



How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment

Daniel Farinotti^{1,2}, Douglas Brinkerhoff³, Garry K.C. Clarke⁴, Johannes J. Fürst⁵, Holger Frey⁶, Prateek Gantayat⁷, Fabien Gillet-Chaulet⁸, Claire Girard⁹, Matthias Huss^{1,10}, Paul W. Leclercq¹¹, Andreas Linsbauer^{6,10}, Horst Machguth^{6,10}, Carlos Martin¹², Fabien Maussion¹³, Mathieu Morlighem⁹, Cyrille Mosbeux⁸, Ankur Pandit¹⁴, Andrea Portmann², Antoine Rabatel⁸, RAAJ Ramsankaran¹⁴, Thomas J. Reerink¹⁵, Olivier Sanchez⁸, Peter A. Stenof¹⁶, Sangita Singh Kumari¹⁴, Ward J.J. van Pelt¹⁷, Brian Anderson¹⁸, Toby Benham¹⁹, Daniel Binder²⁰, Julian A. Dowdeswell¹⁹, Andrea Fischer²¹, Kay Helfricht²¹, Stanislav Kutuzov²², Ivan Lavrentiev²², Robert McNabb^{3,11}, G. Hilmar Gudmundsson¹², Huilin Li²³, and Liss M. Andreassen²⁴

¹Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland

²Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

³Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA

⁴Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver BC, Canada

⁵Institute of Geography, Friedrich-Alexander-University Erlangen-Nuremberg (FAU), Erlangen, Germany

⁶Department of Geography, University of Zurich, Zurich, Switzerland

⁷Divecha Centre for Climate Change, Indian Institute of Science, Bangalore, India

⁸Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), Université Grenoble Alpes, CNRS, Grenoble, France

⁹Department of Earth System Science, University of California Irvine, Irvine, CA, USA

¹⁰Department of Geosciences, University of Fribourg, Fribourg, Switzerland

¹¹Department of Geosciences, University of Oslo, Oslo, Norway

¹²British Antarctic Survey, Natural Environment Research Council, Cambridge, UK

¹³Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria

¹⁴Department of Civil Engineering, Indian Institute of Technology, Bombay, India

¹⁵Institute for Marine and Atmospheric Research (IMAU), Utrecht University, Utrecht, The Netherlands

¹⁶Arctic Technology Centre ARTEK, Technical University of Denmark, Kongens Lyngby, Denmark

¹⁷Department of Earth Sciences, Uppsala University, Uppsala, Sweden

¹⁸Antarctic Research Centre, Victoria University of Wellington, Wellington, New Zealand

¹⁹Scott Polar Research Institute, University of Cambridge, Cambridge, UK

²⁰Central Institute for Meteorology and Geodynamics (ZAMG), Vienna, Austria

²¹Institute for Mountain Research, Austrian Academy of Sciences, Vienna, Austria

²²Laboratory of Glaciology, Institute of Geography, Russian Academy of Science, Moscow, Russia

²³State Key Laboratory of Cryospheric Sciences, Tian Shan Glaciological Station, CAREERI, CAS, Lanzhou, China

²⁴Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway

Correspondence to: Daniel Farinotti (daniel.farinotti@ethz.ch)



Abstract. Knowledge of the ice thickness distribution of glaciers and ice caps is an important prerequisite for many glaciological and hydrological investigations. A wealth of approaches has recently been presented for inferring ice thickness from characteristics of the surface. With the Ice Thickness Models Intercomparison eXperiment (ITMIX) we performed the first coordinated assessment quantifying individual model performance. A set of 17 different models showed that individual ice thickness estimates can differ considerably – locally by a spread comparable to the observed thickness. Averaging the results of multiple models, however, significantly improved the results: On average over the 21 considered test cases, comparison against direct ice thickness measurements revealed deviations in the order of $10 \pm 24\%$ of the mean ice thickness (1σ estimate). For models relying on multiple data sets – such as surface ice velocity fields, surface mass balance, or rates of ice thickness change – the results highlighted the sensitivity to input data consistency. Together with the requirement of being able to handle large regions in an automated fashion, the capacity of better accounting for uncertainties in the input data will be a key for an improved next generation of ice thickness estimation approaches.

1 Introduction

The ice thickness distribution of a glacier, ice cap, or ice sheet is a fundamental parameter for many glaciological applications. It determines the total volume of the ice body, which is crucial to quantify water availability or sea-level change, and provides the link between surface and subglacial topography, which is a prerequisite for ice-flow modelling studies. Despite this importance, knowledge about the ice thickness of glaciers and ice caps around the globe is limited – a fact linked mainly to the difficulties in measuring ice thickness directly. To overcome this problem, a number of methods has been developed to infer the total volume and/or the ice thickness distribution of ice masses from characteristics of the surface.

Amongst the simplest methods, so-called *scaling approaches* are the most popular (see Bahr et al., 2015, for a recent review). These approaches explore relationships between the area and the volume of a glacier (e.g. Chen and Ohmura, 1990; Bahr et al., 1997), partially including other characteristics such as glacier length or surface slope (e.g. Lüthi, 2009; Radić and Hock, 2011; Grinsted, 2013). Such approaches, however, yield estimates of the mean ice thickness and total volume of a glacier only.

Methods that yield distributed information about the ice thickness generally rely on theoretical considerations. Nye (1952), for example, noted that for the case of an idealized glacier of infinite width, ice thickness can be calculated from the surface slope using estimates of basal shear stress and assuming perfect plastic behaviour. Nye (1965) successively extended the considerations to valley glaciers of idealized shapes, whilst Li et al. (2012) additionally accounted for the effect of side drag from the glacier margins. Common to these three approaches is the assumption of a constant and known basal shear stress. Haeberli and Hoelzle (1995) were the first suggesting that the latter can be estimated from the glacier elevation range, and the corresponding parameterization has been used in a series of recent studies (e.g. Paul and Linsbauer, 2011; Linsbauer et al., 2012; Frey et al., 2014).

Early approaches that take into account mass conservation and ice flow dynamics go back to Budd and Allison (1975) and Rasmussen (1988), whose ideas were further developed by Fastook et al. (1995) and Farinotti et al. (2009b). The latter approach was successively extended by Huss and Farinotti (2012), who presented the first globally complete estimate for the



ice thickness distribution of individual glaciers. Alternative methods based on more rigorous inverse modelling of glacier ice
35 flow, on the other hand, have often focused on inferring additional properties at the glacier base, besides ice thickness (e.g.
Gudmundsson et al., 2001; Thorsteinsson et al., 2003; Raymond-Pralong and Gudmundsson, 2011; Mosbeux et al., 2016).

In the recent past, the number of methods aiming at estimating the ice thickness distribution from characteristics of the
surface has increased at a rapid pace: Methods have been presented that include additional data such as surface velocities and
mass balance (e.g. Morlighem et al., 2011; McNabb et al., 2012; Clarke et al., 2013; Farinotti et al., 2013; Huss and Farinotti,
40 2014; Gantayat et al., 2014; Brinkerhoff et al., 2016), as well as approaches that make iterative use of more complex forward
models of ice flow (e.g. van Pelt et al., 2013; Michel et al., 2013, 2014), or non-physical methods based on neural network
approaches (Clarke et al., 2009; Haq et al., 2014). This development has led to a situation in which a wealth of approaches is
potentially available, but no assessment comparing the relative strengths and weaknesses of the models exists.

Against this background, the Working Group on glacier ice thickness estimation, hosted by the International Association of
45 Cryospheric Sciences (www.cryosphericciences.org), launched the *Ice Thickness Models Intercomparison eXperiment*
(ITMIX). The experiment aimed at conducting a coordinated comparison between models capable of estimating the ice
thickness distribution of glaciers and ice caps from surface characteristics. Emphasis was put on evaluating the model perfor-
mance when no a-priori information on actual ice thickness is provided. This was to focus on the most widespread application
of such models; that is, the estimation of the ice thickness of an unmeasured glacier.

50 This article presents both the experimental setup of ITMIX and the results of the intercomparison. The accuracy of individual
approaches is assessed in a unified manner, and the strengths and shortcomings of individual models are highlighted. By doing
so, ITMIX not only provides quantitative constraints on the accuracies that can be expected from individual models, but also
aims at setting the basis for developing a new generation of improved ice thickness estimation approaches.

2 Experimental setup

55 ITMIX was conducted as an open experiment, with a call for participation posted on the email distribution list “Cryolist”
(<http://cryolist.org/>) on 13 October 2015. Individual researchers known to have developed a method for estimating
glacier ice thickness were invited personally. Upon registration, participants were granted access to the input data necessary
for the experiment and the corresponding set of instructions.

The input data referred to the surface characteristics of a predefined set of 21 test cases (see next section, Tab. 1, and Fig. 1)
60 and participants were asked to use these data for generating an estimate of the corresponding ice thickness distribution. Results
were collected, and compared to direct ice thickness measurements.

No prior information about ice thickness was provided, and the participants were asked not to make use of published ice
thickness measurements referring to the considered test cases for model calibration. This was to mimic the general case in
which the ice thickness distribution for unmeasured glaciers has to be estimated. The compliance to the above rule relied on
65 honesty.



Participants were asked to treat as many test cases as possible, and to consider data-availability (cf. next section and Tab. 1) as the only factor limiting the number of addressed cases. Details on the considered test cases and the participating models are given in Sections 3 and 4 respectively. An overview of the solutions submitted to the experiment is given in Table 2.

3 Considered test cases and data

- 70 The considered test cases included 15 glaciers and 3 ice caps for which direct ice thickness measurements are available, and 3 synthetically generated glaciers virtually “grown” over known bedrock topographies (more detailed information below). The real-world test cases (see Fig. 1 for geographical distribution) were chosen to reflect different glacier morphologies (cf. Tab. 1) and different climatic regions, whilst the synthetic test cases were included to have a set of experiments for which all necessary information is perfectly known.
- 75 For each test case, the input data provided to the ITMIX participants included at a minimum (a) an outline of the glacier or ice cap, and (b) a gridded digital elevation model (DEM) of the ice surface. Further information was provided on a case-by-case basis depending on data availability, including the spatial distribution of the (i) surface mass balance (SMB), (ii) rate of ice thickness change ($\partial h/\partial t$), and (iii) surface flow velocity. An overview of the data available for individual test cases and the corresponding data sources is given in Table 2 and Table 1, respectively.
- 80 For the real-world test cases, and whenever possible, temporal consistency was ensured between individual data sets. Glacier outlines and DEMs were snapshots for a given point in time, whereas SMB, $\partial h/\partial t$, and velocity fields generally referred to multi-year averages for an epoch as close as possible to the corresponding DEM. Consistent glacier-wide estimates of surface velocities were not available for any of the considered cases. For obtaining a possibly complete coverage, velocities from separate tiles were therefore merged, which often led to discontinuities along the tile margins.
- 85 Ice thickness measurements were only used for quantifying model performance but were not distributed to the ITMIX participants. Bedrock elevations were obtained by subtracting observed ice thicknesses from surface elevations, and the bedrock was assumed to remain unchanged over time. The time periods the individual data sets are referring to is given in Supplementary Table S1. Note that no specific information about the uncertainties associated to individual measurements were available. Reported uncertainties for ice thickness measurements, however, are typically below 5% (Plewes and Hubbard, 2001).
- 90 The synthetic test cases were generated by “growing” ice masses over known bedrock topographies with the *Elmer/Ice* ice flow model (Gagliardini et al., 2013). To do so, selected deglaciated areas located in the European Alps were extracted from local high-resolution DEMs (product *DHM25* by the Swiss Federal Office of Topography), and the flow model forced with a prescribed SMB field. The SMB field was either generated by prescribing an equilibrium-line altitude and two separate SMB elevation gradients for the accumulation and ablation zone (test cases “Synthetic1” and “Synthetic2”), or by constructing
- 95 the field through a multiple linear regression between SMB and terrain elevation, slope, aspect, curvature, and local position (test case “Synthetic3”). In the latter case, the individual regression parameters were defined arbitrarily but such to ensure a plausible range for the resulting SMB field. The *Elmer/Ice* simulations were stopped after the formation of a glacier judged to be of suitable size and shape, and the corresponding $\partial h/\partial t$ and surface velocity fields were extracted. No sliding at the glacier



Table 1. Overview of the test cases considered in ITMIX. Glacier type follows the *GLIMS classification guidance* (Rau et al., 2005). Abbreviations: A = glacier area; SB = simple basin; CB = compound basin; mnt. = mountain; OL = glacier outline, DEM = digital elevation model of the glacier surface, SMB = surface mass balance, Vel. = surface ice flow velocity, $\partial h/\partial t$ = rate of ice thickness change, H = ice thickness measurements, Unpub. = Unpublished data by. References to the data are given.

Test case	Type	A (km ²)	Available data and source
Academy	Ice cap	5587.2	OL, DEM, H: Dowdeswell et al. (2002)
Aqutitsoq	SB valley gl.	2.8	OL, DEM, H: Marcer et al. (in review)
Austfonna	Ice cap	7804.8	OL, DEM: Moholdt and Kääb (2012); $\partial h/\partial t$, SMB: Unpub. G. Moholdt; Vel.: Dowdeswell et al. (2008); H: Dowdeswell et al. (1986)
Brewster	SB mountain gl.	2.5	OL: LINZ (2013); DEM: Columbus et al. (2011); SMB: Anderson et al. (2010); Vel.: Unpub. B. Anderson; H: Willis et al. (2009)
Columbia	CB valley gl.	937.1	OL, DEM, H: McNabb et al. (2012)
Devon	Ice cap	14015.0	OL, DEM, H: Dowdeswell et al. (2004); Vel.: Unpub. GAMMA ⁽¹⁾
Elbrus	Crater mnt. gl.	120.8	OL, $\partial h/\partial t$, H: Unpub. RAS ⁽²⁾ ; DEM: Zolotarev and Khrkovets (2000); SMB: WGMS (1991-2012)
Freya	SB valley gl.	5.3	OL, DEM, H: Unpub. ZAMG ⁽³⁾ ; SMB: Hynek et al. (2015)
Hellstugubreen	CB valley gl.	2.8	OL: Andreassen et al. (2008); DEM, SMB, $\partial h/\partial t$: Andreassen et al. (2016); Vel.: Unpub. NVE ⁽⁴⁾ ; H: Andreassen et al. (2015)
Kesselwandferner	SB mountain gl.	4.1	OL, DEM: Fischer et al. (2015); SMB: Fischer et al. (2014); H: Fischer and Kuhn (2013)
Mocho	Crater mnt. gl.	15.2	OL, H: Geostudios LTA (2014); DEM: ASTER GDEM v2; SMB: Unpub. M. Schaefer
NorthGlacier	SB valley gl.	7.0	OL, DEM, H: Wilson et al. (2013); Vel.: Unpub. G. Flowers
SouthGlacier	SB valley gl.	5.3	OL, DEM, H: Wilson et al. (2013); SMB: Wheler et al. (2014) Vel.: Flowers et al. (2011)
Starbuck	CB outlet gl.	259.7	OL, H: Farinotti et al. (2014); DEM: Cook et al. (2012)
Tasman	CB valley gl.	100.3	OL: LINZ (2013); DEM: Columbus et al. (2011); H: Anderton (1975)
Unteraar	CB valley gl.	22.7	OL, DEM, SMB: Unpub. VAW-ETHZ ⁽⁵⁾ ; Vel.: Vogel et al. (2012); H: Bauder et al. (2003)
Urumqi	SB mountain gl.	1.6	OL, DEM, H: Wang et al. (2016)
Washmawapta	Cirque mnt. gl.	0.9	OL, DEM, H: Sanders et al. (2010)
Synthetic1	CB valley gl.	10.3	OL, DEM, SMB, Vel., $\partial h/\partial t$, H: Unpub. C. Martin and D. Farinotti
Synthetic2	CB mountain gl.	35.3	OL, DEM, SMB, Vel., $\partial h/\partial t$, H: Unpub. C. Martin and D. Farinotti
Synthetic3	Ice cap	89.9	OL, DEM, SMB, Vel., $\partial h/\partial t$, H: Unpub. C. Martin and D. Farinotti

⁽¹⁾ GAMMA Remote Sensing Research and Consulting AG, Gümligen, Switzerland; contact person T. Strozzi

⁽²⁾ Russian Academy of Sciences, Institute of Geography, Moscow, Russia; contact person S. Kutuzov

⁽³⁾ Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria; contact person D. Binder

⁽⁴⁾ Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway; contact person L.M. Andreassen

⁽⁵⁾ Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland; contact person A. Bauder



Table 2. Overview of provided and used data, as well as test cases considered by individual models. Names of ice caps are flagged with an asterisk (*). Models are named after the modeller submitting the results; alternative model identifiers that have been used in the literature are given in parenthesis. Abbreviations: OL+DEM= Glacier outline and digital elevation model of the surface; SMB = surface mass balance; Vel. = surface ice flow velocity; $\partial h/\partial t$ = rate of ice thickness change. For “Vel.”, a distributed field of flow speeds (s) and flow directions (d), or individual point measurements (p) were provided. “x” (“.”) indicates that the given information was (not) provided/used. In the columns “Data used”, “(x)” indicates that the information was used when available, but that it is not strictly necessary for model application. References for the data source are given in Table 1.

Provided data	Academy*	Aqutikitsiq	Austfonna*	Brewster	Columbia	Devon*	Elbrus	Freya	Hellsstugubreen	Kesselwandferner	Mochu	NorthGlacier	SouthGlacier	Starbuck	Tasman	Unteraar	Urumqi	Washmawapta	Synthetic1	Synthetic2	Synthetic3	TOTAL cases	Data used				
	OL+DEM	SMB	Vel.	$\partial h/\partial t$	OL+DEM	SMB	Vel.	$\partial h/\partial t$																			
Brinkerhoff	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	3	x	x	x	x	
Brinkerhoff-v2	.	.	.	x	.	.	.	x	x	x	.	.	x	.	.	x	x	.	x	x	x	10	x	x	(x)	.	
Farinotti (ITEM)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	21	x	.	.	.	
Fuerst	.	.	x	x	.	.	.	x	x	x	5	x	x	x	x	
Gantayat	.	.	x	.	.	x	x	x	.	.	x	x	x	7	x	.	x	.	
Gantayat-v2	.	.	x	.	.	x	x	x	.	.	x	x	x	7	x	.	x	.	
GCbedstress	x	x	.	x	.	.	.	x	x	x	x	x	.	.	x	x	x	x	x	x	x	15	x	(x)	.	(x)	
GCneuralnet	.	x	x	x	x	.	.	x	x	.	.	x	x	x	10	x	.	.	.	
Huss (HF-model)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	21	x	.	.	.	
Linsbauer (GlabTop)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	21	x	.	.	.	
Machguth (GlabTop2)	x	x	.	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	20	x	.	.	.	
Maussion (OGGM)	x	x	x	x	x	x	x	x	x	x	x	x	.	.	x	x	x	x	x	x	.	19	x	.	.	.	
Morlighem	x	x	x	.	.	x	.	x	x	x	.	x	x	x	10	x	x	(x)	.	
RAAJgantayat	.	.	.	x	x	.	.	x	x	.	.	x	.	.	5	x	.	x	.	
RAAJglabtop2	x	.	.	x	x	.	.	.	x	.	4	x	.	.	.	
Rabatel	x	.	.	1	x	x	x	.
VanPeltLeclercq	.	.	.	x	.	.	x	x	x	x	x	x	.	.	x	x	x	10	x	x	(x)	.	
TOTAL models	6	7	7	9	5	7	6	9	9	10	8	9	9	4	11	15	8	6	16	15	13	189					



Figure 1. Overview of the considered real-world test cases. Note that some names are shortened for convenience (Academy = Academy of Sciences Ice Cap; Devon = Devon Ice Cap; Mocho = Glaciér Mocho-Choshuenco; Unteraar = Unteraargletscher; Urumqi = Urumqi Glacier No. 1).

base was assumed, and all three resulting geometries were close to steady state. Note that, to avoid numerical instabilities, the DEM used for prescribing the bedrock topography had to be smoothed significantly. For anonymizing the individual locations, the original coordinates were removed, and the individual tiles arbitrarily rotated and shifted in elevation.

All data provided as input to the ITMIX participants, as well as the results submitted by individual models, will be provided as an electronic supplement to this article. The direct ice thickness measurements were additionally included in the Glacier Thickness Database (GlaThiDa) version 2 (WGMS, 2016).

4 Participating models

The ITMIX call for participation was answered by 13 research groups providing results from 15 different models in total (Tab. 2). Two modelling approaches (the ones by Gantayat et al. (2014) and Frey et al. (2014)) were used twice, with independent results stemming from two different groups. In general, four model categories can be distinguished, i.e. models that: (1) are based on mass-conservation only (“Brinkerhoff”, “Farinotti”, “Fuerst”, “GCbedstress”, “Huss”, “Maussion”, “Morlighem”, “Rabatel”), (2) additionally solve the momentum equation (“Brinkerhoff-v2”, “VanPeltLeclercq”), (3) directly apply the shallow ice approximation (“Linsbauer”, “Machguth”, “Gantayat”, “Gantayat-v2”, “RAAJgantayat”, “RAAJglabtop2”), and (4) are based on artificial neural networks (“GCneuralnet”). A brief description of individual models is provided below.

4.1 “Brinkerhoff” – Brinkerhoff et al. (2016)

The method of Brinkerhoff et al. (2016) poses the problem of finding bedrock elevations in the context of Bayesian inference. The model uses the prior hypotheses that bed elevations and the ice flux divergence ($SMB - \partial h / \partial t$) can be modelled as Gaussian random fields with assumed covariance but unknown mean. The choice of covariance function enforces strong prior information about smoothness. Depth averaged velocities are found by solving the continuity equation. The likelihood assumes that velocities and the ice flux divergence are normally distributed with known covariance around the supplied data. With prior and



likelihood in hand, the model uses the Metropolis-Hastings algorithm (Hastings, 1970) to generate samples from the posterior
120 distribution of bed elevations.

For ITMIX, only the maximum likelihood solution is reported (not the full posterior distribution). Because of the Gaussian distribution of observations and priors, the choice of observation and prior covariance does not affect the maximum likelihood solution. The model was applied only to the synthetic cases, as velocity fields provided for the real cases had either insufficient spatial coverage or non-physical behaviour incompatible with the assumed model physics.

125 4.2 “Brinkerhoff-v2” – Brinkerhoff (unpublished)

The unpublished model “Brinkerhoff-v2” solves the inverse problem *find a bedrock topography $B(x, y)$ that minimizes*

$$I = \int_{\Omega} \frac{(\bar{S} - \bar{S}_{\text{obs}})^2}{2} d\Omega, \quad (1)$$

where \bar{S}_{obs} is a set of surface elevations that are Gaussian smoothed over approximately half the glacier width, and \bar{S} is $S = B + H$ smoothed similarly. The ice thickness H is found by simultaneously solving a mixed finite-/spectral-element
130 discretized form of the first-order momentum and mass conservation equations (Brinkerhoff and Johnson, 2015). The minimization procedure is performed with a pseudo-transient continuation method, where the functional gradient is approximated by

$$\delta I = \int_{\Omega} (\bar{S} - \bar{S}_{\text{obs}}) \delta B d\Omega. \quad (2)$$

Rather than solving the forward equations to steady state before updating B , the bedrock elevation is continuously updated
135 while solving the forward model. This process occurs until B and H have both converged and the average smoothed surface elevation misfit falls below 1 m.

When surface velocity observations were available, the basal traction field β^2 was also solved for using an adjoint-based quasi-Newton method (Brinkerhoff and Johnson, 2013). A fixed-point iteration was used to ensure that the updated β^2 and B are mutually consistent. In the absence of velocity observations, β^2 was tuned such that approximately half of the resulting
140 surface velocity field is due to sliding. A uniform flow rate factor of $A = 3.17 \times 10^{-24} \text{ Pa}^{-3} \text{ s}^{-1}$ was used. The model was applied to all cases that were topologically contiguous and for which effective mass balance was available.

4.3 “Farinotti” (ITEM) – Farinotti et al. (2009b)

The method of Farinotti et al. (2009b) (also referred to as “ITEM”, e.g. Farinotti et al., 2009a; Gabbi et al., 2012), is based on mass conservation and principles of ice flow dynamics. Basically, the approach estimates the ice volume flux across profiles
145 located along manually prescribed ice flow lines, and converts it into ice thickness by using Glen’s flow law (Glen, 1955). For any point along a given flowline, the ice volume flux is approximated by integrating the ice flux divergence (difference between SMB and $\partial h / \partial t$, sometimes referred to as “apparent mass balance”) upstream of that point, whereas only the area within manually prescribed “ice flow catchments” is considered. The ice flux divergence is assumed to have a linear dependence on



elevation, and two separate gradients are prescribed for the glacier ablation and accumulation zone. The ice thickness obtained
150 along the individual flow lines is then interpolated across the glacier, and the local surface slope is used to modulate the
resulting local ice thickness.

With the exception of the correction factor C (cf. Equation 7 in Farinotti et al., 2009b), the same model parameters as in
Farinotti et al. (2009a) (see their Tab. 2) were used for ITMIX. C was set to 0.65, i.e. a value about 15 % higher than in the
original publications. This is because Gabbi et al. (2012) suggested a systematic overestimation in the derived ice thickness
155 distributions, and because a higher value of C translates into a lower ice thickness. Although the approach was originally
designed for mountain glaciers only, it was applied to all 21 test cases, including ice caps.

4.4 “Fuerst” – Fürst et al. (unpublished)

The unpublished approach by Fürst et al. relies on the Elmer/Ice model (Gagliardini et al., 2013), extended by the mass conser-
vation method of Morlighem et al. (2011) (see Sec. 4.13). This inverse approach aims at finding an optimal ice thickness field
160 by minimizing a predefined cost function. The minimization makes use of the adjoint system of the mass conservation equa-
tion, and adjusts both surface velocities and SMB fields. The cost function is a linear combination of components penalising
(i) negative thickness values, (ii) the mismatch between modelled and observed surface velocities and SMB, (iii) strong spatial
variations in the inferred thickness field, and (iv) the adjusted velocity field.

For accommodating the real-world cases within ITMIX, the above procedure was adapted for specific cases. For Unteraar,
165 for example, no velocity information was available in the glacier accumulation area, and the velocity-mismatch penalty was
therefore reduced in the corresponding areas. For Austfonna, on the other hand, the provided ice flow directions were highly
variable. Flow directions were thus prescribed from surface slopes. Since the approach relies on extensive input from surface
observations, only five test cases were considered (the three synthetic ones and two real-world cases).

4.5 “Gantayat” – Gantayat et al. (2014)

170 The approach by Gantayat et al. (2014) is based on the shallow ice approximation (e.g. Cuffey and Paterson, 2010) and Glen’s
flow law (Glen, 1955). It solves the equation

$$u_s = u_b + \frac{2A}{n+1} (f\rho g H \sin \alpha)^n \quad (3)$$

for ice thickness H , where u_s and u_b are surface and basal velocities respectively, A is the flow rate factor, $n = 3$ the creep
exponent, f a shape factor, $\rho = 900 \text{ kg m}^{-3}$ the ice density, $g = 9.81 \text{ m s}^{-2}$ the gravitational acceleration, and α the surface
175 slope. The equation is solved within 100 m elevation bands and the result is smoothed with a kernel of 3×3 grid cells to obtain
the final ice thickness distribution.

For ITMIX, $A = 3.2 \times 10^{-24} \text{ Pa}^{-3} \text{ s}^{-1}$ and $f = 0.8$ where chosen for all test cases; u_s was obtained from the provided
surface velocity fields; and $u_b = 0.25 u_s$ was assumed (Gantayat et al., 2014). Because of the data requirements, only 7 test
cases were considered.



180 4.6 “Gantayat-v2” – modified Gantayat et al. (2014)

“Gantayat-v2” is a modified version of the approach by Gantayat et al. (2014). Instead of solving Equation 3 for elevation bands, the same equation is first solved along discrete points of manually digitized branchlines (e.g. Linsbauer et al., 2012), and the resulting ice thickness is spatially interpolated by assuming zero ice thickness at the glacier margin. For the interpolation, the ANUDEM-algorithm (Hutchinson, 1989) is used.

185 For ITMIX, individual branchlines were generated requiring (a) a lateral spacing between adjacent lines of about ~ 200 m, (b) a minimal distance of ~ 100 m from the glacier margin, and (c) that branchlines from individual glacier tributaries gradually merge with the branchlines of the main stream. The same 7 test cases as for “Gantayat” were considered, and the same parameters used.

4.7 “GCbedstress” – Clarke et al. (2013)

190 The bed-stress method of Clarke et al. (2013) shares many conceptual features with Farinotti et al. (2009b) but differs in its implementation. Glacier flowsheds are hand-delineated and then transversely dissected by ladder-like “rungs” that represent flux gates oriented roughly perpendicular to the local ice flow direction. Ice discharge through these flux gates is calculated by integrating the apparent balance ($SMB - \partial h / \partial t$) over the upstream area associated with each gate. Discharge values associated with each rung are then applied to intervening cells by interpolation in a process that is equivalent to inserting rungs. The
195 average ice flux per unit width of channel is found by dividing ice discharge by the channel width for each cell. This width is taken to be the sum of the distance from the nearest channel boundaries to the left and right of the downflow direction. “Raw” ice thickness estimates are obtained from the ice flux using Glen’s law (flow rate factor $A = 2.4 \times 10^{-24} \text{ Pa}^{-3} \text{ s}^{-1}$), together with the estimated surface slope for each cell and the inclined slab flow assumption. The raw estimates are then smoothed by minimizing a cost function that negotiates a tradeoff between accepting the raw estimates or maximizing the smoothness of the
200 solution. Zero ice thickness is used as a boundary condition at ice-free margins.

When SMB fields were not provided within ITMIX, these were constructed assuming a linear variation with elevation above and below the estimated equilibrium line altitude. When $\partial h / \partial t$ fields were lacking these were assumed to vanish, or take a constant value, or vary linearly with elevation, depending on available information. No sliding was assumed, which leads to systematically higher ice thickness estimates than in the case of sliding. The tradeoff between raw estimates and smoothness
205 was found to depend on the grid spacing in a manner unforeseen by Clarke et al. (2013) so that the assigned parameter χ (see their Eq. 6) differs among the considered test cases.

4.8 “GCneuralnet” – Clarke et al. (2009)

The artificial neural net (ANN) method of Clarke et al. (2009) is based on the assumption that presently glacierized areas denuded of their ice cover would resemble nearby ice-free landscapes. This assumption relies on the geomorphic premise that
210 “landscape signatures of glaciation are the expression of regional influences such as geology, climate, and the intensity of past glaciations” (Clarke et al., 2009). The ice thickness is estimated considering the minimum range distance from an on-glacier



site to enclosing valley walls in eight compass directions (45° aperture) using an azimuthal stencil that has two elevation layers. The maximum search range for the stencil is limited by the map dimensions and other considerations. To train the ANN the stencil is centred on an ice-free cell and range distances to valley walls are measured for each sector and layer of the stencil as
215 the thickness of ice cover is increased.

The surface DEMs provided within ITMIX are unsuitable for direct application of the method. This is because the DEMs either lack elevation data beyond the glacier margins, or because the tight framing may cause the stencil to probe the frame edge. The provided map domains were therefore artificially expanded by reflecting elevation and ice mask data at the frame boundaries. This expedient falsifies the topography and ice cover beyond the frame boundaries. It can only be justified because
220 it allows stencil calculations to proceed and the ANN method to be included in ITMIX. Note, moreover, that the geomorphic premise makes the ice caps (Academy, Austfonna and Devon) and crater mountain glaciers (Elbrus and Mocho) considered within ITMIX unsuitable candidates for the ANN method. This is because little about the nature of their subglacial topography can be inferred from the geometric character of the surrounding ice-free terrain.

4.9 “Huss” (HF-model) – Huss and Farinotti (2012)

225 The method of Huss and Farinotti (2012) further develops the approach by Farinotti et al. (2009b). It avoids the digitization of glacier flowlines, includes additional physics (e.g. basal sliding, longitudinal variations in the valley shape factor, influence of ice temperature and climatic regime), and is applicable at the global scale. Glacier hypsometry and surface characteristics (mean slope and width) are evaluated for 10 m elevation bands, and all calculations are performed using this simplified 2D shape. Apparent mass balance gradients for the ablation and accumulation area (see Farinotti et al., 2009b) are estimated based
230 on the continentality of the glacier, which is derived from local equilibrium line altitudes. Ice volume fluxes along the glacier are converted into ice thickness using an integrated form of Glen’s flow law. The variations in the valley shape factor and the basal shear stress in the longitudinal glacier profile are taken into account. Simple parameterizations describe both the temperature-dependence of the flow rate factor and the variability in basal sliding. Calculated mean elevation-band thickness is extrapolated to each cell of a regular grid considering local surface slope, and the distance from the glacier margin. For
235 marine-terminating glaciers, a fixed ice volume flux is prescribed at the glacier terminus.

For ITMIX, all model parameters were set to the values used in Huss and Farinotti (2012). For the synthetic cases, geographical position and continentality defining the spatial variation of the model parameters, were not provided within ITMIX. These variables were thus estimated from the supplied surface mass balance distribution.

4.10 “Linsbauer” (GlabTop) – Linsbauer et al. (2009, 2012)

240 The *Glacier bed Topography (GlabTop)* method (labelled “Linsbauer” throughout the manuscript) by Linsbauer et al. (2009, 2012) uses an empirical relation between average basal shear stress τ and glacier elevation range (Haeberli and Hoelzle, 1995) for calculating the ice thickness at individual points along manually digitized glacier branchlines. From τ and the zonal surface slope α (computed within 50 m elevation bins along the branchlines) the ice thickness h is calculated as $h = \tau / (f \rho g \sin \alpha)$, where f is a shape factor, $\rho = 900 \text{ kg m}^{-3}$ the ice density, and $g = 9.81 \text{ m s}^{-2}$ gravitational acceleration. The dependence on



245 α implies thin ice where the glacier surface is steep and thick ice where it is flat. A distributed ice thickness is obtained by
interpolating the estimated point-values within the glacier outlines.

For ITMIX, branchlines covering all glacier branches and tributaries were digitized manually. A maximal value of $\tau =$
150 kPa was assumed and $f = 0.8$ was set. The remaining parameters, as well as the interpolation algorithms used for mod-
elling, are the same as in Linsbauer et al. (2012). Note that GlabTop was designed for alpine glaciers and has not been applied
250 for ice caps so far. Within ITMIX, however, it was applied to all 21 test cases, including ice caps.

4.11 “Machguth” (GlabTop2) – Frey et al. (2014)

GlabTop2, labelled with “Machguth” and fully described in Frey et al. (2014), is based on the same concept as the “Linsbauer”
model (see previous section). In particular, local ice thickness is calculated from an estimate of the basal shear stress and the
surface slope. The laborious process of manually drawing branchlines, however, is rendered obsolete by computing the surface
255 slope from the average slope of all grid cells within a predefined elevation buffer. The method is entirely grid-based and first
calculates the ice thickness at a set of randomly selected grid cells. In a second step, this thickness is interpolated to the entire
glacier area. To achieve realistic glacier cross-sections, the interpolation scheme assigns a minimum, non-zero ice thickness to
all grid cells directly adjacent to the glacier margin.

For ITMIX, the identical settings as in Frey et al. (2014) were used. A maximal ice thickness of 1000 m, however, was intro-
260 duced to avoid excessive glacier thickness in very flat areas of ice caps. To avoid influence of small-scale surface undulations on
modelled ice thickness, all provided DEMs were down-sampled to 75 m cell size for the calculations, and then re-interpolated
to the original resolution. The method is fully automated and fast, but includes a non-physical, tunable parameter controlling
the random point sampling. The model has been shown to be well-suited for mountain glaciers (Frey et al., 2014) but was not
applied to ice caps so far. All test cases besides Austfonna were considered.

265 4.12 “Maussion” (OGGM) – Maussion et al. (unpublished)

The *Open Global Glacier Model (OGGM)* by Maussion et al. (unpublished) implements a procedure which extends and fully
automatizes the method of Farinotti et al. (2009b). Automatization is achieved by generating and ordering multiple flowlines
according to Kienholz et al. (2014). For each grid point along these flowlines, the drainage area and both the local slope and
glacier width are computed. The ice volume flux is calculated by integrating the SMB field within the corresponding upstream
270 drainage area, where SMB is derived from monthly temperature and precipitation (extracted from the CRU dataset; Harris
et al., 2014) with the temperature index model by Marzeion et al. (2012). The resulting ice volume flux is then converted
into an ice thickness following Farinotti et al. (2009b). Further information on OGGM, including its code, can be found at
www.oggm.org.

OGGM has Glen’s flow rate factor A as a free calibration parameter. For ITMIX, A was calibrated with the “observed”
275 average glacier thicknesses reported in *GlaThiDa v1* (Gärtner-Roer et al., 2014). During calibration, *GlaThiDa* entries referring
to any of the ITMIX test cases were omitted, and the resulting value of A applied to all test cases. Note that (a) the primary
goal of OGGM is the dynamical modelling of glaciers and the estimation of their total volume, and not the estimation of a



distributed subglacial topography, and (b) both the inversion and the calibration procedures were developed for valley glaciers. For ITMIX, ice caps were handled as the sum of individual glacier basins as provided by the Randolph Glacier Inventory v5.0
280 (Arendt et al., 2015).

4.13 “Morlighem” – Morlighem et al. (2011)

The method of Morlighem et al. (2011) was originally designed to fill gaps between ground-penetrating radar measurements over the Greenland and Antarctic ice sheets. Based on mass conservation, it computes the ice thickness by requiring the ice flux divergence to be balanced by the rate of thickness change and the net surface and basal mass balances. The strength of the
285 method lies in the capability of including direct ice thickness measurements (e.g. Morlighem et al., 2014), which is achieved by optimizing the ice flux divergence and depth-averaged velocities to minimize the misfit between observed and modelled thicknesses.

For ITMIX, the optimization sequence was not used, as no ice thickness measurements were provided. The algorithm thus relied solely on mass conservation. For the test cases including surface velocities, these were assumed to be equal to the depth
290 averaged velocities. In the other cases, the shallow ice approximation was used together with an assumption of no-sliding to convert the computed ice mass flux into ice thickness (flow rate factor $A = 9.3 \times 10^{-25} \text{ Pa}^{-3} \text{ s}^{-1}$). The method was only applied to test cases providing SMB.

4.14 “RAAJgantayat” – Adapted from Gantayat et al. (2014)

“RAAJgantayat” is a set of solutions derived with an independent re-implementation of the approach by Gantayat et al.
295 (2014) (see Sec. 4.5 for a description). This particular version was applied to four valley glaciers (Tasman, Unteraar, Brewster, NorthGlacier) and one synthetic case (Synthetic1).

For Brewster and NorthGlacier, the required surface flow velocity fields were obtained by interpolating the provided point velocities with a standard inverse-distance weighting technique. For the other cases, the available velocity field were used directly. To avoid unrealistically large ice thicknesses, distributed surface slope values derived from the provided DEMs were
300 filtered so that values below 2° were eliminated. For all test cases, the shape factor (flow rate factor) was set to $f = 0.75$ ($A = 3.2 \times 10^{-24} \text{ Pa}^{-3} \text{ s}^{-1}$). All other parameters (including ice density, creep exponent, gravitational acceleration) were set to the same values as in Gantayat et al. (2014).

4.15 “RAAJglabtop2” – Adapted from Linsbauer et al. (2009)

Similarly as above, “RAAJglabtop2” is a set of solutions stemming from an independent re-implementation of an existing
305 model, i.e. the model GlabTop2 by Frey et al. (2014) (see Sec. 4.11). Results were generated for three real-world geometries (Tasman, NorthGlacier, Unteraar) and one synthetic test case (Synthetic2).



For the simulations, a shape factor of $f = 0.74$ was used. The average basal shear stress τ , derived from the glacier's elevation range, was set to 130, 150, 150, and 97 kPa for the test cases North Glacier, Tasman, Unteraar, and Synthetic2, respectively. All other parameters were set to the values given in Frey et al. (2014).

310 **4.16 “Rabatel” – Rabatel et al. (unpublished)**

The method of Rabatel et al. (unpublished) consists of four main steps: (1) delineation of the glacier's central flow line on the basis of the surface DEM; (2) delineation of individual cross profiles perpendicular to the central flow line, the latter being defined by using available surface flow velocities; (3) quantification of the ice volume flux across the profiles and conversion into ice thickness; and (4) interpolation over the glacier surface area.

315 As in Farinotti et al. (2009b), a gridded surface DEM and two prescribed vertical gradients are used to compute a distributed “apparent mass balance” field. The ice volume flux across each transversal profile is then computed by integrating that field in the upstream area of each profile. The flux is subsequently converted into an average ice thickness for the profile assuming that the depth-averaged flow velocity corresponds to 90 % (an adjustable parameter) of the velocity at the surface (Cuffey and Paterson, 2010). This average thickness is then distributed along the cross profile using the profile's surface flow velocities as
320 additional information. A complete ice thickness distribution is finally obtained by interpolating the values obtained for various profiles. Because of the required input data and because of time constraints, only the test case “Synthetic1” was considered within ITMIX.

4.17 “VanPeltLeclercq” – Van Pelt and Leclercq (unpublished)

The approach by Van Pelt and Leclercq (unpublished) iteratively derives a distributed glacier bed topography by minimizing
325 the mismatch between modelled and observed glacier surface elevations (e.g. Leclercq et al., 2012; Michel et al., 2013; van Pelt et al., 2013). Following van Pelt et al. (2013), repeated time-dependent model runs are performed with a model for ice dynamics. The latter is based on the vertically integrated shallow ice approximation (e.g. Hutter, 1983), including Weertman sliding (Huybrechts, 1991), and is part of the ICEDYN package (Reerink et al., 2010). Model runs are stopped at the time the provided surface DEMs refer to. After every run, the misfit between modelled and observed surface elevation is computed. The
330 bed topography is then adjusted by a fraction of this misfit, thus resulting in a new bed topography for a next model iteration. In test-cases providing velocity data, the iterative procedure is stopped when a minimum velocity misfit is achieved. In absence of such data, the procedure is terminated when the average surface elevation misfit drops below 5 m.

Since ITMIX does not provide SMB time series, the model was run with constant forcing until equilibrium. To match modelled and observed glacier extents, the SMB fields of the individual test cases were adjusted with a constant, glacier-
335 optimized offset. Only test cases providing SMB information were considered. When velocity observations were available, these were additionally used to tune the model parameters affecting basal sliding and deformational flow.



5 Results and discussion

In total, 189 different solutions were submitted to ITMIX (Tab. 2). Three models (“Farinotti”, “Huss”, “Linsbauer”) were able to handle all 21 test cases, one model handled 20 cases (“Machguth”), and one model handled 19 cases (“Maussion”). Data availability was the main factor hindering the consideration of additional test cases. This is particularly true for the approaches (a) “Brinkerhoff”, “Brinkerhoff-v2”, “Morlighem”, and “VanPeltLeclercq”, requiring SMB at least, (b) “Gantayat”, “Gantayat-v2”, and “RAAJgantayat”, requiring surface velocity fields, (c) “Fuerst”, requiring SMB, $\partial h/\partial t$ and velocity fields simultaneously and (d) “GCneuralnet”, requiring surrounding ice-free terrain for algorithm training. For the approaches “GCbedstress”, “RAAJglabtop2”, and “Rabatel”, the time required for model set up was a deterrent for considering additional test cases.

5.1 Between-model intercomparison

Locally, the solutions provided by the different models can differ considerably. As an example, Figure 2 provides an overview of the solutions generated for the test case “Unteraar” (the real-world case considered by the largest number of models). The large differences between the solutions are particularly evident when comparing the average composite ice thickness (i.e. the composite of the local average thickness computed from the ensemble of provided solutions, Fig. 2a) with the local ensemble spread (Fig. 2b). Often, the local spread is larger than the local average. This observation holds true for most of the other test cases as well (not shown).

Figures 2c and 2d provide insights into the composition of the ensemble spread by presenting the composites of the minimum and maximum provided thicknesses, respectively. The models providing the most extreme solutions are depicted in Figures 2e and 2f. In the “Unteraar” example, the approach “GCneuralnet” (“Fuerst”) tends to provide the smallest (largest) local ice thickness of the ensemble (light green and yellow area in Fig. 2e and 2f, respectively). For the specific case, closer inspection shows that the very low ice thicknesses estimated by “GCneuralnet” are associated with the debris covered parts of the glacier, and to the steep slopes delimiting these parts in particular. This is an artefact introduced by the specific setup of the stencil used within the ANN method. In fact, Clarke et al. (2009) found that including steep ice in the definition of valley walls can be advantageous for ANN training. An unforeseen consequence is that steep ice walls close to debris-covered glacier ice are interpreted as valley walls as well, thus causing the surrounding ice thickness to be too thin. Flagging debris-covered glacier parts and treating them as a special case could be an option for alleviating this issue. For “Fuerst”, large ice thicknesses (locally reaching 1000 m) mostly occur in the accumulation area. This is the area for which no measured ice flow velocities were available, thus precluding precise model constraint. The generated inconsistencies are propagated in flow direction, and cause artefacts for areas in which velocity data are available as well. For the particular test case, the approach also provides the minimal ice thickness for large areas, indicating that important oscillations are present in the estimated ice thickness field.

The overall tendency for individual models to provide “extreme” solutions is shown in Figure 3. Two models (“Rabatel” and “GCbedstress”) seem to be particularly prone to predict large ice thicknesses, providing the largest ice thickness of the ensemble for 33 and 25 % of the area they considered. Although for “Rabatel” the basis of the statement is weak (only one test case considered) possible explanations lie in (a) the possible overestimation of the area contributing to the ice volume



370 flux of individual profiles, and (b) the assumed relation between depth-averaged and surface flow velocity (cf. Sec. 4.16). For
“GCbedstress” the possible reason are less clear. The no-sliding assumption – which causes systematically higher thickness
estimates than if sliding is assumed – could be a reason. The model, however, seems not to be particularly sensitive to it:
Assuming that half of the surface velocity is due to sliding decreases the mean estimated thickness by 13 % only (not shown).

Very small ice thicknesses are often predicted by the models “Maussion” and “GCneuralnet”. The two models provided
375 the smallest ice thickness of the ensemble in 30 and 23 % of the considered area, respectively. For “Maussion”, the result is
mainly driven by the ice thickness predicted for ice caps (Academy, Austfonna, Devon) and large glaciers (Columbia, Elbrus).
This is likely related to the applied calibration procedure (cf. Sec. 4.12), which is based on data included in GlThiDa v1. The
observations in that dataset, in fact, mostly refer to smaller glaciers (Gärtner-Roer et al., 2014). For “GCneuralnet”, it can be
noted that the smallest ice thicknesses are often predicted along the glacier centrelines (not shown). Besides the previously
380 discussed issue related to steep ice in proximity of e.g. medial moraines, the ad-hoc solution adopted to allow the ANN stencil
to be trained (see Sec. 4.8) might be an additional cause.

Although the above observations provide insights into the general behaviour of individual models, it should be noted that a
tendency of providing extreme results is not necessarily an indicator of poor model performance. Actual model performance,
in fact, can only be assessed through comparison against direct observations (see next section).

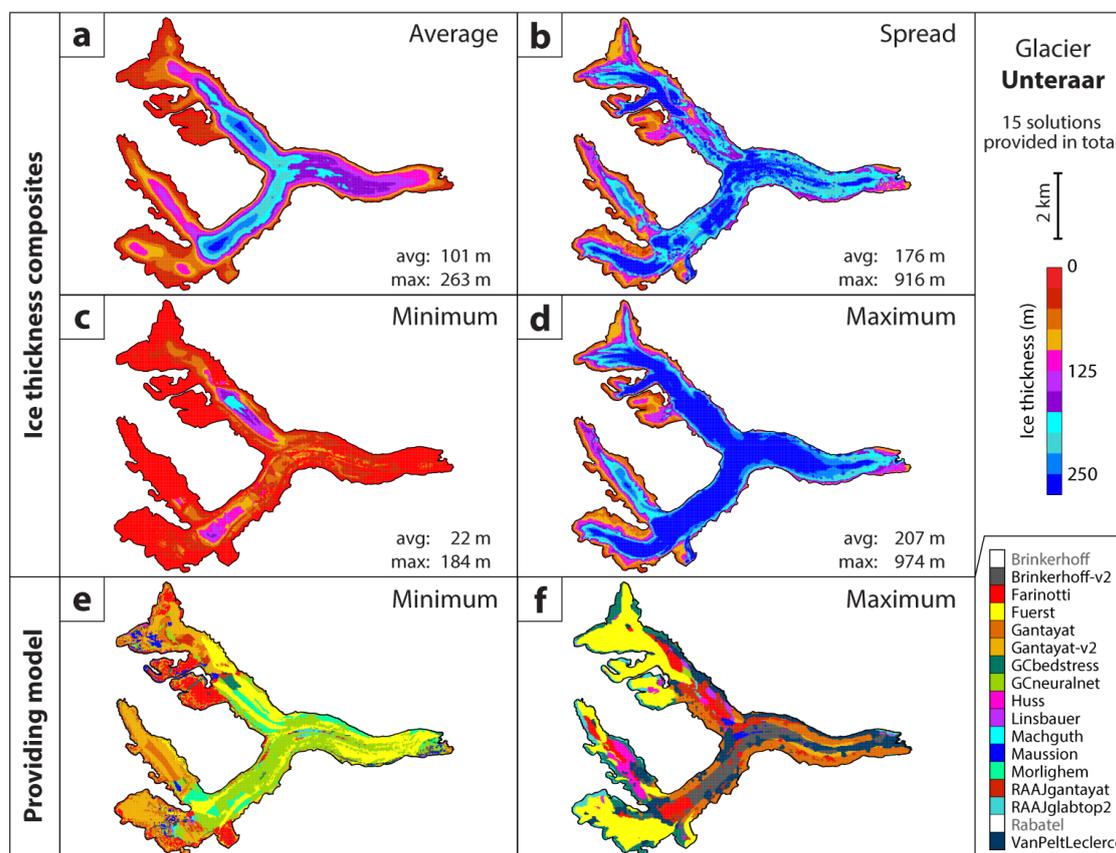


Figure 2. Overview of the range of solutions provided by the ensemble of models. The example refers to the test case “Unteraar”. The first four panels show composites for the (a) average, (b) spread, (c) minimal, and (d) maximal ice thickness distribution of the 15 submitted solutions. The model providing the minimal and maximal ice thickness for a given location is depicted in panels (e) and (f). Models that did not consider the specific test case are greyed out on the bottom right legend.

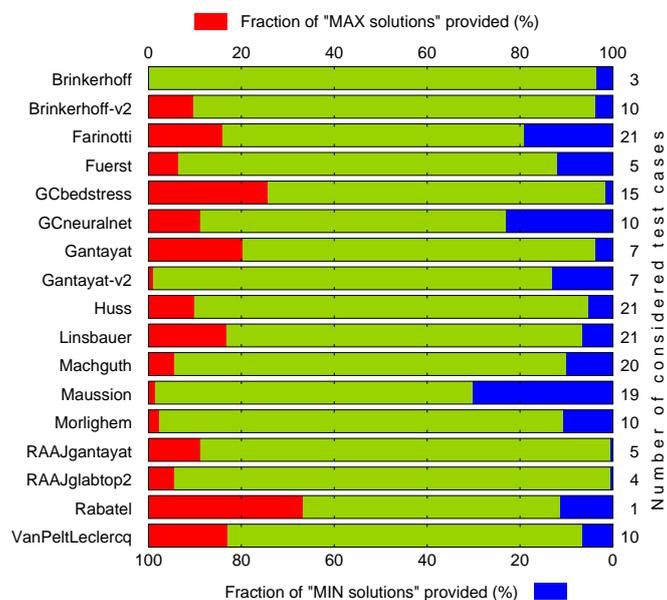


Figure 3. Share of “extreme results” provided by individual model versions. An “extreme result” is defined as either the minimum (MIN) or maximum (MAX) ice thickness occurring in the ensemble of solutions provided for a given test case. The share is based on test case area and assigns equal weights to all cases (a 10 % “fraction of MAX solutions provided” indicates, for example, that on average, the model generated the maximal ice thickness for 10 % of the area of any considered case). The number of test cases considered by individual models is given.

385 5.2 Comparison to ice thickness measurements

The solutions submitted by individual models are compared to ice thickness measurements in Figures 4 and 5. For every glacier, the figures show one selected profile along and one across the main ice flow direction. The previously noted large spread between individual solutions re-emerges, as well as the tendency of individual models to produce rather large oscillations. The spread is particularly pronounced for ice caps (Academy, Austfonna, Devon) and for across-flow profiles (Fig. 5).

390 In light of the large spread, it is important to note that the average solution of the model ensemble (thick green line in the figures) matches the direct measurements rather closely in most of the cases (average deviation below 10 % in 17 out of 21 cases). This suggests that the deviations of the individual models are independent and not biased, for example, by sets of common assumptions. For a set of independent random realizations of the same variable, in fact, Poisson’s *law of large numbers* predicts the average result to converge to the expected value (the “true bedrock” in this case) with increasing number
 395 of realizations.

The positive effect of averaging the results of individual models is best seen in Figure 6. On average over the individual model solutions, the difference between modelled and measured ice thickness is $-17 \pm 36 \%$ (1σ estimate) of the mean glacier thickness (first boxplot in the “ALL” group). This value reduces to $+10 \pm 24 \%$ when the average composite solution is considered, and is close to the value obtained when selecting the best single solution for every test case individually (third and second
 400 boxplots of the group, respectively).



Two notable exceptions in the above considerations are given by the test cases “Unteraar” and “Tasman”, for which the ensembles of solutions (15 and 11 solutions provided, respectively) converge to a significantly smaller ice thickness than observed (median deviations of -84% and -65% , respectively). Two common features that might partially explain the observation are (a) the significant debris cover of the two glaciers, that might bury ice thicker than what would be expected from the present-day SMB fields, and (b) the branched nature of the glaciers, that might be insufficiently captured by the models. Both hypotheses, however, are difficult to test further, as the remaining cases show very different morphological characteristics. An erroneous interpretation of the actual ice thickness measurements, on the other hand, seems unlikely. This is particularly true for Unteraar, for which the reported quality of original radio-echo soundings is high and independent verifications through borehole measurements were performed (Bauder et al., 2003).

“Urumqi” and “Washmawapta”, for which 8 and 6 individual solutions were provided respectively, are the other two cases for which the average ice thickness composite differs largely from the observations (median deviations of -71% and -125% , respectively; Figs. 4, 5, and 6). For “Washmawapta” – a cirque glacier mostly fed by steep ice-free headwalls (Sanders et al., 2010) – it is interesting to note that the “Farinotti” approach is the only one predicting ice thickness in the observed range. This suggests that the concept of “ice flow catchments”, which is used in the approach for accommodating areas outside the glacier margin that contribute to snow accumulation (cf. Farinotti et al., 2009b), is an effective workaround for taking such areas into account. Failure of doing so, in fact, causes the ice volume flux (and thus the ice thickness) to be underestimated. For “Urumqi”, on the other hand, the reasons for the substantial underestimation of actual ice thickness are less clear. Potentially, they could be linked to (a) the cold nature of the glacier (e.g. Maohuan et al., 1989), which requires thicker ice to produce a given surface velocity (note that most models assumed flow rate factors for temperate ice), and (b) the artefacts in the provided DEM (note the step-like features in the surface shown in Fig. 4), which lead to locally very high surface slope and thus low ice thickness.

The comparison between Figures 4 and 5 also suggests that, in general, the ice thickness distribution along-flow is better captured than the distribution across-flow. This is likely due to the combination of the fact that most participating approaches include considerations about mass conservation, and that virtually all models include surface slope as a predictor for the local ice thickness. These two factors, in fact, have a stronger control on the along-flow ice thickness distribution than they have across-flow.

The results also indicate that, compared to real-world cases, the ice thickness distribution of the three synthetic cases is better reproduced. On average over individual solutions, the difference to the correct ice thickness is $-17 \pm 20\%$ (Fig. 6). This difference reduces to $-15 \pm 11\%$ for the average composites, i.e. to a 1σ -spread reduced by a factor of two. Again, two factors provide the most likely explanation. On the one hand, the model used for generating the synthetic cases is built upon the same theoretical knowledge as the models used for generating the ice thickness estimates. On the other hand and more importantly, the input data from which the ice thickness distribution is inferred are known without any uncertainty in the synthetic cases. The latter is in contrast to the data available for the real-world cases: Whilst the provided DEMs, $\partial h/\partial t$ fields, and outlines can be considered of good quality, SMB fields are often the product of the inter- and extrapolation of sparse in-situ measurements. The inconsistencies that may arise between $\partial h/\partial t$ and SMB, together with the previously mentioned



discontinuities in the available velocity fields (cf. Sec. 3), are obviously problematic for methods that use this information. Two additional observations that might be related to the better model performance in the synthetic cases are (1) that the no-sliding assumption adopted in most models was adequate for the considered synthetic cases, but does not hold true in the real-world ones, and (2) that synthetic glacier geometries were close to steady state. Testing the importance of the second consideration is not possible with the data at hand, and would require the generation of transient synthetic geometries. It has to be noted, however, that apart from “VanPeltLeclercq” (cf. Sec. 4.17) none of the participating models are explicitly based on a steady state assumption.

In relative terms, the average composite solutions seem to better predict (smaller interquartile range, IQR) the ice thickness distribution of ice caps than that of glaciers. In fact, the 1σ -deviations from the measurements for ice caps and glaciers are of $12 \pm 16\%$ and $12 \pm 34\%$, respectively (Fig. 6). This might be surprising at first, but Figure 4 illustrates that for all three considered ice caps, the average composites are the results of a relatively small set (6 or 7) of largely differing solutions. This issue is particularly evident for the ice cap interiors, for which two model clusters emerge, predicting extremely high and extremely low ice thicknesses, respectively. The relatively small IQR of the ensemble mean, thus, appears to be rather fortuitous, and calls for additional work in this domain. Note, moreover, that the relative accuracy is expressed in relation to the mean ice thickness. In absolute terms, the abovementioned values translate into average deviations in the order of 48 ± 63 m for ice caps and 11 ± 27 m for glaciers. Obviously, these values are strongly affected by the particular test cases included in the intercomparison, and should not be expected to hold true in general.

To put the average model performance into context, the results are compared to a benchmark model based on volume-area scaling (last boxplot in Fig. 6). The “model” neglects spatial variations in thickness altogether, and simply assigns the mean ice thickness predicted by a scaling relation to the whole glacier. For the scaling relation, we use the form $\bar{h} = cA^{\gamma-1}$, where h (m) and A (km^2) are the mean ice thickness and the area of the glacier, respectively. The parameters c and γ are set to $c = 0.034$ and $\gamma = 1.36$ for glaciers (Bahr et al., 2015), and to $c = 0.054$ and $\gamma = 1.25$ for ice caps (Radić and Hock, 2010).

This simple model deviates from the measured ice thickness by $-42 \pm 59\%$, which is a spread (bias) more than twice (four times) as large as estimated for the average composites of the model ensemble ($10 \pm 24\%$). This result is reassuring as it suggests that the individual models have actual skill in estimating both the relative ice thickness distribution and the total glacier volume of individual glaciers.

5.3 Individual model performance

The considerations in the previous section refer mainly to the average composite solution provided by the ensemble of models. Running a model ensemble, however, can be very impractical. This opens the question on whether individual models can be recommended for particular settings, or whether a single best model can be identified.

To address this question, we propose two separate rankings. Both are based on the (I) average, (II) median, (III) interquartile range, and (IV) 95% confidence interval (95% CI) of the distribution of the deviations between modelled and measured ice thicknesses (Fig. 7).

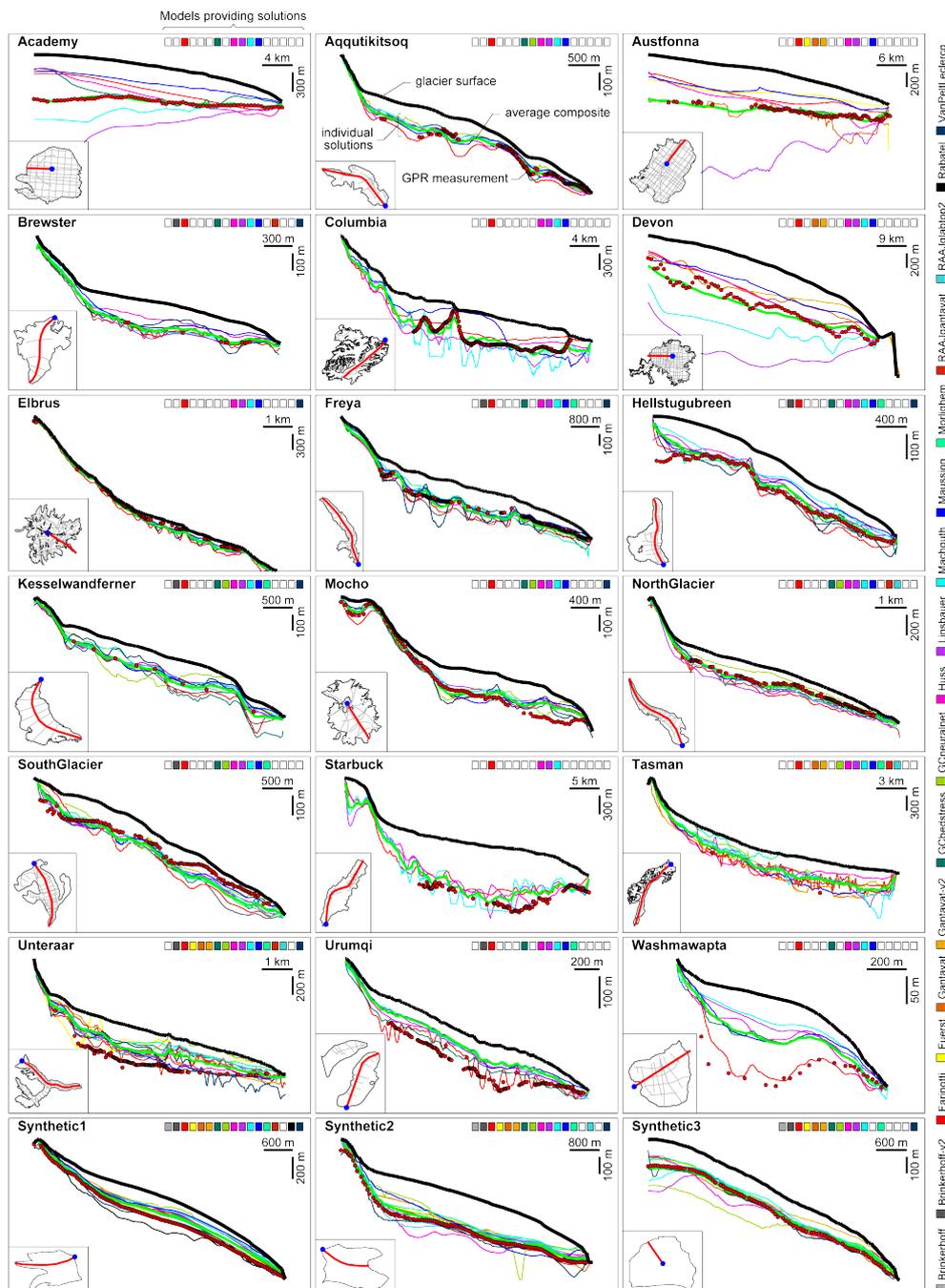


Figure 4. Comparison between estimated and measured bedrock topographies. For every test case, a longitudinal profile showing the glacier surface (thick black line), the bedrock solution of individual models (coloured lines), the average composite solution (thick green line), and the available GPR measurements (black-encircled red dots) are given. The coloured squares on the upper left of the panels indicate which models provided solutions for the considered test case (see legend on the right margin for colour key). The location of the profiles are shown on the small map on the bottom left of the panels (red), and the beginning of the profile (blue dot) is to the left. Available ice thickness measurements are shown in grey.

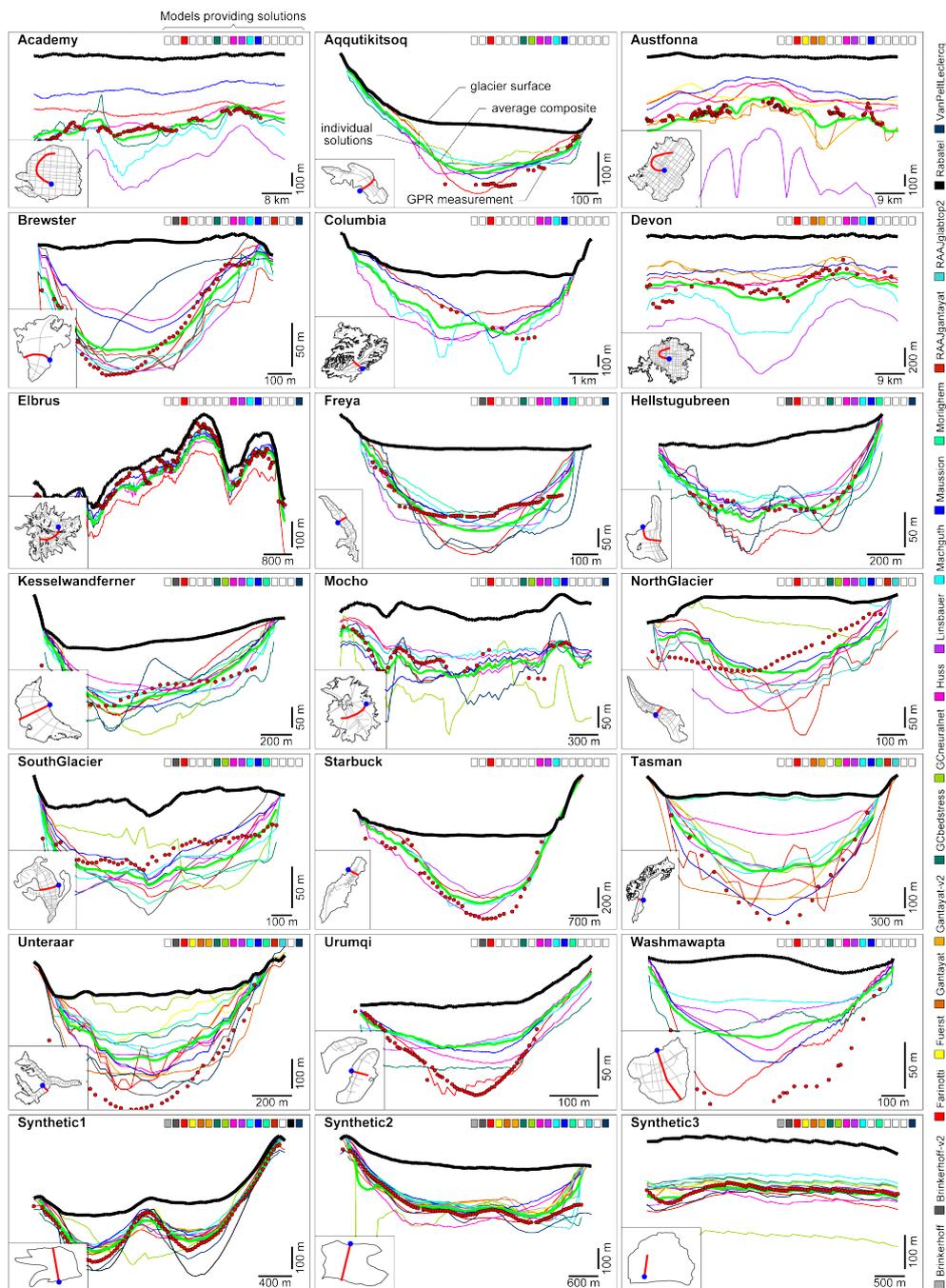


Figure 5. Same as Figure 4, for a series of cross-sectional profiles.

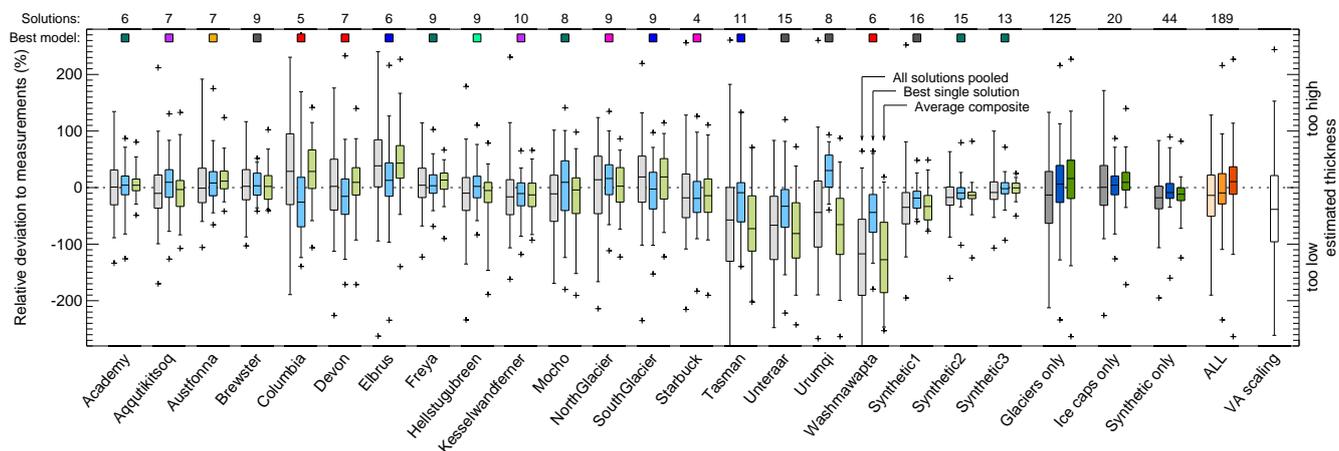


Figure 6. Effect of merging individual model solutions. For every test case, the distribution of the deviations between modelled and measured ice thicknesses is shown for the case in which (i) the individual point-to-point comparisons of all available solutions are pooled (grey boxplots), (ii) only the provided single best solution is considered (blue boxplots), and (iii) the deviations are computed from the average composite thickness of all model solutions (green boxplots). Deviations are expressed relative to the mean ice thickness. The best single solution is computed by summing the ranks for the (a) average deviation, (b) median deviation, (c) interquartile range, and (d) 95 % confidence interval. The distributions of the deviations when grouping glaciers, ice caps, and synthetic glaciers separately are additionally shown, as are the results when grouping all test cases together (ALL). When forming the groups, point-to-point deviations for every test case are resampled so that every test case has the same weight. The last boxplot to the right refers to the case in which the mean ice thickness is predicted by volume-area scaling (see Section 5.2). The upper part of the panel provides the number of considered model solutions, and the model providing the single best solution (see Fig. 4 for colour key). Boxplots show minimum and maximum values (crosses), the 95 % confidence interval (whiskers), the interquartile range (box) and the median (lines within box).

The first ranking considers the individual test cases separately. All models considering a particular test case are first ranked
 470 separately for the four indicators (I-IV). When a model does not include a particular test case, no ranks are assigned. For every
 model, the four indicators are then averaged individually over all test cases. The final rank is computed by computing the mean
 of these average ranks (Tab. 3). The ranking rewards models with a consistently high performance over a large number of test
 cases.

The second ranking is only based on the average model performance. In this case, ranks for the above indicators (I-IV)
 475 are assigned to the ensemble of point-to-point deviations of the various models (last row of boxplots in Fig. 7; same weight
 between test cases ensured). The ranks for the four individual indicators are then averaged to obtain the overall rank (Tab. 4).
 In contrast to the first option, this ranking does not consider the test cases individually, and does not account for the number of
 considered test cases. A model considering only one test case but performing perfectly on it, for example, would score highest.

The ranking result of every model on a case-by-case basis is given in Supplementary Table S2. The distributions of the
 480 deviations between modelled and measured ice thicknesses for every model and considered test case are given in Figure 7 and
 Supplementary Figure S1.



Combined over the two rankings, the model “Brinkerhoff-v2” scores highest (3rd and 1st rank, respectively). The good score is mainly driven by the comparatively small model spread (IQR and 95 % CI) and bias (Tab. 4). The small model bias (−3 % average deviation), however, arises from a partial compensation between positive bias for glaciers (+5 %) and negative bias
485 for the synthetic cases (−22 %) (Tab. 4). Unfortunately, the model did not consider any ice cap, thus hampering any statement on model performance in this particular setting. Ice caps were not considered mainly because of the absence of the necessary data.

Apart from the model “Brinkerhoff-v2”, the first positions in the first ranking are occupied by models that consider a large number of test cases (Tab. 3). The model by “Maussion” is rated highest. Similar to “Brinkerhoff-v2”, the good result is driven
490 by the small IQRs and 95 % CIs, in particular for glaciers and ice caps. In the second ranking, the model is severely penalised (11th rank) for its large bias (−36 % on average; Tab. 4). The bias is particularly prominent in the case of ice caps and the synthetic cases (−42 and −45 % average deviation, respectively), and may be related to the fact that the “Maussion” model was developed and calibrated by using data from valley glaciers only. For the synthetic cases in particular, the calibration with real-world glaciers (i.e. cases that include sliding) seems to be a likely explanation for a systematic underestimation of the ice
495 thickness. This, however, appears to be only a partial explanation, as such a negative bias is apparent for most approaches, i.e. also for approaches that explicitly assumed no sliding (e.g. “GCbedstress”, “Morlighem”; cf. Sec. 4).

In general, the model bias can be interpreted as an indicator for the performance of the models in reproducing the total glacier ice volume. The latter is not discussed explicitly as the computation of a “measured volume” would need the available measurements to be interpolated over large distances. Seven of the considered models show a bias of less than 8 % (Tab. 4).
500 An interesting case in this respect is given by the model by Gantayat et al. (2014), which yields small biases (−4 and −8 %) for both considered implementations (“Gantayat” and “RAAJgantayat”, respectively). The relative low overall ranks assigned to these models (ranks 10 and 14 in the first ranking, ranks 3 and 12 in the second, respectively) are an expression of the relatively small number of considered test cases (first ranking), and the relatively large model spread (second ranking). Of interest is also the observation that the version of the model considering multiple flowlines (“Gantayat-v2”) yields a significant
505 higher bias (−32 % on average) than the approach based on one flowline only, despite a moderate decrease in model spread. The increase is particularly visible for real-world glaciers, for which the bias changes from +4 % to −61 %. This might hint at the difficulty in correctly subdividing a given glaciers into individual flowlines, and could be an indication that the rather mechanistic procedure used in this case (cf. Sec. 4.6) is insufficient for achieving a sensible subdivision.

The difficulty in correctly interpreting the overall model bias is well illustrated in the case of the “Linsbauer” model: The
510 model yields the smallest bias over the entire set of considered test cases (−1 % on average), but is the result of a compensation between (a) a moderate negative bias for glaciers and the synthetic test cases (both −16 %), and (b) a large positive bias for ice caps (+91 %).

Together with “Brinkerhoff-v2”, the model “Farinotti” is the second one included in the first five places of both rankings (ranks 4 and 5, respectively; Tabs. 3 and 4). The relatively high ranking is due to a combination of comparatively high model
515 performance (small bias and spread) and large number of considered test cases. The consideration of all test cases, however, should not be interpreted as capability of handling large samples of glaciers in this case. The application of the model, in fact,



requires a significant amount of manual input (cf. Sec. 4.3). This is in contrast to the fully automated methods of “Maussion”, “Huss”, and “Machguth”. In this respect it is interesting to note that the model by “Huss” slips from the 2th rank in the first ranking to the 8th in the second one. The relatively low score in the second ranking is mainly an expression of the comparatively large confidence intervals (Tab. 4). Combined over the two rankings, however, the model can be considered as the best amongst the fully automated approaches.

The model “GCbedstress” (5th and 6th in the two rankings) ranks highest when only ice caps are considered. The average deviation of $3 \pm 17\%$ indeed suggests a very high model performance. However, it has to be noted that the result is based on one test case only (Tab. 2). For the models considering more ice caps, the results are heterogeneous and difficult to interpret, as models showing small IQRs show large bias, and vice versa (Tab. 4).

The model “GCneuralnet” is found at the other end of the ranking (penultimate and last ranks, respectively). The average deviation of $-39 \pm 52\%$ highlights both the large bias and large spread of the estimates. Obviously, the performance of approaches based on ANN are highly dependent on the data set used for algorithm training. The large deviations might therefore be an expression of the issues encountered with the provided DEMs (cf. Sec. 4.8), rather than an indication of generally low model performance. As already noted, however, the general absence of ice-free analogues for ice caps or crater glaciers makes the approach unsuitable for this kind of morphologies.

An interesting result emerges when considering the IQRs and 95 % CIs in the synthetic test cases (cf. Tab. 4): Approaches that include SMB, $\partial h/\partial t$ or velocity information in addition to the glacier outline and the DEM of the surface (e.g. approaches by “Brinkerhoff”, “Fuerst”, “Morlighem”, or “VanPeltLeclerq”) yield the smallest spreads around the average deviation (IQR $< 22\%$ for all mentioned models). This is in rather marked contrast to the real-world cases, in which the IQRs are about five times larger (average IQR for the same models = 105 %), and similar considerations apply when analysing the 95 % CIs. As already noted, this is most likely linked to the differences in data quality. Whilst the input data for the synthetic cases are perfectly known, the data available for the real-world cases were necessarily retrieved from various, independent data sources (cf. Tab. 1). This often led to problems in the mutual consistency of the surface fields, and caused particular difficulties to those models requiring all of the information. The capability of accounting for observational uncertainties, as in the “Brinkerhoff” approach for example, hence seems to be an important prerequisite when handling real-world cases. Similarly, having access to reliable uncertainty estimates for any particular dataset would be important. Emphasis has to be put in this domain if significant advances are to be achieved.

6 Conclusions

ITMIX was the first coordinated intercomparison of approaches that estimate the ice thickness of glacier and ice caps from surface characteristics. The goal was to assess model performance for cases in which no a-priori information about ice thickness is available. The experiment included 15 glaciers and 3 ice caps spread across a range of different climatic regions, as well as 3 synthetically generated test cases.

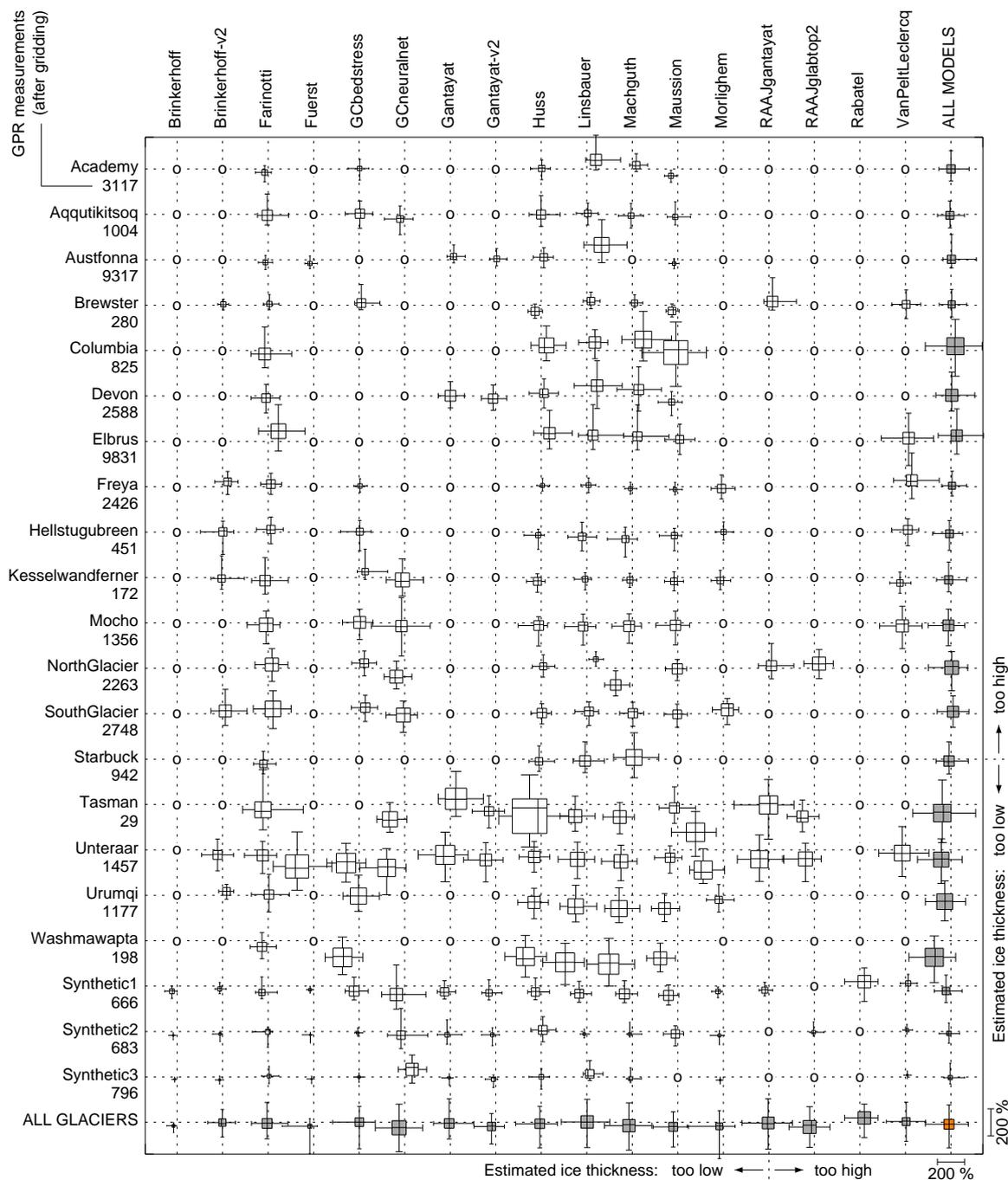


Figure 7. Difference between estimated and measured ice thicknesses. For every test case (rows) and every model (columns), the distribution of the point-by-point deviations between estimated and measured ice thicknesses is shown. Differences are expressed relative to the mean ice thickness. A 100 % deviation means, for example, that the modelled ice thickness deviates from the measured one by one mean ice thickness. Circles indicate that no solution was submitted. Boxplots show the 95 % confidence interval (whiskers), the interquartile range (box) and the median (lines within box). The boxplots are squared to facilitate the comparison within models and within test cases. Note the scale in the bottom right corner.



Table 3. Ranking of individual models based on case-by-case performance. Displayed values are mean ranks for the average (avg), median (med), interquartile range (IQR) and 95 % confidence interval (95 %) of the deviations from ice thickness measurements. For every group (glaciers, ice caps, and synthetic cases), the average of the above three ranks is given (AVG). Values for “ALL” correspond to the average of the three groups. The three best average results for every group are given in bold. *n* is the number of considered test cases per model.

Model - version	n	Glaciers only					Ice caps only					Synthetic only					ALL				
		avg	med	IQR	95%	AVG	avg	med	IQR	95%	AVG	avg	med	IQR	95%	AVG	avg	med	IQR	95%	AVG
Maussion	19	4.6	4.3	2.9	3.0	3.7	5.3	5.3	1.3	1.0	3.2	13.0	11.0	12.5	10.5	11.8	5.6	5.2	3.7	3.5	4.5
Huss	21	4.5	4.3	4.3	3.5	4.1	2.0	2.0	4.3	5.0	3.3	4.7	3.3	12.0	12.7	8.2	4.1	3.8	5.4	5.0	4.6
Brinkerhoff-v2	10	3.6	3.6	4.6	5.3	4.2	-	-	-	-	-	7.7	6.7	3.3	4.3	5.5	4.8	4.5	4.2	5.0	4.6
Farinotti	21	3.7	3.9	5.5	5.6	4.7	2.7	3.3	2.7	3.0	2.9	5.3	6.0	9.0	10.7	7.8	3.8	4.1	5.6	6.0	4.9
GCbedstress	15	4.6	4.8	5.7	5.5	5.2	2.0	1.0	1.0	5.0	2.2	4.0	5.3	8.3	7.3	6.2	4.3	4.7	5.9	5.9	5.2
VanPeltLeclercq	10	4.1	4.3	7.1	7.9	5.9	-	-	-	-	-	4.3	4.0	5.3	4.7	4.6	4.2	4.2	6.6	6.9	5.5
Linsbauer	21	5.4	5.3	3.5	3.9	4.5	6.7	6.7	6.7	6.7	6.7	11.0	8.3	9.0	8.0	9.1	6.4	6.0	4.8	4.9	5.5
Machguth	20	5.9	5.6	4.1	3.7	4.8	4.5	4.5	5.0	4.5	4.6	8.7	10.7	10.0	10.3	9.9	6.2	6.2	5.0	4.8	5.5
Brinkerhoff	3	-	-	-	-	-	-	-	-	-	-	9.7	9.3	2.3	2.0	5.8	9.7	9.3	2.3	2.0	5.8
RAAJgantayat	5	4.8	4.8	7.8	9.8	6.8	-	-	-	-	-	4.0	3.0	6.0	5.0	4.5	4.6	4.4	7.4	8.8	6.3
RAAJglabtop2	4	7.0	6.7	6.7	6.3	6.7	-	-	-	-	-	1.0	3.0	9.0	8.0	5.2	5.5	5.8	7.2	6.8	6.3
Fuerst	5	13.0	14.0	15.0	15.0	14.2	4.0	6.0	2.0	2.0	3.5	7.7	8.0	1.3	3.3	5.1	8.0	8.8	4.2	5.4	6.6
Morlighem	10	6.7	6.7	6.0	4.9	6.1	-	-	-	-	-	12.0	11.3	4.3	3.7	7.8	8.3	8.1	5.5	4.5	6.6
Gantayat	7	4.0	4.0	11.0	10.0	7.2	3.5	2.0	5.5	4.0	3.8	8.0	7.7	8.7	9.0	8.3	5.6	5.0	8.4	7.9	6.7
Gantayat-v2	7	7.0	8.0	3.0	6.5	6.1	2.5	2.5	4.0	3.0	3.0	11.0	10.0	9.3	8.7	9.8	7.4	7.3	6.0	6.4	6.8
GCneuralnet	10	7.1	7.9	7.3	6.9	7.3	-	-	-	-	-	9.7	14.0	14.0	14.7	13.1	7.9	9.7	9.3	9.2	9.0
Rabatel	1	-	-	-	-	-	-	-	-	-	-	5.0	5.0	16.0	15.0	10.2	5.0	5.0	16.0	15.0	10.2

ITMIX attracted 13 research groups with 17 different model versions of various degrees of sophistication and data require-
 550 ments. The 189 solutions submitted in total provided insights into the performance of the various models, and the accuracies
 that can be expected from their application.

The submitted results highlighted the large deviations between individual solutions. The local spread often exceeded the
 local ice thickness. Caution is thus required when interpreting the results of individual models, especially if they are applied
 to individual sites. Substantial improvements in terms of accuracy, however, could be achieved when combining the results of
 555 different models. Locally, the mean deviation between an average composite solution and the measured ice thickness was in
 the order of $10 \pm 24\%$ of the mean ice thickness (1σ estimate). This hints at the random nature of individual model errors,
 and suggest that ensembles of models could help in improving the estimates. For applications at the large scale – such as the
 estimation of the ice thickness distribution of an entire mountain range and beyond –, however, reducing the uncertainties
 through such a strategy will be challenging. This is because only few models are currently capable of operating at the regional
 560 or global scale.

The intercomparison also allowed statements about the performance of individual models. The model “Brinkerhoff-v2” was
 detected as the best single model, with average deviations for real-world glaciers in the order of $-3 \pm 27\%$. Some caution has



Table 4. Ranking of individual models based on average model performance. Displayed values are the average (avg), median (med), interquartile range (IQR) and 95 % confidence interval (95 %) of the percental deviations from ice thickness measurements. The three best results for every column is given in bold. n is the number of considered test cases per model. AVG is the average of the ranks assigned to the values in the three “ALL” columns.

Model - version	n	Glaciers only				Ice caps only				Synthetic only				ALL				AVG
		avg	med	IQR	95%	avg	med	IQR	95%	avg	med	IQR	95%	avg	med	IQR	95%	
Brinkerhoff-v2	10	5	6	±33	±120	-	-	-	-	-22	-19	±10	±39	-3	-9	±27	±102	3.8
Brinkerhoff	3	-	-	-	-	-	-	-	-	-29	-26	±10	±36	-29	-26	±10	±36	5.8
Gantayat	7	4	-7	±72	±200	16	14	±29	±93	-22	-18	±22	±68	-4	-7	±35	±148	6.0
VanPeltLeclercq	10	2	4	±43	±168	-	-	-	-	15	13	±11	±47	6	9	±29	±145	6.0
Farinotti	21	1	-3	±45	±148	-21	-21	±22	±76	-13	-3	±27	±76	-4	-7	±36	±135	6.5
GCbedstress	15	-4	7	±39	±168	3	4	±17	±77	-15	-9	±17	±71	-6	1	±32	±157	6.5
Linsbauer	21	-16	-4	±46	±167	91	79	±48	±159	-16	-17	±20	±106	-1	2	±46	±180	8.0
Huss	21	-21	-13	±39	±171	13	11	±26	±90	-9	-10	±27	±89	-14	-8	±35	±154	8.5
Gantayat-v2	7	-61	-62	±41	±140	-5	-3	±29	±90	-30	-25	±22	±64	-32	-28	±31	±112	8.8
Fuerst	5	-113	-135	±86	±228	-26	-30	±16	±47	-24	-22	±9	±28	-42	-26	±14	±131	9.0
Maussion	19	-34	-26	±36	±142	-42	-39	±19	±60	-45	-43	±35	±89	-36	-31	±33	±131	10.5
RAAJgantayat	5	-3	3	±50	±201	-	-	-	-	-28	-27	±22	±50	-8	-10	±43	±186	10.8
Machguth	20	-34	-23	±52	±190	39	33	±33	±132	-26	-24	±15	±91	-26	-18	±45	±175	11.0
Rabatel	1	-	-	-	-	-	-	-	-	29	34	±46	±123	29	34	±46	±123	11.5
RAAJglabtop2	4	-42	-43	±62	±162	-	-	-	-	-2	-9	±13	±53	-32	-21	±48	±151	12.2
Morlighem	10	-55	-26	±48	±237	-	-	-	-	-32	-28	±11	±41	-47	-27	±26	±215	12.5
GCneuralnet	10	-55	-50	±53	±181	-	-	-	-	0	-15	±63	±158	-39	-40	±52	±176	15.8

to be expressed, however, since the model considered only about half of the provided test cases and was not applied to any ice cap. The model “Huss” scored highest amongst the automated methods capable of handling large sets of glaciers. With average deviations of $-14 \pm 35 \%$, the approach ranged mid-way when considering point-to-point deviations from measurements. For ice caps, the model “GCbedstress” showed very promising results (average deviations of $3 \pm 17 \%$), although generalizing this observation would be speculative, as the approach considered only one test case.

Somewhat surprisingly, models that include SMB, $\partial h/\partial t$ or surface flow velocity fields in addition to the glacier outline and DEM did not perform better when compared to approaches requiring less data, in particular for real-world cases. Inconsistencies between available datasets – which are often acquired with very different techniques, spatial footprints, and temporal resolutions – appeared to be the most likely cause. The better model performance in the set of synthetic cases, for which input data were known precisely, supports this thesis. This highlights the importance for mutually consistent data sets, and suggests that improved observational capabilities could help to improve the performance of the next generation of ice thickness estimation methods. Similarly, improving the model’s capability of taking into account uncertainties in the input data should be considered a priority.



Besides improved data concerning glacier surface characteristics, a key for developing a new generation of ice thickness estimation models will be the data base against which the models can be calibrated and validated. The data utilized within ITMIX will be available as a supplement to this paper, but a much larger effort is ongoing in collaboration with the World Glacier Monitoring Service. With the initiation of the Glacier Thickness Database (Gärtner-Roer et al., 2014; WGMS, 2016),
580 the first steps towards a freely accessible, global database of ice thickness measurements have been undertaken. We anticipate that this effort, together with a second phase of ITMIX targeting at how to best integrate sparse thickness measurements to improve model performance, will foster the development of improved ice thickness estimation approaches.

Author contributions. DF and HL designed ITMIX. DF coordinated the experiment, performed the model evaluations, prepared the figures, and wrote the paper with contributions by DB, GKCC, JF, PG, FG-C, MH, PWL, AL, HM, FM, MM, AR, RR, OS, and WJJvP. DB,
585 GKCC, JJF, HF, PG, FG-C, CG, MH, PWL, AL, HM, CMa, FM, MM, CMo, APa, APo, AR, RR, OS, SSK, and WJJvP participated to the experiment providing modelling results. BA, LMA, TB, DB, JAD, DF, AF, KH, SK, IL, HL, RMcN, and PAS provided the data necessary for the real-world test cases. CMa and DF generated the synthetic test cases. MH, LMA and GHG provided advice during experimental setup and evaluation.

Acknowledgements. ITMIX was made possible by the International Association of Cryospheric Sciences (IACS). We are indebted to the
590 European Space Agency's project *Glaciers_cci*, as well as to Elisa Bjerre, Emiliano Cimoli, Gwenn Flowers, Marco Marcer, Geir Moholdt, Johnny Sanders, Marius Schäfer, Konrad Schindler, Tazio Strozzi, and Christoph Vogel for providing data necessary for the experiment. DF acknowledges support from the Swiss National Science Foundation (SNSF). MM and CG were funded by the National Aeronautics and Space Administration, Cryospheric Sciences Program, grant NNX15AD55G. PWL was funded by the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement No. 320816. JAD acknowledges the
595 UK Natural Environment Research Council for supporting the airborne radio-echo sounding surveys of the three Arctic ice caps included in the experiment (grants GR3/4663, GR3/9958, GR3/12469 and NE/K004999/1). AR and OS acknowledge the contribution of Labex OSUG@2020 (Investissements d'avenir - ANR10 LABX56) and the VIP_Mont-Blanc project (ANR-14-CE03-0006-03). FG-C and CMo acknowledge support from French National Research Agency (ANR) under the SUMER (Blanc SIMI 6) 2012 project ANR-12BS06-0018.



References

- 600 Anderson, B., Mackintosh, A., Stumm, D., George, L., Kerr, T., Winter-Billington, A., and Fitzsimons, S.: Climate sensitivity of a high-precipitation glacier in New Zealand, *Journal of Glaciology*, 56, 114–128, doi: 10.3189/002214310791190929, 2010.
- Anderton, P.: *Tasman Glacier 1971-73*, Hydrological Research: Annual Report 33., Published by the Ministry of Works and Development for the National Water and Soil Conservation Organization of New Zealand, 1975.
- Andreassen, L., Paul, F., Kääb, A., and Hausberg, J.: Landsat-derived glacier inventory for Jotunheimen, Norway, and deduced glacier
605 changes since the 1930s, *The Cryosphere*, 2, 131–145, doi: 10.5194/tc-2-131-2008, 2008.
- Andreassen, L., Huss, M., Melvold, K., Elvehøy, H., and Winsvold, S.: Ice thickness measurements and volume estimates for glaciers in Norway, *Journal of Glaciology*, 61, 763–775, doi: 10.3189/2015JoG14J161, 2015.
- Andreassen, L. M., Elvehøy, H., Kjølmoen, B., and Engeset, R. V.: Reanalysis of long-term series of glaciological and geodetic mass balance for 10 Norwegian glaciers, *The Cryosphere*, 10, 535–552, doi: 10.5194/tc-10-535-2016, 2016.
- 610 Arendt, A., , and 87 others: Randolph Glacier Inventory [5.0]: A Dataset of Global Glacier Outlines, Version 5.0, Global Land Ice Measurements from Space (GLIMS), Boulder, Colorado, USA, digital Media, 2015.
- Bahr, D., Meier, M., and Peckham, S.: The physical basis of glacier volume-area scaling, *Journal of Geophysical Research*, 102, 20355–20362, doi: 10.1029/97JB01696, 1997.
- Bahr, D., Pfeffer, W., and Kaser, G.: A review of volume-area scaling of glaciers, *Reviews of Geophysics*, 53, 95–140, doi:
615 10.1002/2014RG000470, 2015.
- Bauder, A., Funk, M., and Gudmundsson, G. H.: The ice thickness distribution of Unteraargletscher (Switzerland), *Annals of Glaciology*, 37, 331–336, doi:10.3189/172756403781815852, 2003.
- Brinkerhoff, D. and Johnson, J.: Data assimilation and prognostic whole ice sheet modelling with the variationally derived, higher order, open source, and fully parallel ice sheet model VarGlaS, *The Cryosphere*, 7, 1161–1184, doi: 10.5194/tc-7-1161-2013, 2013.
- 620 Brinkerhoff, D. and Johnson, J.: Dynamics of thermally induced ice streams simulated with a higher-order flow model, *Journal of Geophysical Research: Earth Surface*, 120, 1743–1770, doi: 10.1002/2015JF003499, 2015.
- Brinkerhoff, D., Aschwanden, A., and Truffer, M.: Bayesian inference of subglacial topography using mass conservation, *Frontiers in Earth Science*, 4, 1–15, doi: 10.3389/feart.2016.00008, 2016.
- Budd, W. and Allison, I.: An empirical scheme for estimating the dynamics of unmeasured glaciers, in: *Proceedings of the Moscow Symposium "Snow and Ice"*, pp. 246–256, International Association of Hydrological Sciences, publication No. 104, 1975.
- 625 Chen, J. and Ohmura, A.: Estimation of Alpine glacier water resources and their change since the 1870s, in: *Hydrology in Mountainous Regions*, edited by Lang, H. and Musy, A., pp. 127–135, IAHS Publ. No. 193., proceedings of two Lausanne symposia, 1990.
- Clarke, G., Berthier, E., Schoof, C., and Jarosch, A.: Neural networks applied to estimating subglacial topography and glacier volume, *Journal of Climate*, 22, 2146–2160, doi: 10.1175/2008JCLI2572.1, 2009.
- 630 Clarke, G., Anslow, F., Jarosch, A., Menounos, B., Bolch, T., and Berthier, E.: Ice volume and subglacial topography for western Canadian glaciers from mass balance fields, thinning rates, and a bed stress model, *Journal of Climate*, 26, 4282–430, doi: 10.1175/JCLI-D-12-00513.1, 2013.
- Columbus, J., Sirguey, P., and Tenzer, R.: A free, fully assessed 15-m DEM for New Zealand, *Survey Quarterly*, 66, 16–19, 2011.
- Cook, A., Murray, T., Luckman, A., Vaughan, D., and Barrand, N.: A new 100-m Digital Elevation Model of the Antarctic Peninsula derived
635 from ASTER Global DEM, *Earth System Science Data*, 4, 129–142, doi: 10.5194/essd-4-129-2012, 2012.



- Cuffey, K. and Paterson, W.: *The Physics of Glaciers*, Elsevier, Oxford, U.K., fourth edn., 2010.
- Dowdeswell, J., Drewry, D., Cooper, A., Gorman, M., Liestøl, O., and Orheim, O.: Digital mapping of the Nordaustlandet ice caps from airborne geophysical investigations, *Annals of Glaciology*, 8, 51–58, 1986.
- Dowdeswell, J. A., Bassford, R. P., Gorman, M. R., Williams, M., Glazovsky, A. F., Macheret, Y. Y., Shepherd, A. P., Vasilenko, Y. V., Savatyuquin, L. M., Hubberten, H.-W., and Miller, H.: Form and flow of the Academy of Sciences Ice Cap, Severnaya Zemlya, Russian High Arctic, *Journal of Geophysical Research: Solid Earth*, 107, EPM 5–1–EPM 5–15, doi: 10.1029/2000JB000129, 2002.
- Dowdeswell, J. A., Benham, T. J., Gorman, M. R., Burgess, D., and Sharp, M. J.: Form and flow of the Devon Island Ice Cap, Canadian Arctic, *Journal of Geophysical Research: Earth Surface*, 109, F02 002, doi: 10.1029/2003JF000095, 2004.
- Dowdeswell, J. A., Benham, T. J., Strozzi, T., and Hagen, J. O.: Iceberg calving flux and mass balance of the Austfonna ice cap on Nordaustlandet, Svalbard, *Journal of Geophysical Research: Earth Surface*, 113, F03 022, doi: 10.1029/2007JF000905, 2008.
- Farinotti, D., Huss, M., Bauder, A., and Funk, M.: An estimate of the glacier ice volume in the Swiss Alps, *Global and Planetary Change*, 68, 225–231, doi: 10.1016/j.gloplacha.2009.05.004, 2009a.
- Farinotti, D., Huss, M., Bauder, A., Funk, M., and Truffer, M.: A method to estimate ice volume and ice thickness distribution of alpine glaciers, *Journal of Glaciology*, 55, 422–430, doi: 10.3189/002214309788816759, 2009b.
- Farinotti, D., Corr, H., and Gudmundsson, G.: The ice thickness distribution of Flask Glacier, Antarctic Peninsula, determined by combining radio-echo soundings, surface velocity data, and flow modelling, *Annals of Glaciology*, 54, 18–24, doi: 10.3189/2013AoG63A603, 2013.
- Farinotti, D., King, E., Albrecht, A., Huss, M., and Gudmundsson, G.: The bedrock topography of Starbuck Glacier, Antarctic Peninsula, as measured by radio-echo soundings, *Annals of Glaciology*, 55, 22–28, doi: 10.3189/2014AoG67A025, 2014.
- Fastook, J., Brecher, H., and Hughes, T.: Derived bedrock elevations, strain rates and stresses from measured surface elevations and velocities: Jakobshavn Isbrae, Greenland, *Journal of Glaciology*, 41, 161–173, doi: 10.3198/1995JoG41-137-161-173, 1995.
- Fischer, A. and Kuhn, M.: Ground-penetrating radar measurements of 64 Austrian glaciers between 1995 and 2010, *Annals of Glaciology*, 54, 179–188, doi: 10.3189/2013AoG64A108, 2013.
- Fischer, A., Markl, G., Schneider, H., Abermann, J., and Kuhn, M.: Glacier mass balances and elevation zones of Kesselwandferner, Ötztal Alps, Austria, 1952/1953 to 2010/2011, *Pangaea*, doi: 10.1594/PANGAEA.818757, 2014.
- Fischer, A., Seiser, B., Stocker Waldhuber, M., Mitterer, C., and Abermann, J.: Tracing glacier changes in Austria from the Little Ice Age to the present using a lidar-based high-resolution glacier inventory in Austria, *The Cryosphere*, 9, 753–766, doi: 10.5194/tc-9-753-2015, 2015.
- Flowers, G. E., Roux, N., Pimentel, S., and Schoof, C. G.: Present dynamics and future prognosis of a slowly surging glacier, *The Cryosphere*, 5, 299–313, doi: 10.5194/tc-5-299-2011, 2011.
- Frey, H., Machguth, H., Huss, M., Huggel, C., Bajracharya, S., Bolch, T., Kulkarni, A., Linsbauer, A., Salzmann, N., and Stoffel, M.: Estimating the volume of glaciers in the Himalayan–Karakoram region using different methods, *The Cryosphere*, 8, 2313–2333, doi: 10.5194/tc-8-2313-2014, 2014.
- Gabbi, J., Farinotti, D., Bauder, A., and Maurer, H.: Ice volume distribution and implications on runoff projections in a glacierized catchment, *Hydrology and Earth System Sciences*, 16, 4543–4556, doi: 10.5194/hess-16-4543-2012, 2012.
- Gagliardini, O., Zwinger, T., Gillet-Chaulet, F., Durand, G., Favier, L., de Fleurian, B., Greve, R., Malinen, M., Martin, C., Raback, P., Ruokolainen, J., Sacchetti, M., Schafer, M., Seddik, H., and Thies, J.: Capabilities and performance of Elmer/Ice, a new-generation ice sheet model, *Geoscientific Model Development*, 6, 1299–1318, doi: 10.5194/gmd-6-1299-2013, 2013.



- Gantayat, P., Kulkarni, A., and Srinivasan, J.: Estimation of ice thickness using surface velocities and slope: case study at Gangotri Glacier, India, *Journal of Glaciology*, 60, 277–282, doi: 10.3189/2014JoG13J078, 2014.
- 675 Gärtner-Roer, I., Naegeli, K., Huss, M., Knecht, T., Machguth, H., and Zemp, M.: A database of worldwide glacier thickness observations, *Global and Planetary Change*, 122, 330–344, doi: 10.1016/j.gloplacha.2014.09.003, 2014.
- Geostudios LTA: Estimación de volúmenes de hielo mediante sondeos de radar en zonas norte, central y sur [Estimation of ice volumes using radar soundings in the north, central and south region], S.I.T. No. 338, 2014.
- Glen, J.: The creep of polycrystalline ice, *Proceedings of the Royal Society of London, Series A*, 228, 519–538, 1955.
- 680 Grinsted, A.: An estimate of global glacier volume, *The Cryosphere*, 7, 141–151, doi: 10.5194/tc-7-141-2013, 2013.
- Gudmundsson, G. H., Thorsteinsson, T., and Raymond, C. F.: Inferring bed topography and stickiness from surface data on ice streams, *Eos Trans. AGU*, 82, fall Meet. Suppl., Abstract IP21A-0687, 2001.
- Haerberli, W. and Hoelzle, M.: Application of inventory data for estimating characteristics of and regional climate-change effects on mountain glaciers: a pilot study with the European Alps, *Annals of Glaciology*, 21, 206–212, doi: 10.3198/1995AoG21-1-206-212, 1995.
- 685 Haq, M., Jain, K., and Menon, K.: Modelling of Gangotri glacier thickness and volume using an artificial neural network, *International Journal of Remote Sensing*, 35, 6035–6042, doi: 10.1080/01431161.2014.943322, 2014.
- Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, *International Journal of Climatology*, 34, 623–642, doi 10.1002/joc.3711, 2014.
- Hastings, W.: Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika*, 57, 97–109, doi: 10.1093/biomet/57.1.97, 1970.
- 690 Huss, M. and Farinotti, D.: Distributed ice thickness and volume of all glaciers around the globe, *Journal of Geophysical Research*, 117, F04010, doi: 10.1029/2012JF002523, 2012.
- Huss, M. and Farinotti, D.: A high-resolution bedrock map for the Antarctic Peninsula, *The Cryosphere*, 8, 1261–1273, doi: 10.5194/tc-8-1261-2014, 2014.
- 695 Hutchinson, M.: A new procedure for gridding elevation and stream line data with automatic removal of spurious pits, *Journal of Hydrology*, 106, 211–232, doi: 10.1016/0022-1694(89)90073-5, 1989.
- Hutter, K.: *Theoretical glaciology; material science of ice and the mechanics of glaciers and ice sheets*, D. Reidel Publishing Company, Tokyo, Terra Scientific Publishing Company, 1983.
- Huybrechts, P.: *The Antarctic ice sheet and environmental change: a three-dimensional modelling study*, Ph.D. thesis, Vrije Universiteit of 700 Brussel, doi: 10013/epic.12054.d001, 1991.
- Hynek, B., Hillerup Larsen, S., Binder, D., Weyss, G., Citterio, M., Schöner, W., and Ahlstrøm, A.: In-situ glacier monitoring in Zackenberg (NE Greenland): Freya Glacier and A.P. Olsen Ice Cap, in: *EGU General Assembly Conference Abstracts*, vol. 17, pp. EGU2015–15477, 2015.
- Kienholz, C., Rich, J., Arendt, A., and Hock, R.: A new method for deriving glacier centerlines applied to glaciers in Alaska and northwest 705 Canada, *The Cryosphere*, 8, 503–519, doi: 10.5194/tc-8-503-2014, 2014.
- Leclercq, P., Pitte, P., Giesen, R., Masiokas, M., and Oerlemans, J.: Modelling and climatic interpretation of the length fluctuations of Glacier Frias (north Patagonian Andes, Argentina) 1639-2009 AD, *Climate of the Past*, 8, 1385–1402, doi: 10.5194/cp-8-1385-2012, 2012.
- Li, H., Ng, F., Li, Z., Qin, D., and Cheng, G.: An extended "perfect-plasticity" method for estimating ice thickness along the flow line of mountain glaciers, *Journal of Geophysical Research: Earth Surface*, 117, F01020, doi: 10.1029/2011JF002104, 2012.



- 710 Linsbauer, A., Paul, F., Hoelzle, M., Frey, H., and Haerberli, W.: The Swiss Alps without glaciers - A GIS-based modelling approach for reconstruction of glacier beds, in: Proceedings of Geomorphometry 2009, pp. 243–247, Zurich, available from www.geomorphometry.org/Linsbauer2009, 2009.
- Linsbauer, A., Paul, F., and Haerberli, W.: Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach., *Journal of Geophysical Research*, 117, F03 007, doi: 10.1029/2011JF002313, 2012.
- 715 LINZ: New Zealand Mapping Series 260 (NZMS 260) Map Sheets, CC-By Land Information New Zealand (LINZ), Wellington, New Zealand, Data retrievable at <https://data.linz.govt.nz/layer/1579-nzms-260-map-sheets/>, last accessed on 03.05.2016, 2013.
- Lüthi, M.: Transient response of idealized glaciers to climate variations, *Journal of Glaciology*, 55, 918–930, doi: 10.3189/002214309790152519, 2009.
- Maohuan, H., Zhongxiang, W., Baolin, C., and Jiankang, H.: Some dynamics studies on Urumqi Glacier No. 1, Tianshan Glaciological Station, China, *Annals of Glaciology*, 12, 70–73, 1989.
- 720 Marcer, M., Stentoft, P., Bjerre, E., Cimoli, E., Bjørk, A., Stenseng, L., and Machguth, H.: Three decades of mass change on a small Greenlandic glacier using ground penetrating radar, structure from motion and aerial photogrammetry, *Arctic Antarctic and Alpine Research*, doi:, in review.
- Marzeion, B., Jarosch, A., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, *The Cryosphere*, 6, 1295–1322, doi:10.5194/tc-6-1295-2012, 2012.
- 725 McNabb, R., Hock, R., O’Neel, S., Rasmussen, L., Ahn, Y., Braun, M., Conway, H., Herreid, S., Joughin, I., Pfeffer, W., Smith, B., and Truffer, M.: Using surface velocities to calculate ice thickness and bed topography: A case study at Columbia Glacier, Alaska, *Journal of Glaciology*, 58, 1151–1164, doi: 10.3189/2012JoG11J249, 2012.
- Michel, L., Picasso, M., Farinotti, D., Funk, M., and Blatter, H.: Estimating the ice thickness of mountain glaciers with an inverse approach using surface topography and mass-balance, *Inverse Problems*, 29, 035 002, doi: 10.1088/0266-5611/29/3/035002, 2013.
- 730 Michel, L., Picasso, M., Farinotti, D., Funk, M., and Blatter, H.: Estimating the ice thickness of shallow glaciers from surface topography and mass-balance data with a shape optimization algorithm, *Computers & Geosciences*, 66, 182–199, doi: 10.1016/j.cageo.2014.01.012, 2014.
- Moholdt, G. and Käab, A.: A new DEM of the Austfonna ice cap by combining differential SAR interferometry with icesat laser altimetry, *Polar Research*, 31, Art. nr. 18 460, doi: 10.3402/polar.v31i0.18460, 2012.
- 735 Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Dhia, H. B., and Aubry, D.: A mass conservation approach for mapping glacier ice thickness, *Geophysical Research Letters*, 38, L19 503, doi: 10.1029/2011GL048659, 2011.
- Morlighem, M., Rignot, E., Mouginot, J., Seroussi, H., and Larour, E.: Deeply incised submarine glacial valleys beneath the Greenland Ice Sheet, *Nature Geoscience*, 7, 418–422, doi: 10.1038/ngeo2167, 2014.
- 740 Mosbeux, C., Gillet-Chaulet, F., and Gagliardini, O.: Comparison of adjoint and nudging methods to initialise ice sheet model basal conditions, *Geoscientific Model Development*, 9, 2549–2562, doi: 10.5194/gmd-9-2549-2016, 2016.
- Nye, J.: The mechanics of glacier flow, *Journal of Glaciology*, 2, 82–93, doi: 10.3198/1952JoG2-12-82-93, 1952.
- Nye, J.: The flow of a glacier in a channel of rectangular, elliptic or parabolic cross-section, *Journal of Glaciology*, 5, 661–690, 1965.
- 745 Paul, F. and Linsbauer, A.: Modeling of glacier bed topography from glacier outlines, central branch lines, and a DEM, *International Journal of Geographic Information Sciences*, 26, 1173–1190, doi: 10.1080/13658816.2011.627859, 2011.
- Plewes, L. and Hubbard, B.: A review of the use of radio-echo sounding in glaciology, *Progress in Physical Geography*, 25, 203–236, doi: 10.1177/030913330102500203, 2001.



- Radić, V. and Hock, R.: Regional and global volumes of glaciers derived from statistical upscaling of glacier inventory data, *Journal of Geophysical Research*, 115, F01 010, doi: 10.1029/2009JF001373, 2010.
- 750 Radić, V. and Hock, R.: Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise, *Nature Geoscience*, 4, 91–94, doi: 10.1038/ngeo1052, 2011.
- Rasmussen, L.: Bed topography and mass-balance distribution of Columbia Glacier, Alaska, U.S.A., Determined from sequential aerial photography, *Journal of Glaciology*, 34, 208–216, doi: 10.3198/1988JoG34-117-208-216, 1988.
- Rau, F., Mauz, F., Vogt, S., Khalsa, S. J. S., and Raup, B.: Illustrated GLIMS glacier classification manual, Institut für Physische Geographie
755 Freiburg, Germany, and National Snow and Ice Data Center, Boulder, USA, version 1.0, 2005.
- Raymond-Pralong, M. and Gudmundsson, G.: Bayesian estimation of basal conditions on Rutford Ice Stream, West Antarctica, from surface data, *Journal of Glaciology*, 57, 315–324, doi: 10.3189/002214311796406004, 2011.
- Reerink, T., Kliphuis, M., and van de Wal, R.: Mapping technique of climate fields between GCM's and ice models, *Geoscientific Model Development*, 3, 13–41, doi: 10.5194/gmd-3-13-2010, 2010.
- 760 Sanders, J. W., Cuffey, K. M., Macgregor, K. R., Kavanaugh, J. L., and Dow, C. F.: Dynamics of an alpine cirque glacier, *American Journal of Science*, 310, 753–773, doi: 10.2475/08.2010.03, 2010.
- Thorsteinsson, T., Raymond, C. F., Gudmundsson, G. H., Bindschadler, R. A., Vornberger, P., and Joughin, I.: Bed topography and lubrication inferred from surface measurements on fast-flowing ice streams, *Journal of Glaciology*, 49, doi: 10.3189/172756503781830502, 2003.
- van Pelt, W., Oerlemans, J., Reijmer, C., Pettersson, R., Pohjola, V., Isaksson, E., and Divine, D.: An iterative inverse method to estimate
765 basal topography and initialize ice flow models, *The Cryosphere*, 7, 987–1006, doi: 10.5194/tc-7-987-2013, 2013.
- Vogel, C., Bauder, A., and Schindler, K.: Optical flow for glacier motion estimation, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-3, 359–364, doi: 10.5194/isprsannals-I-3-359-2012, 2012.
- Wang, P., Li, Z., Li, H., Yao, H., Xu, C., Zhou, P., Jin, S., and Wang, W.: Analyses of recent observations of Urumqi Glacier No. 1, Chinese Tianshan Mountains, *Environmental Earth Sciences*, 75, 1–11, doi: 10.1007/s12665-016-5551-3, 2016.
- 770 WGMS: Glacier mass balance bulletins - Bulletins No. 1-12 (1988-1989, 2010-2011), ICSU (WDS) / IUGG (IACS) / UNEP / UNESCO / WMO, World Glacier Monitoring Service, Zurich, Switzerland, available online at http://wgms.ch/products_gmbb/. Publication based on database version: doi:10.5904/wgms-fog-2013-11, 1991-2012.
- WGMS: Glacier Thickness Database 2.0, I. Gärtner-Roer, L.M. Andreassen, E. Bjerre, D. Farinotti, A. Fischer, M. Fischer, K. Helfricht, M. Huss, T. Knecht, S. Kutuzov, J. Landmann, I. Lavrentiev, H. Li, Z. Li, H. Machguth, K. Naegeli, F. Navarro, A. Rabatel, P. Stentoft, M. Zemp (eds.), World Glacier Monitoring Service, Zurich, Switzerland, doi: 10.5904/wgms-glathida-2016-07, 2016.
- 775 Wheler, B., MacDougall, A., Flowers, G., Petersen, E., Whitfield, P., and Kohfeld, K.: Effects of temperature forcing provenance and extrapolation on the performance of an empirical glacier-melt model, *Arctic, Antarctic, and Alpine Research*, 46, 379–393, doi:10.1657/1938-4246-46.2.379, 2014.
- Willis, I., Lawson, W., Owens, I., Jacobel, B., and Autridge, J.: Subglacial drainage system structure and morphology of Brewster Glacier, New Zealand, *Hydrological Processes*, 23, 384–396, doi: 10.1002/hyp.7146, 2009.
- 780 Wilson, N., Flowers, G., and Mingo, L.: Comparison of thermal structure and evolution between neighboring subarctic glaciers, *Journal of Geophysical Research: Earth Surface*, 118, 1443–1459, doi: 10.1002/jgrf.20096, 2013.
- Zolotarev, E. and Kharkovets, E.: Elbrus glaciation at the end of XX century - Digital orthophotomap of Elbrus 1997, *Data of Glaciological studies*, 89, 175–181, in Russian, 2000.