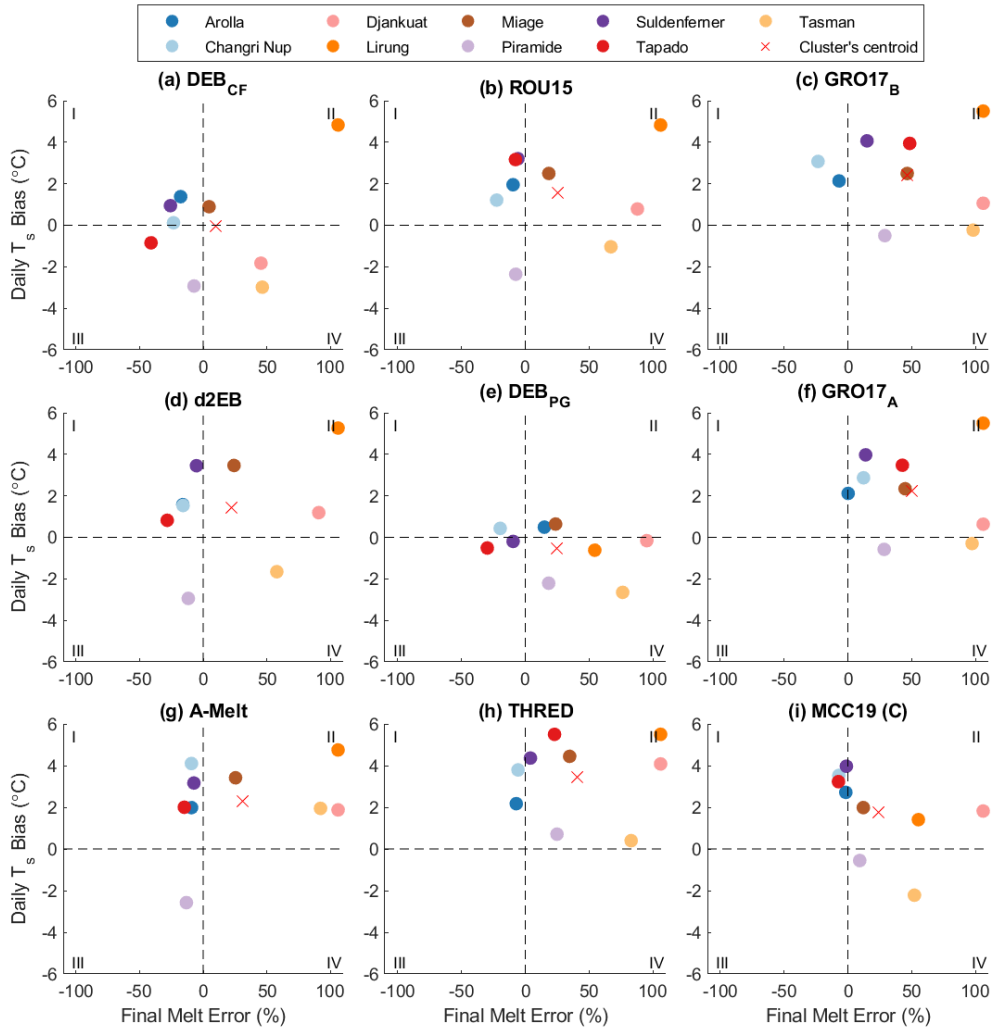


Figure S4. Consistency of model performance across the two validation datasets. For each model (a-i), each site is scattered based on their daily temperature bias and melt error. Dashed lines correspond to the zero line for both axes, and separate the plot in four quadrants. Quadrants above the horizontal dashed line indicate overestimation of surface temperature. Quadrants to the right of the vertical dashed line indicate overestimation of melt. Values higher or lower than the axes limit are shown on the axis limit.

3.2. Individual model performance

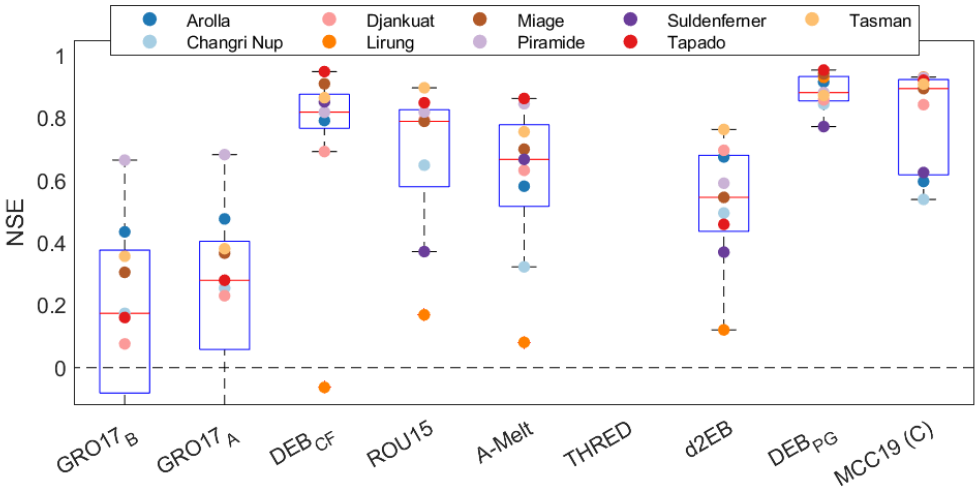


Figure S5. Validation of the energy balance model simulations against surface temperature, at all sites. Statistical boxes for each model show the distribution of the Nash Sutcliffe Efficiency of surface temperature at each site. Note that the THRED model is run at daily resolution and therefore is not validated with NSE as discussed in the text. GRO17$_A$ and GRO17$_B$ are

negative at Lirung (-4.1 and -3.4, respectively) and Suldenferner (-0.55 and -0.46, respectively) but are not displayed for better visibility of the plot.

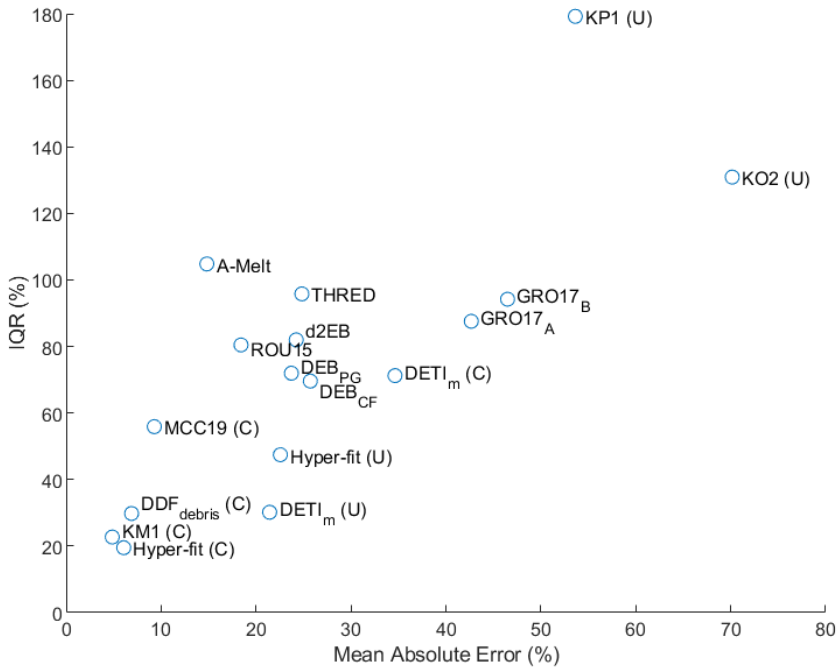


Figure S6. Alternative visualization for the model ranking of Table 4.

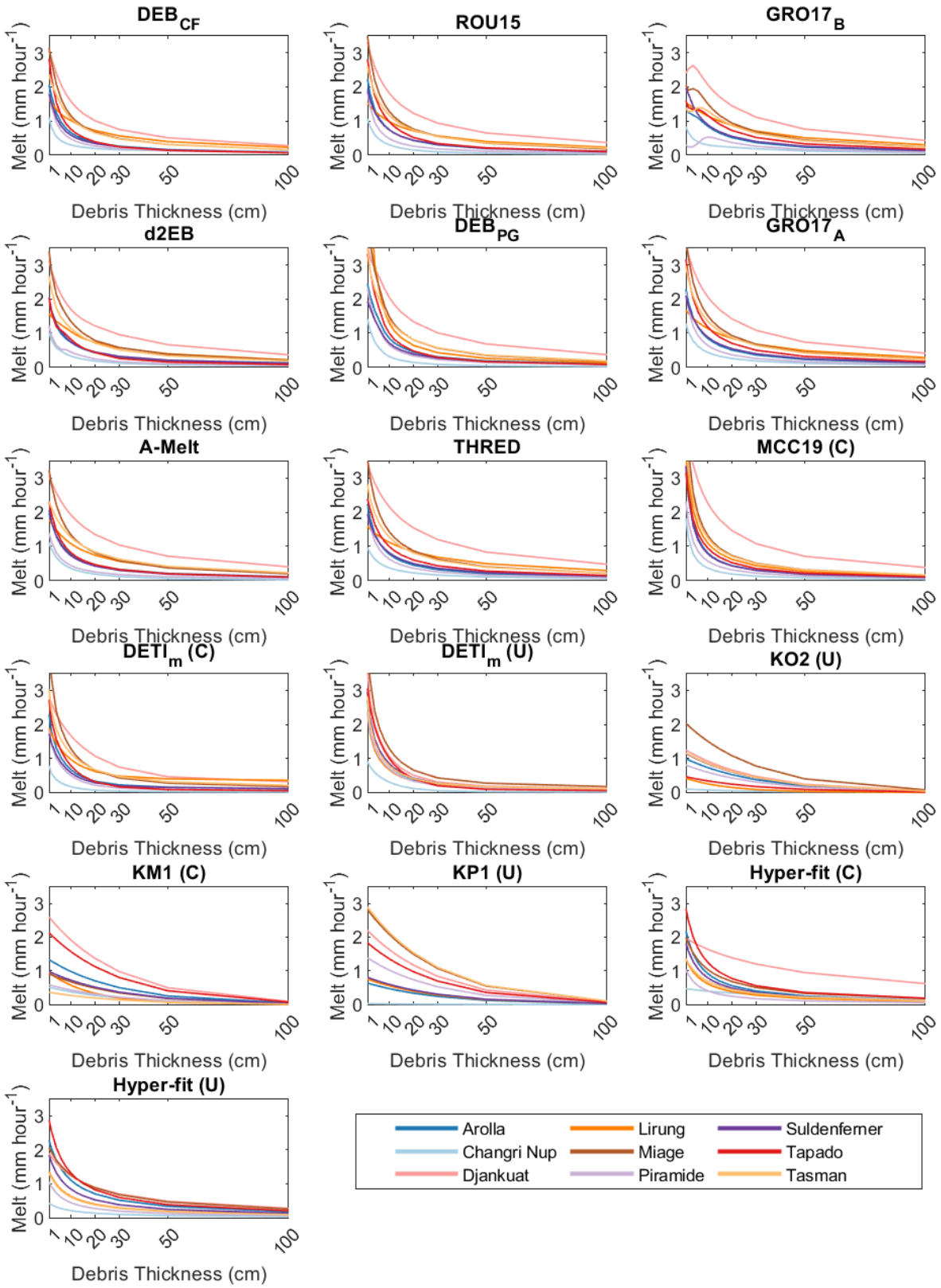


Figure S7. Østrem curves for each model at all sites.

4. Discussion

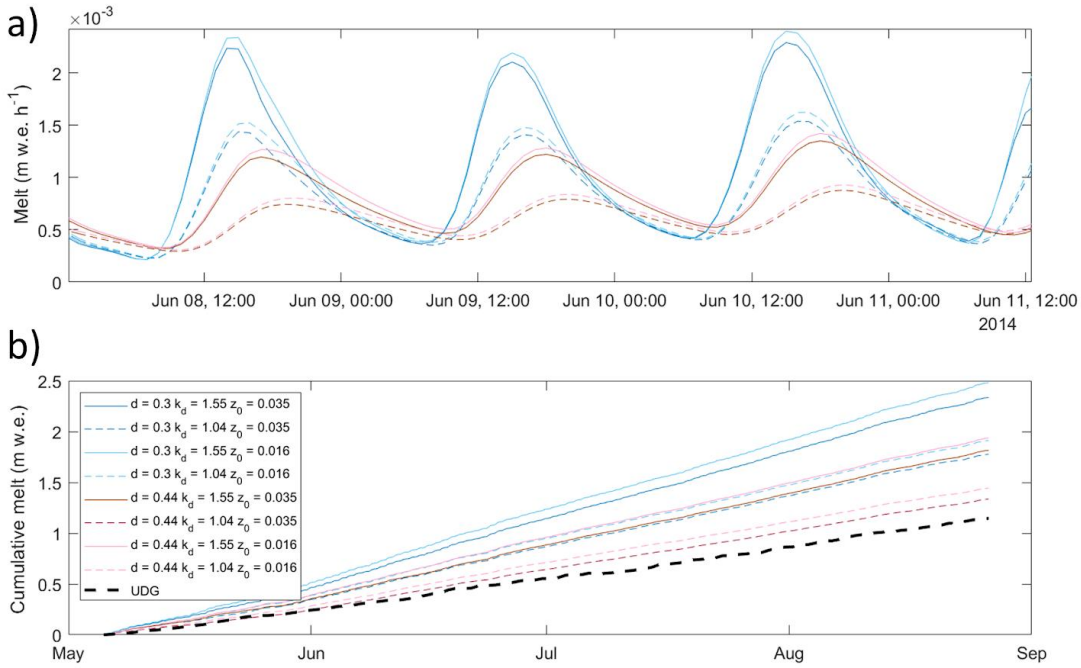


Figure S8. Melt rates simulated by the DEB_{CF} model at Lirung, with debris thickness (d) = 30 cm (thickness at the AWS location) and d = 44 cm (thickness at the UDG location), and with debris conductivity and surface roughness from Lirung (k = 1.55 and z_0 = 0.035) and Miage (k = 1.04 and z_0 = 0.016) as well as combinations of them. Panel a) shows hourly melt for three days of the simulation period, b) shows the cumulative melt of all scenarios compared to the UDG record.

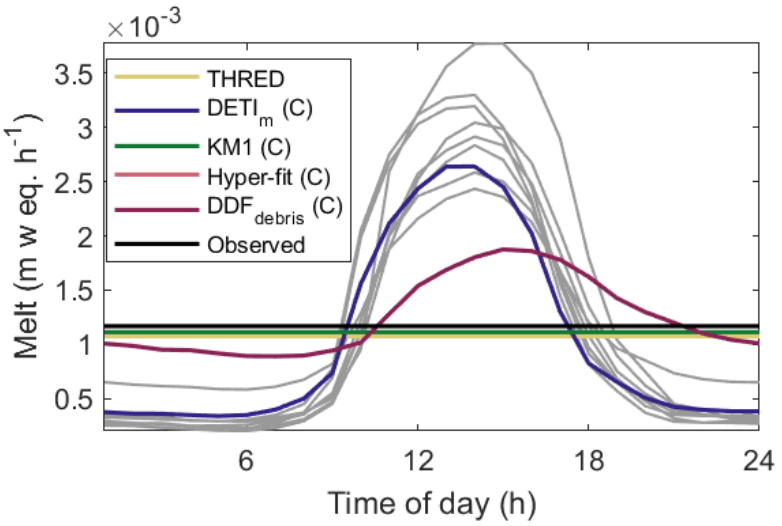


Figure S9. Illustrations of the temporal sub-daily cycle of melt of energy balance (and simplified energy balance) models (in grey) versus simple temperature index approaches (coloured), for one study site, Arolla (debris thickness = 6 cm). The cross (x) indicates the observed daily melt at the site. Daily melt simulated by the models operating at the daily time step are also indicated to show that daily values arise from overestimation of melt at night and underestimation of melt during the day (relative to the energy-balance models operating at hourly resolution. Hyper-fit and DDF_{debris} overlap each other.

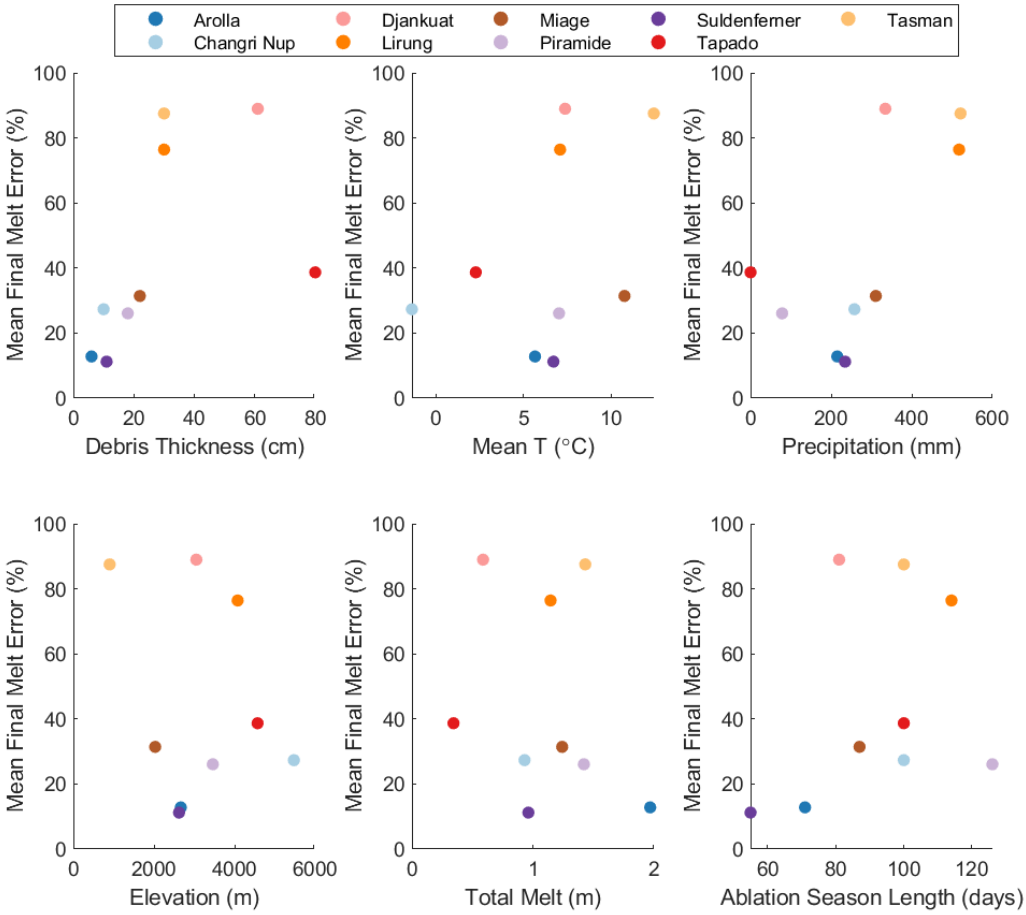


Figure S10. Mean melt error against site characteristics. For each of: debris thickness, mean air temperature, total precipitation, elevation, total melt, and simulation period length, the mean absolute error of the ensemble of models at each site is shown.

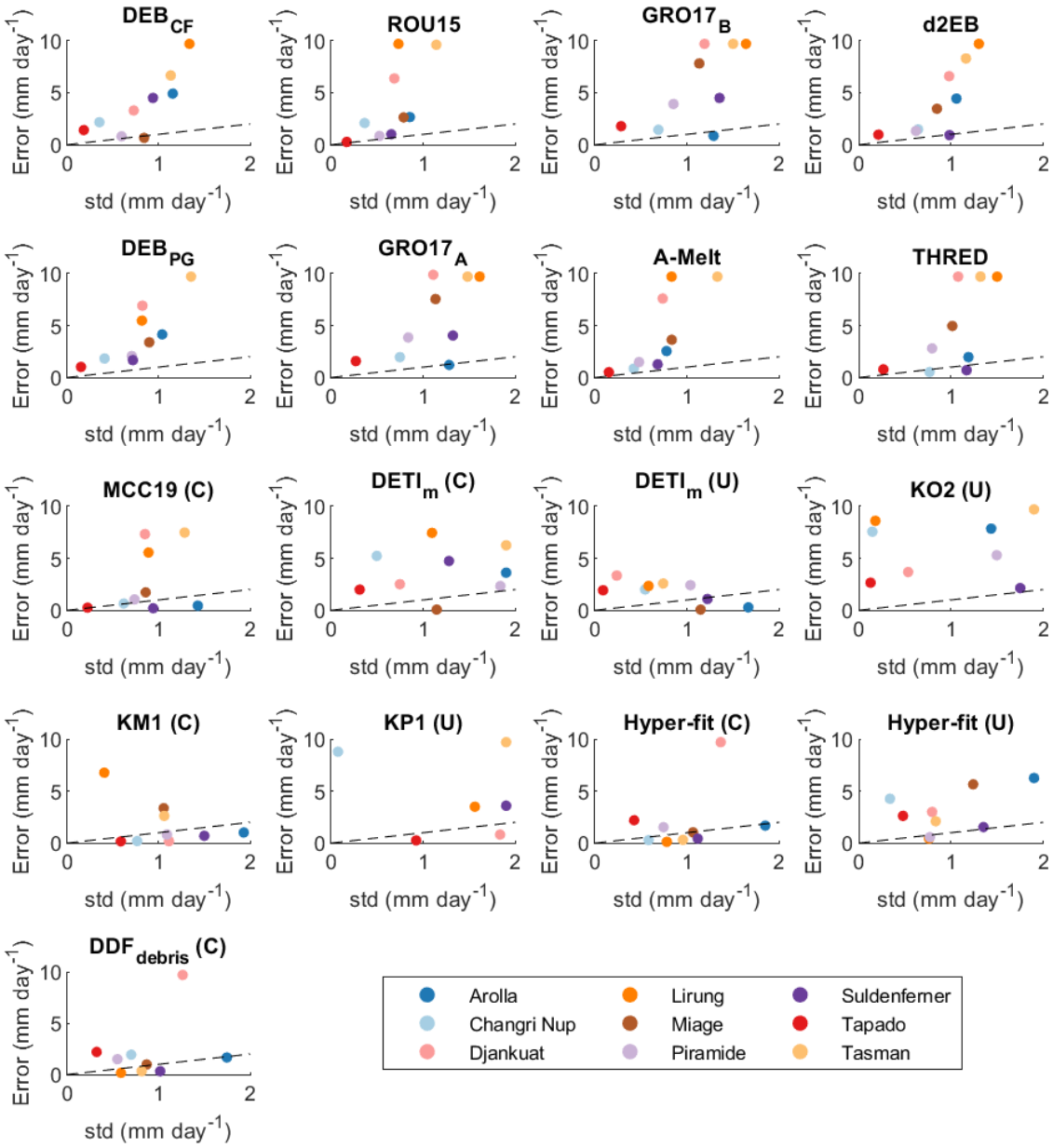


Figure S11. Modelled melt error (y-axis) vs model uncertainty (assumed equal to the standard deviation of the Monte Carlo Simulations, x-axis). Dashed line is the 1:1 line, so wherever a point is above the line, it means the error is larger than the Monte Carlo Uncertainty.

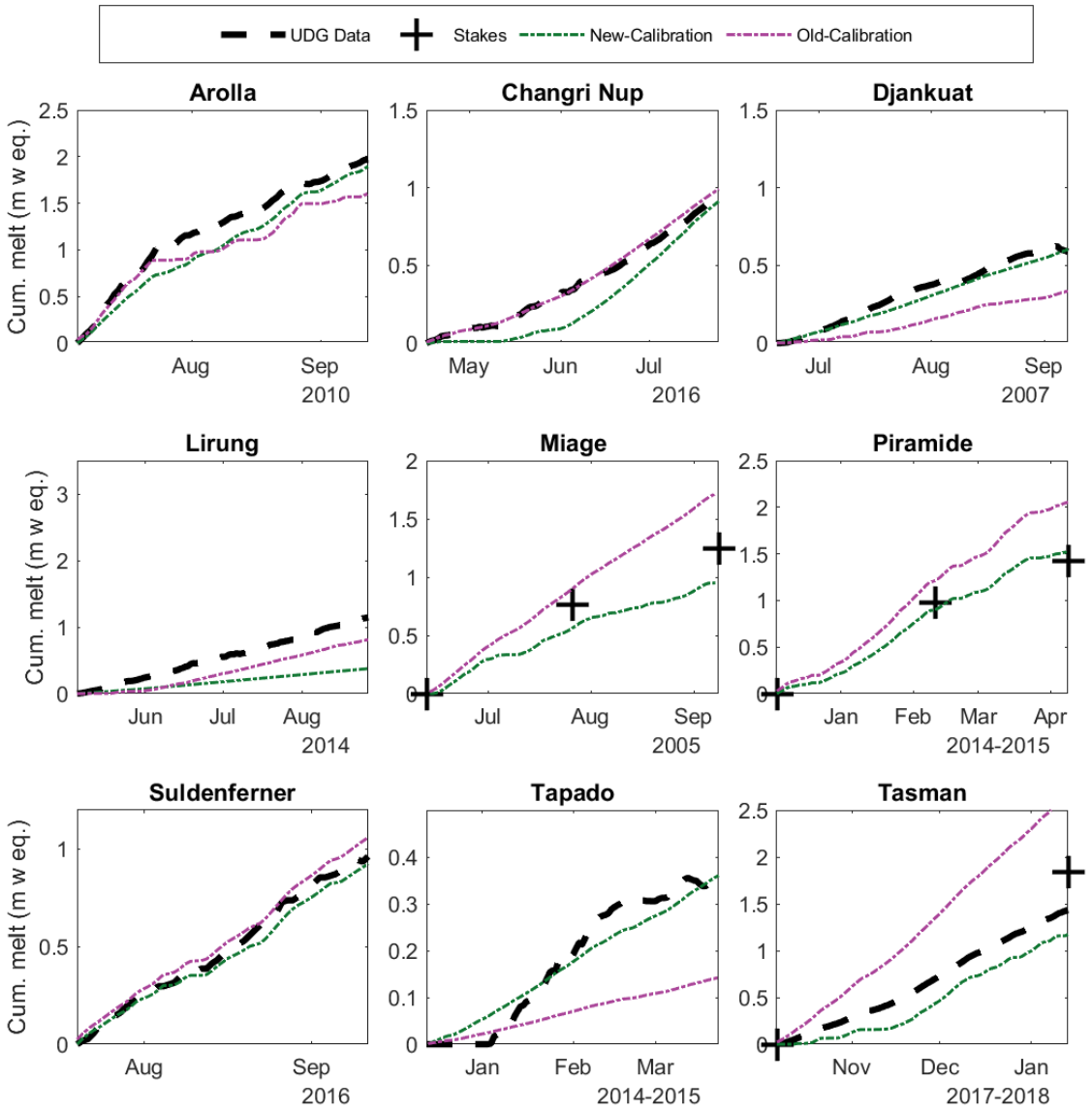


Figure S12. Melt simulations from KM1 model, new versus old calibration scheme: In the new calibration, the calibration strategy is set up to match the total final melt, and in the old calibration, to match mean daily melt rates.

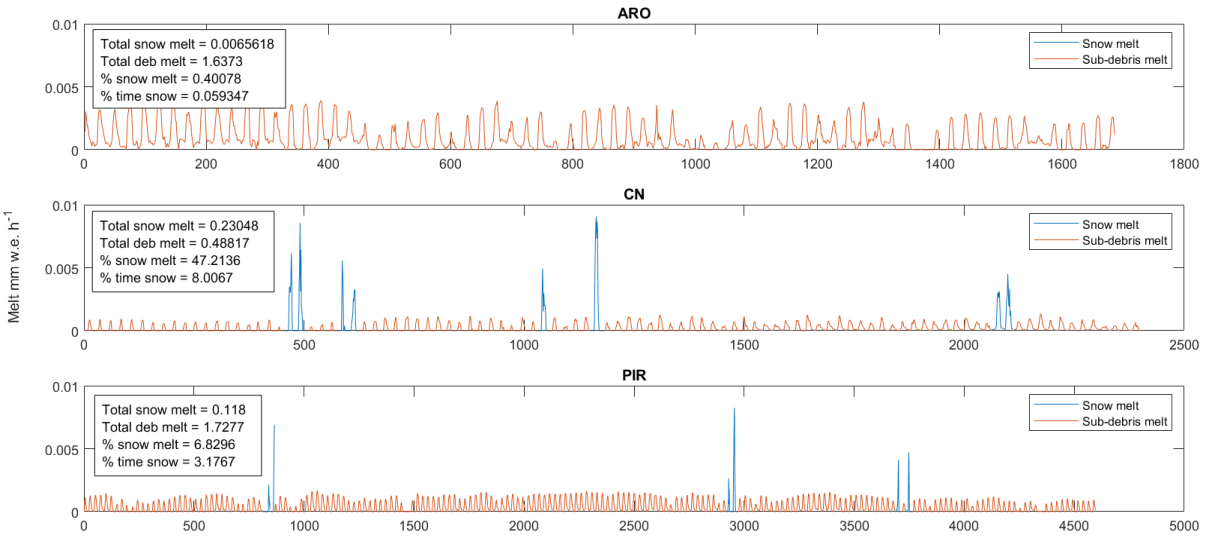


Figure S13. Snowmelt calculated by the DEB_{CF} model at the three sites where occasional snowfalls occurred during the period of record selected for this intercomparison.

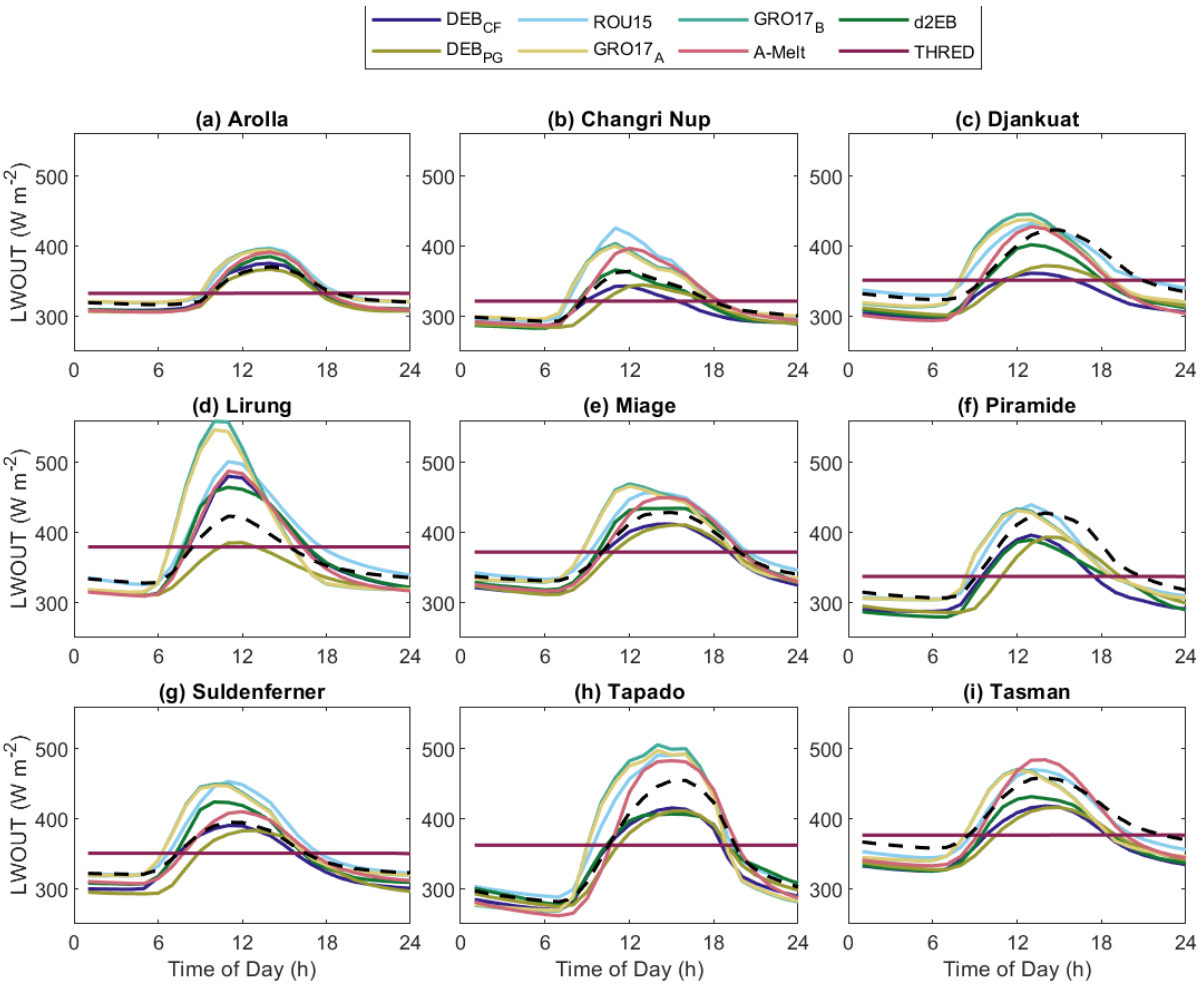


Figure S14. Outgoing longwave radiation flux (LW_{out}) computed by each model, from the internally simulated debris surface temperature. The dashed black line is the observed LW_{out} from the sensor. Notice that $GRO17_A$ and $GRO17_B$ overlap each other.

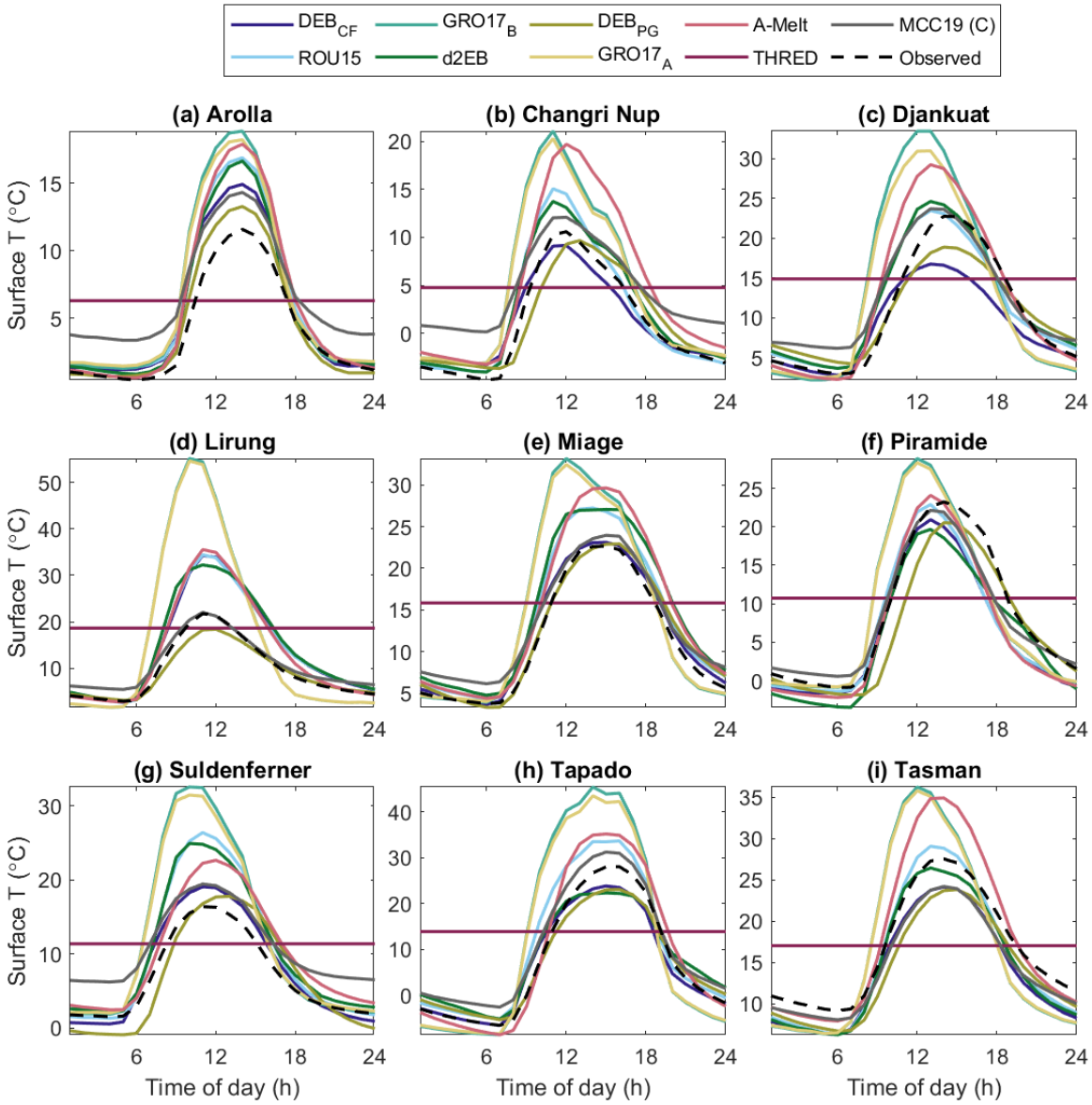


Figure S15. Mean daily cycle of debris surface temperature computed by each model at each site. All models simulate internally simulated surface temperature.

Table S28. Average daily energy fluxes simulated by the models at Arolla, Changri Nup, and Djankuat, during the period of simulation. L↑ is the outgoing longwave radiation flux. H is sensible heat flux. LE latent heat flux. G is ground heat flux. P is precipitation flux. All fluxes in $W\ m^{-2}$.

Glacier	Model	L↑	H	LE	G	P
Arolla	$GRO17_A$	342.48	-13.53	-7.98	103.31	-
	$GRO17_B$	342.39	-13.97	-	102.86	-
	DEB_{CF}	329.17	-35.02	-3.56	-	0.12
	ROU15	342.91	-12.75	-3.18	-96.85	0.13
	A-Melt	332.35	-5.63	-17.46	-	-
	d2EB	332.86	-19.17	-0.38	101.14	-
	THRED	330.23	-1.65	-33.20	-	-
	DEB_{PG}	324.81	-12.12	0.00	-114.10	-
	MCC19	334.11	-	-	-	-
Changri Nup	$GRO17_A$	331.59	-82.29	-29.85	40.86	-
	$GRO17_B$	330.68	-84.64	-	39.40	-
	DEB_{CF}	305.97	-91.21	-35.07	-	-0.11
	ROU15	334.67	-62.29	-28.10	-15.31	-0.19
	A-Melt	325.28	-49.10	-29.27	-	-
	d2EB	321.56	-88.19	0.00	34.08	-
	THRED	312.71	-33.97	-85.42	-	-
	DEB_{PG}	306.88	-113.82	0.00	-14.44	-
	MCC19	319.14	-	-	-	-
Djankuat	$GRO17_A$	362.48	-104.26	0.00	56.39	-
	$GRO17_B$	360.69	-108.02	-	54.28	-
	DEB_{CF}	324.08	-81.01	-74.99	-	-0.04
	ROU15	369.58	-94.94	-2.44	-53.49	-0.30
	A-Melt	343.52	-53.69	-65.23	-	-
	d2EB	351.50	-102.49	0.00	68.41	-
	THRED	339.22	-59.95	-59.94	-	-
	DEB_{PG}	331.50	-118.99	0.00	-57.89	-
	MCC19	362.24	-	-	-	-

Table S29. Average daily energy fluxes simulated by the models at Lirung, Miage and Piramide, during the period of simulation. L↑ is the outgoing longwave radiation flux. H is sensible heat flux. LE latent heat flux. G is ground heat flux. P is precipitation flux. All fluxes in $W\ m^{-2}$.

Glacier	Model	L↑	H	LE	G	P
Lirung	$GRO17_A$	383.01	-31.26	0.00	84.39	-
	$GRO17_B$	380.45	-36.85	-	81.09	-
	DEB_{CF}	367.93	-44.27	-3.68	-	-0.40
	ROU15	386.36	-31.13	-2.65	-73.28	-0.34
	Λ-Melt	366.31	-29.64	-23.04	-	-
	d2EB	379.53	-28.45	0.00	89.10	-
	THRED	369.12	-17.34	-6.35	-	-
	DEB_{PG}	338.42	-99.07	0.00	-53.69	-
	MCC19	360.95	-	-	-	-
Miage	$GRO17_A$	381.38	-68.05	0.00	73.32	-
	$GRO17_B$	380.65	-69.57	-	72.49	-
	DEB_{CF}	355.83	-92.21	-13.50	-	0.02
	ROU15	383.52	-59.81	-12.29	-65.73	-0.04
	Λ-Melt	369.58	-36.13	-45.66	-	-
	d2EB	372.19	-75.07	0.00	74.90	-
	THRED	369.71	-55.01	-28.73	-	-
	DEB_{PG}	354.41	-91.23	0.00	-65.60	-
	MCC19	369.54	-	-	-	-
Piramide	$GRO17_A$	347.59	-85.50	-0.30	47.63	-
	$GRO17_B$	347.24	-86.25	-	47.09	-
	DEB_{CF}	323.68	-83.05	-37.03	-	0.05
	ROU15	348.06	-51.87	-41.89	-34.21	0.05
	Λ-Melt	0.00	-20.75	-94.55	-	-
	d2EB	337.26	-89.65	-0.92	48.34	-
	THRED	323.09	-16.03	-156.86	-	-
	DEB_{PG}	325.80	-98.58	0.00	-41.36	-
	MCC19	352.75	-	-	-	-

Table S30. Average daily energy fluxes simulated by the models at Suldenferner, Tapado and Tasman, during the period of simulation. L↑ is the outgoing longwave radiation flux. H is sensible heat flux. LE latent heat flux. G is ground heat flux. P is precipitation flux. All fluxes in $W\ m^{-2}$.

Glacier	Model	L↑	H	LE	G	P
Suldenferner	$GRO17_A$	362.85	-30.61	0.00	72.11	-
	$GRO17_B$	362.45	-31.62	-	71.45	-
	DEB_{CF}	334.55	-73.23	-3.68	-	-0.08
	ROU15	365.24	-30.33	-3.00	-64.99	-0.06
	Λ-Melt	345.45	-26.40	-27.03	-	-
	d2EB	350.75	-38.43	0.06	72.33	-
	THRED	347.67	-30.18	-22.69	-	-
	DEB_{PG}	329.17	-64.80	0.00	-56.15	-
	MCC19	347.62	-	-	-	-
Tapado	$GRO17_A$	360.30	-165.68	0.00	17.93	-
	$GRO17_B$	358.27	-168.27	-	17.24	-
	DEB_{CF}	330.72	-200.47	0.00	-	0.00
	ROU15	367.57	-158.29	0.00	-13.71	0.00
	Λ-Melt	348.30	-75.77	-103.78	-	-
	d2EB	362.46	-164.41	0.00	16.49	-
	THRED	339.74	-81.81	-157.93	-	-
	DEB_{PG}	331.81	-190.80	0.00	-11.20	-
	MCC19	349.09	-	-	-	-
Tasman	$GRO17_A$	386.28	-65.56	0.00	97.70	-
	$GRO17_B$	386.06	-66.09	-	97.38	-
	DEB_{CF}	363.10	-87.10	-14.82	-	0.29
	ROU15	392.19	-50.48	-13.64	-91.83	0.36
	Λ-Melt	388.14	-44.98	-78.49	-	-
	d2EB	376.79	-71.07	-0.01	100.05	-
	THRED	368.63	-31.66	-62.07	-	-
	DEB_{PG}	364.19	-81.49	0.00	-95.02	-
	MCC19	397.90	-	-	-	-

References

Anderson, L. S. and Anderson, R. S.: Modeling debris-covered glaciers: Response to steady debris deposition, The Cryosphere, 10, 1105–1124, doi:10.5194/tc-10-1105-2016, 2016.

Ayala, A., Pellicciotti, F., MacDonell, S., McPhee, J., and Burlando, P.: Patterns of glacier ablation across N orth- C entral C hile: Identifying the limits of empirical melt models under sublimation-favorable conditions, Water Resour. Res., 53, 5601–5625, doi:10.1002/2016WR020126, 2017.

Ayala, A., Pellicciotti, F., MacDonell, S., McPhee, J., Vivero, S., Campos, C., and Egli, P.: Modelling the hydrological response of debris-free and debris-covered glaciers to present climatic conditions in the semiarid Andes of central Chile, Hydrol. Process., 30, 4036–4058, doi:10.1002/hyp.10971, 2016.

Bozhinskiy, A. N., Krass, M. S., and Popovnin, V. V.: Role of Debris Cover in the Thermal Physics of Glaciers, J. Glaciol., 32, 255–266, doi:10.3189/S0022143000015598, 1986.

Brock, B. W., Mihalcea, C., Kirkbride, M. P., Diolaiuti, G., Cutler, M. E. J., and Smiraglia, C.: Meteorology and surface energy fluxes in the 2005–2007 ablation seasons at the Miage debris-covered glacier, Mont Blanc Massif, Italian Alps, J. Geophys. Res. Atmos., 115, 2009JD013224, doi:10.1029/2009JD013224, 2010.

Carenzo, M., Pellicciotti, F., Mabillard, J., Reid, T., and Brock, B. W.: An enhanced temperature index model for debris-covered glaciers accounting for thickness effect, Adv. Water Resour., 94, 457–469, doi:10.1016/j.advwatres.2016.05.001, 2016.

Chand, M. B. and Kayastha, R. B.: Study of thermal properties of supraglacial debris and degree-day factors on Lirung Glacier, Nepal, Sci. Cold Arid Reg., 10, 357–368, doi:10.3724/SP.J.1226.2018.00357, 2018.

Elagina, N., Rets, E., Korneva, I., Toropov, P., and Lavrentiev, I.: Simulation of mass balance and glacial runoff of Mount Elbrus from 1984 to 2022, Hydrol. Sci. J., doi:10.1080/02626667.2025.2516080, 2025.

Evatt, G. W., Abrahams, I. D., Heil, M., Mayer, C., Kingslake, J., Mitchell, S. L., Fowler, A. C., and Clark, C. D.: Glacial melt under a porous debris layer, J. Glaciol., 61, 825–836, doi:10.3189/2015JoG14J235, 2015.

Fujita, K. and Sakai, A.: Modelling runoff from a Himalayan debris-covered glacier, Hydrol. Earth Syst. Sci., 18, 2679–2694, doi:10.5194/hess-18-2679-2014, 2014.

Fyffe, C. L., Reid, T. D., Brock, B. W., Kirkbride, M. P., Diolaiuti, G., Smiraglia, C., and Diotri, F.: A distributed energy-balance melt model of an alpine debris-covered glacier, J. Glaciol., 60, 587–602, doi:10.3189/2014JoG13J148, 2014.

Groos, A. R. and Mayer, C.: glacierSMBM: Glacier Surface Mass Balance Model, R package (v0.1), https://CRAN.R-project.org/package=glacierSMBM, 2017.

Groos, A. R., Mayer, C., Smiraglia, C., Diolaiuti, G., and Lambrecht, A.: A first attempt to model region-wide glacier surface mass balances in the Karakoram: Findings and future challenges, Geogr. Fis. Dinam. Quat., 40, 137–159, doi:10.4461/GFDQ.2017.40.10, 2017.

Hock, R.: Temperature index melt modelling in mountain areas, J. Hydrol., 282, 104–115, doi:10.1016/S0022-1694(03)00257-9, 2003.

Kayastha, R. B., Takeuchi, Y., Nakawo, M., and Ageta, Y.: Practical prediction of ice melting beneath various thickness of debris cover on Khumbu Glacier, Nepal, using a positive degree-day factor, 264, 71–81, 2000.

Kirkbride, M.: Ice Flow Vectors on the Debris-Mantled Tasman Glacier, 1957–1986, Geogr. Ann. A. Phys. Geogr., 77, 147–157, doi:10.1080/04353676.1995.11880435, 1995.

Kuzmin, P. P.: The Process of Snow Melting, Leningrad: Gidrometizdat, 1961.

Lambrecht, A., Mayer, C., Hagg, W., Popovnin, V., Rezepkin, A., Lomidze, N., and Svanadze, D.: A comparison of glacier melt on debris-covered glaciers in the northern and southern Caucasus, The Cryosphere, 5, 525–538, doi:10.5194/tc-5-525-2011, 2011.

Lejeune, Y., Bertrand, J.-M., Wagnon, P., and Morin, S.: A physically based model of the year-round surface energy and mass balance of debris-covered glaciers, J. Glaciol., 59, 327–344, doi:10.3189/2013JoG12J149, 2013.

McCarthy, M. J.: Quantifying supraglacial debris thickness at local to regional scales, https://doi.org/10.17863/CAM.41172, 2018.

McCarthy, M.: A simplified energy-balance model for ice melt below debris (Version 1.0.0), GitHub, https://github.com/mchl-mccrthy/seb_model, 2025.

Ayala, Á., McPhee, J., and MacDonell, S.: 8_Tapado Data [Dataset], https://doi.org/10.5281/ZENODO.3362402, 2019.

Shaw, T., McPhee, J., and MacDonell, S.: 6_Pirámide Data [Dataset], https://doi.org/10.5281/ZENODO.3056072, 2019.

Miles, E. S., Steiner, J. F., and Brun, F.: Highly variable aerodynamic roughness length (z 0) for a hummocky debris-covered glacier, J. Geophys. Res. Atmos., 122, 8447–8466, doi:10.1002/2017JD026510, 2017.

Nicholson, L. and Benn, D. I.: Calculating ice melt beneath a debris layer using meteorological data, J. Glaciol., 52, 463–470, doi:10.3189/172756506781828584, 2006.

Oerlemans, J.: Glaciers and Climate Change, A.A. Balkema Publishers, 2001.

Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M., and Corripio, J.: An enhanced temperature-index glacier melt model including the shortwave radiation balance: Development and testing for Haut Glacier d'Arolla, Switzerland, J. Glaciol., 51, 573–587, doi:10.3189/172756505781829124, 2005.

Reid, T. D. and Brock, B. W.: An energy-balance model for debris-covered glaciers including heat conduction through the debris layer, J. Glaciol., 56, 903–916, doi:10.3189/002214310794457218, 2010.

Reid, T. D., Carenzo, M., Pellicciotti, F., and Brock, B. W.: Including debris cover effects in a distributed model of glacier ablation, J. Geophys. Res. Atmos., 117, 2012JD017795, doi:10.1029/2012JD017795, 2012.

Rets, E. and Kireeva, M.: Hazardous hydrological processes in mountainous areas under the impact of recent climate change: Case study of Terek River basin, 340, 126–134, 2010.

Robertson, E. C.: General and Engineering Geology of the Northern Part of Pueblo, Colorado (Open-File Report 88–441), U.S. Geological Survey, https://pubs.usgs.gov/of/1988/0441/report.pdf, 1988.

Röhl, K.: Characteristics and evolution of supraglacial ponds on debris-covered Tasman Glacier, New Zealand, J. Glaciol., 54, 867–880, doi:10.3189/002214308787779861, 2008.

Rounce, D. R., Quincey, D. J., and McKinney, D. C.: Debris-covered glacier energy balance model for Imja–Lhotse Shar Glacier in the Everest region of Nepal, The Cryosphere, 9, 2295–2310, doi:10.5194/tc-9-2295-2015, 2015.

Rye, C. J., Arnold, N. S., Willis, I. C., and Kohler, J.: Modeling the surface mass balance of a high Arctic glacier using the ERA-40 reanalysis, J. Geophys. Res. Earth Surf., 115, 2009JF001364, doi:10.1029/2009JF001364, 2010.

Steiner, J. F., Kraaijenbrink, P. D. A., and Immerzeel, W. W.: Distributed Melt on a Debris-Covered Glacier: Field Observations and Melt Modeling on the Lirung Glacier in the Himalaya, Front. Earth Sci., 9, 678375, doi:10.3389/feart.2021.678375, 2021.

Steiner, J. F., Litt, M., Stigter, E. E., Shea, J., Bierkens, M. F. P., and Immerzeel, W. W.: The Importance of Turbulent Fluxes in the Surface Energy Balance of a Debris-Covered Glacier in the Himalayas, Front. Earth Sci., 6, 144, doi:10.3389/feart.2018.00144, 2018.

Sturm, M., Holmgren, J., König, M., and Morris, K.: The thermal conductivity of seasonal snow, J. Glaciol., 43, 26–41, doi:10.3189/S0022143000002781, 1997.

Wagnon, P.: 2_Changri Nup Data [Dataset], https://doi.org/10.5281/ZENODO.3048780, 2019.

Winter-Billington, A., Moore, R. D., and Dadic, R.: Evaluating the transferability of empirical models of debris-covered glacier melt, J. Glaciol., 66, 978–995, doi:10.1017/jog.2020.57, 2020.