Physically-constrained data-driven inversions to infer the bed topography beneath glaciers flows. Application to East Antarctica

Jerome Monnier, Jiamin Zhu

▶ To cite this version:

Jerome Monnier, Jiamin Zhu. Physically-constrained data-driven inversions to infer the bed topography beneath glaciers flows. Application to East Antarctica. 2020. hal-01926620v2

HAL Id: hal-01926620

https://hal.science/hal-01926620v2

Preprint submitted on 28 Feb 2020 (v2), last revised 7 Jun 2021 (v4)

Noname manuscript No.
(will be inserted by the editor)

Physically-constrained data-driven inversions to infer the bed topography beneath glaciers flows. Application to East Antarctica

Jérôme Monnier · **Jiamin Zhu**

Received: date / Accepted: date

Abstract A method to infer the bed topography beneath glaciers from surface measurements (e.g. elevation from altimetry and velocity from InSAR) plus sparse thickness measurements is developed and assessed. The method relies on an original non-isothermal Reduced Uncertainty (RU) version of the Shallow Ice Approximation (SIA) equation which natively integrates the surface measurements. The flow model presents a single dimensionless multi-physics parameter γ; it takes into account basal slipperiness and varying vertical rate factor profiles (thus vertical thermal variations). Next the inversions rely on a statistical learning model (deep neural network) based on the in-situ thickness measurements combined with two Variational Data Assimilation (VDA) processes. The inversion method is demonstrated to be robust. It is valid for moderately sheared flows (presenting a moderate basal slipperiness), thus it can be applied to inland ice-sheets areas. Numerical results are presented in East Antarctica Ice Sheet areas where the bed elevation may be highly uncertain (Bedmap2 values). The estimations are valid for wave lengths greater than $\sim 10\bar{h}$ (due to the long wave assumption, shallow flow model) with a resolution at $\sim \bar{h}$ ($\bar{h}$ a characteristic thickness value).

Keywords Variational data assimilation · reduced flow model · deep learning · topography · inference · glaciers · Antarctica.

1 Introduction

Bed topography elevation is a fundamental data to set up dynamic ice flow models; moreover if combined with the surface topography measurements (e.g. acquired by altimetry in ice-sheets) it straightforwardly provides the ice volume. In Antarctica and Greenland ice-sheets, ice thickness measurements are available along airborne radio-echo sounding tracks see e.g. CReSIS RDS database [1]. The measurements are particularly dense in fast ice costal stream areas. On the contrary they are very sparse inland, they are even nonexistent deep inland, see [4,14] and references therein. In other respects, numerous satellites have provided accurate measurements of ice sheets surfaces: altimeters provide surface elevation H at $\approx \pm 20$ cm for 1 km^2 pixels, see e.g. [22,4], radar interferometers (InSar) provide surface velocity $\mathbf{u}_H$ see e.g. [43]. Next the challenge is to infer the bed elevation values between the thickness measurements; this is done by using a model.
The most simple model one can employ is the ordinary Kriging; that is a Gaussian process regression algorithm providing the best linear unbiased prediction of the intermediate values. It is statistical interpolation, a purely data-driven model (the most employed method in geo-statistics). It has enabled to provide bed elevation estimations e.g. inland Greenland and Antarctica, [3,14]. In Antarctica, the authors of [14] suggest that for cells located less than 20km from a measurement the estimation error is an increasing function of the distance; beyond 20 km the error would be uncorrelated with distance. For cells located more than 50 km from any measurement, the thickness estimation is based on the gravity-field inversion therefore presenting very large uncertainties ($\pm \sim 1000$m according to [14]).
To reduce uncertainties on the bed topography estimations, combining physical-based models with datasets is definitively a good direction. Then a challenge is to employ a flow model sufficiently complex to be accurate but simple enough to lead to well-posed (and stable) inverse problems in the sense not leading to severe equifinality issues. In inverse geophysical flows modelling by data assimilation, equifinality issues are the common pitfall to avoid, see e.g. [5]. In other respect the available large glaciers depth datasets invite to employ deep neural networks.

J. Monnier
INSA & Institut de Mathématiques de Toulouse (IMT), France
E-mail: jerome.monnier@insa-toulouse.fr

J. Zhu
INSA & Institut de Mathématiques de Toulouse (IMT), France

---

[1] CReSIS RDS Data http://data.cresis.ku.edu (NASA, NSF, Kansas fundings).

In fast ice streams, the flows may be considered as plug flows (pure sliding) then the inversion of the depth-integrated mass equation enables to fill up the gaps downstream (and upstream) the measurements, see [46,40]. The measurements locations (e.g. the flight tracks during airborne campaigns) have to be cross-lines and relatively dense. Indeed it is well known that this transport equation is intrinsically unstable to invert and it propagates errors, see e.g. [36] and references therein. To locally damp this feature, artificial diffusion regularising the equation may be introduced. The inversion of mass conservation proposed in [8,46], next combined with surface measurements by Variational Data Assimilation (VDA) in [41,42], has enabled to estimate the bed elevation under ice streams along flow lines in Greenland, see [40, 42]. In [7] a Bayesian framework is applied to provide probability distributions of thickness assuming Gaussian covariance structures of input data (the algorithm being equivalent to ordinary Kriging if no data is available). This Bayesian approach is particularly suitable however in this case the estimations are based on mass conservation only.

For fully sheared flows, the isothermal SIA flow model (with no slip at bottom) has been inverted e.g. in [37,21], providing robust estimations but relevant in a very restricted flow regime only. In moderately sliding flows (equivalently moderately sheared flows), the slipperiness at bottom has to be taken into account in the bed inversion therefore transforming the inverse problem to a severely ill-posed one. Indeed in this case, the flow models present the basal slipperiness in addition to the rate factor (internal deformation). [25] inverts the 1D depth-integrated SIA equation with slipperiness and a shape factor (modelling the 3D features of the flow) leading to an ill-posed inverse problem (the inversions are performed by imposing empirical constant values for the few unknown parameters). Based on the inversion of the complete hybrid SIA-SSA system PISM [55], [51] proposes an empirical iterative method to fit with the surface elevation by calibrating the bed elevation only. In [38] the SIA flow model with slipperiness at bottom is inverted by distinguishing different regimes, enabling to define well-posed inverse problems, hence providing stable and robust inversions. However the rate factor (depending in particular on the temperature field) is supposed to be constant; unfortunately this assumption is often unrealistic, in particular in ice-sheets glaciers.

Inland ice-sheets areas presenting mid-range surface velocities ($\approx [10-80]$ m/y in Antarctica), the flows are moderately slipping / moderately sheared; they cannot accurately be modelled using plug flows models or fully sheared flows models. Indeed these areas, the measured surface features (elevation and velocity) are the signature of both the slipperiness at bottom and internal deformation; the latter depending on the constitutive ice behaviour and the vertical thermal profiles. As a consequence inverting the surface data in moderately sheared flows is highly challenging.

A comparison of various inverse methods to estimate the bed elevation beneath glaciers (equivalently the ice thickness) is presented in [12]. The comparison relies on numerous test cases representing a large spectrum of ice flow regimes (but no ice-sheet is considered). One feature of the considered benchmarks is that no prior thickness value is supposed to be known. The 15 inter-compared methods are classified by resolution type and not by domain of validity (e.g. in function of the flow regime). Numerical comparisons are presented; no analysis of the equifinality issue(s) is proposed (that is well-posed vs ill-posed inverse formulations).

The present study aims at solving the following inverse problem: estimating the ice thickness in moderately sheared flows while taking into account the complete physics of the flow. To do so, a key ingredient is the Reduced Uncertainty SIA (RU-SIA) model first derived in [39]. This flow model is dedicated to the present inverse problem by natively integrating the surface measurements. It is a multi-physics depth-integrated shallow flow model while respecting a "well-balanced complexity" to be inverted. Indeed it is a 2D shallow flow model (and not 3D with a mobile surface) but relatively complete since mass and momentum equations are considered. Moreover the RU-SIA equation takes into account the non uniform internal properties (in particular due to the varying vertical temperature profiles) while it presents a single uncertain dimensionless parameter, see [38,39].

Inland moderately sheared flows have been poorly covered during the airborne campaigns (they are difficult and expansive to flight over). Therefore the chosen flow model needs to be stable and robust when inverted even in lack of local in-situ data. This mathematical feature is as important as the model consistency; this feature is all but trivial to obtain, see e.g. [2,36,48,38]. Moreover it would be highly valuable that the flow model inversion is as insensitive as possible to the measurements locations.

Note that an effective bed topography only can be infered from the surface signature. Indeed the glaciers flows act as low-band filters: the bed variations are filtered by the flow with filtering features depending on the flow regime, see [18, 34,35] for detailed analysis. To a flow regime and a flow model corresponds a minimal inferable wave length [19,34,35].

In the present study, a new inversion method to infer the bed elevation from surface measurements (elevation, velocity) and sparse in-situ thickness values is developed. It is applied to six large inland East Antartica Ice Sheet (EAIS) areas. This inversion method relies on the RU-SIA equation derived in [39], two advanced VDA processes and a deep Neural Network (statistical learning).The latter aims at estimating the single dimensionless parameter γ of the RU-SIA model from the in-situ ice thickness measurements (here available along the flight tracks of airborne campaigns).

The present inversion method may be applied to any glaciers flows in ice-sheets or ice-caps as soon as the flow model assumption is satisfied and the estimation of γ is robust. Then the computed thickness estimations are valid at $\approx 10 \times \bar{h}$ wave length, $\bar{h}$ a characteristic thickness value.

In [39], the proposed inversion strategy is similar as the present one since based on the two VDA processes and an estimation of γ too. However in [39] γ is estimated by a simple Kriging method. This was possible in the single considered test case since a clear correlation is observed between γ and one of the observable (namely $|\mathbf{u}_H|$, see [39] Fig. 7). Such clear correlation may be not observed in other contexts. Then in the present inversion method, a more general Neural Network Residual Kriging (NNRK) algorithm ([11,30]) is employed. The present assessment experiments of this method (assessment based on cross-validations) performed on numerous large EAIS areas have demonstrated that it is robust and accurate even in the lack of clear correlation between γ and one of the observable.
In other respect the present study details a remarkable relationship between the mass inversion method [46,40,42] enabling to estimate the bed elevation under fast streams, and the present RU-SIA based method. This relationship opens promising perspectives to mix these two complimentary methods, then to obtain smooth estimations between their respective domain of validity.

The present test areas have been selected since they respect the domain of validity of the method; moreover estimating more accurately the bed elevation there is interesting since global warming may threaten EAIS stability, particularly around some of these areas, see e.g. [13].

The outline of the article is as follows. In Section 2, the Reduced Uncertainty (RU) equation developed in [38,39] is recalled; its domain of validity is highlighted; uncertainty range on its uncertain dimensionless parameter γ is presented. Next, the inversion method developed in three steps is detailed: Step 1) and Step 3) aim at inverting the RU-SIA equation by VDA (physical-based inversions); Step 2) aims at estimating γ using a ANNK algorithm (deep learning). In Section 4, the considered six large EAIS areas (named Ant$p$, $p = 1,..,6$) are presented; basis of the Bedmap2 estimations [14] are recalled. Each computational steps of the inversion method is analysed in detail. In Section 5, the robustness of the obtained estimations is assessed for Ant1 and Ant3 cases; in particular their sensitivities with respect to the presence or not of additional flight tracks. A conclusion is proposed in Section 6. As a complimentary material, the computed thickness estimations for the four other areas Ant$p$, $p = 2,4,..,6$, are presented in Appendix.

2 Method

In this section the inversion method to estimate the ice thickness h (equivalently the bed topography elevation b) is detailed. It is done in three steps. Steps 1) and 3) aim at assimilating all surface data (altimetry, InSAR and climatic term Source Mass Balance) plus the in-situ thickness measurements in the RU-SIA flow model. Step 2) aims at estimating the single dimensionless parameter γ from the in-situ measurements only (measurements available along the flight tracks of airborne campaigns). The final output of the inversion method is the ice thickness h, therefore the bed topography elevation b.

2.1 The RU-SIA flow model

The RU-SIA equation is obtained by reformulating the depth-integrated SIA model with basal slipperiness (see e.g. [17] Chapter 5) but with a non constant rate factor and by natively integrating the surface measurements (elevation and velocity), see [39]. The resulting 2D depth-integrated flow model is original; it is relevant for large scale sheared flows with moderate slipperiness at bottom and with non constant vertical temperature profile. The various uncertain multi-physics parameters (constitutive law exponent, flow regime, temperature dependent term) are gathered into the single dimensionless parameter γ. As a consequence, all the physical parametrisation uncertainty is represented by this original dimensionless parameter. The basic RU-SIA model assumptions are the same as those for the SIA model (classical lubrication theory) with basal slipperiness that is: the flow is necessarily sheared (normal stress components are negligible) and it is ”shallow” (long wave assumption).

2.1.1 The equations and the dimensionless parameter γ

The surface slope is denoted by $\mathscr{S} = |\nabla H|$; $\mathbf{u}_H$ is the surface velocity and $\mathbf{u}_b$ is the velocity at bottom (basal velocity). The depth (ice thickness) is denoted by h, $h = (H - b)$ with H the ice surface elevation and b the bed elevation, see Fig. 1. θ is a potential mean slope value in the (x, y)-plane ($\theta = 0$ in the forthcoming test areas).

The depth-integrated flow model SIA model with slipperiness at bottom (see e.g. [17] Chapter 5) is derived in a non isothermal case in [38] providing the so-called xSIA (x for extended) equation. Next in [39], by taking advantage of the

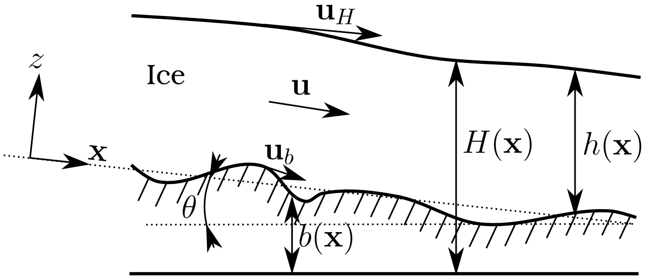

Fig. 1 Schematic vertical view of the gravitational ice flow and notations

measured surface features (elevation and velocity norm), the xSIA model is re-formulated to obtain the RU-SIA model (RU for Reduced Uncertainty). The RU-SIA equation reads as follows:

$$-div\left(\frac{|\mathbf{u}_H|}{\mathscr{S}}\gamma h\nabla H\right)=\dot{a} \tag{1}$$

The RHS $\dot{a}$ is the classical one defined by: $\dot{a}=(\partial_t h-a)$ with a the mass balance (accumulation and ablation). γ is the dimensionless parameter of the model, its expression is:

$$\gamma=\left(1-\frac{c_A R_s}{(q+2)}\right) \tag{2}$$

R_s is the slip ratio describing the flow regime; it is defined as: $R_s=1-\dfrac{|\mathbf{u}_b|}{|\mathbf{u}_H|}$.
The parameter c_A is defined by:

$$c_A=[(q+2)-(q+1)R_A] \tag{3}$$

where q is the constitutive power-law exponent ($q=3$ in the classical Glen's law (see e.g. [17] Chapter 5, [38]) and $R_A=\dfrac{\bar{A}}{\underline{A}}$.
The parameters $\bar{A}(\mathbf{x})$ and $\underline{A}(\mathbf{x})$ are the depth-integrated quantities naturally appearing if the rate factor A depends on $(\mathbf{x},z)$. A highly depends on z as soon as the temperature vertical profile varies. The expressions of these parameters are as follows, see [38]:

$$\bar{A}(\mathbf{x})=\frac{(q+2)}{h^{q+2}(\mathbf{x})}\left(\int_b^H\int_b^z A(\mathbf{x},\xi)(H(\mathbf{x})-\xi)^q d\xi dz\right)^{-1} \tag{4}$$

$$\underline{A}(\mathbf{x})=\frac{(q+1)}{h^{q+1}(\mathbf{x})}\int_b^H A(\mathbf{x},z)(H(\mathbf{x})-z)^q dz \tag{5}$$

If the vertical profile of A is constant then: $\bar{A}(\mathbf{x})=A(\mathbf{x})=\underline{A}(\mathbf{x})\ \forall\mathbf{x}$. Therefore in the isothermal case, A is a constant and: $\bar{A}=A=\underline{A}$; moreover $R_A=1=c_A$. This is the case described in the literature (excepted in [38,39]).
For the classically employed value $q=3$, $\gamma=\left(1-\frac{1}{5}R_s\right)$.
Recall that $R_s=1$ for a fully sheared flow (no slip at bottom) and $R_s=0.5$ for a mid-sheared mid-sliding flow.

In (1), the term $\frac{|\mathbf{u}_H|}{\mathscr{S}}$ is the observational term, e.g. provided by InSAR and altimetry surface measurements. Assuming that h (or equivalently b) is given, Eqn (1) in variable H contains γ as the single uncertain parameter. The linear diffusive equation in H is closed by Dirichlet condition at boundaries: H given e.g. by altimetry.

2.1.2 Domain of validity of the model

The shallowness of the flow is estimated through the geometrical ratio $\varepsilon=\frac{H^*}{L^*}$, where H^* and L^* are characteristic flow depth and length respectively. In these depth-integrated asymptotic models, ε has to be small enough, $\varepsilon\lesssim 1/10$ at least, see e.g. [29]. As a consequence this flow model is valid for a minimal wave length $L^*\gtrsim 10H^*$. The flow regime is characterised by the slip ratio R_s. By construction, the SIA-like models (including xSIA and RU-SIA equations) are valid for R_s ranging from ≈ 0.3 to 1, see [23,47,6] for detailed analysis. This estimation in terms of R_s is numerically quantified in real world cases (including EAIS) in [54]. This study is based on the so-called MCL criteria (criteria proposed in [24] and defined as the length scale over which the terms of driving stress and drag are comparable). In particular it can be noticed that the ice-sheet areas presenting surface velocity ranging in $\approx[5-100]$ m/y are accurately modelled by the SIA model as soon as the minimal wave length equals $\approx 10-12$ km in mean. The six test areas Ant$p$ considered in the next

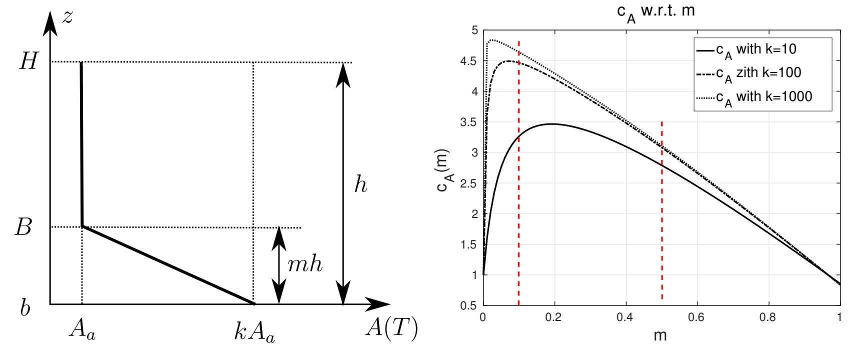

Fig. 2 (Left) A typical vertical profile of rate factor $A(z)$, see (6). (Right) The parameter c_A vs m, see (3), with $k = 10, 100$ and 1000.

sections, see Fig. 4, have been defined from the surface velocities norm values: $|\mathbf{u}_H| \in\approx [10-80]$ m/y. In these cases $H^* \approx 2-3$ km, then the RU-SIA equation is accurate for minimal wave lengths $L^* \approx 20-30$ km.

2.1.3 *Junction with the mass equation & its inversion in fast plug-like flows*

As already mentioned, nice bed elevation estimations are obtained in fast streams by inverting the depth-averaged mass equation $div(h\bar{\mathbf{u}}) = \dot{a}$ with $\bar{\mathbf{u}}$ the depth-averaged velocity, see [42]. In [46,41,42], $\bar{\mathbf{u}}$ is related to $\mathbf{u}_H$ as $\bar{\mathbf{u}} = \tilde{\alpha}\mathbf{u}_H$, with $\tilde{\alpha}$ empirically set. In fast streams (plug like flows), one has $R_s \sim 0$ and $\tilde{\alpha} \lesssim 1$; therefore the uncertainty on the internal deformation (represented in the RU-SIA equation by the parameter c_A) is negligible. As a consequence α may be set close to 1 with a few percent error only.

In [39] it is shown that: $\bar{\mathbf{u}} = -\frac{|\mathbf{u}_H|}{\mathscr{S}}\gamma\nabla H$. Therefore if the slopes $\mathscr{S} = |\nabla H|$ and the velocity are co-linear (this is a commonly admitted assumption) than the parameter $\tilde{\alpha}$ empirically defined in [46,41] is nothing else than the dimensionless parameter γ defined by (2). On the contrary to plug like flows, in moderately sheared / moderately sliding flows, γ varies quite importantly; setting its value empirically is not reasonable anymore. That is why an actual estimation of γ is required.

This equality $\gamma = \tilde{\alpha}$ enables to build up continuous estimations between fast plug-flows (obtained by inverting the mass equation as in [41,42]) and the moderately sheared / moderately sliding flows (obtained by inverting the RU-SIA equation).

2.2 On the uncertainty range of parameter γ

The dimensionless parameter γ defined by (2) depends on various physics parameters: the constitutive law exponent q, the vertical temperature profile through the rate factor $A(z)$ and the flow regime (slip ratio R_s). Recall that the RU-SIA model domain of validity corresponds to $R_s \in [\approx 1./3., 1]$, see [23,47,6]. For a sake of simplicity, q is supposed to be set to the widely employed value for glaciers flows, that is $q = 3$ (Glen's law). In isothermal cases, it follows: $\gamma = (1-0.2R_s)$. Therefore in isothermal cases, the uncertainty on γ is relatively small, $\approx 10\%$ only.

The large majority of glaciers are not isothermal in particular those in ice-sheets. Following the Arrhenius law, see e.g. [17, p.54], and by considering typical ice-sheet vertical temperature profiles in ice-sheets, see [44,45] and e.g. [27,49], we consider the following profile of $A(z)$, Fig. 2 (Left):

$$A(z) = \begin{cases} A_a & \text{for } z \in [B,H] \\ \frac{A_a}{B-b}\left((1-k)z + kB - b\right) & \text{for } z \in [b,B] \end{cases} \tag{6}$$

with A_a and k given constants. We define: $B = (mh+b)$ with $m \in [0,1]$. Then $A(z)$ presents a boundary layer at bottom of thickness $(B-b) = mh$, Fig. 2 (Left). Following [39], the value of c_A vs m for different values of k is presented in Fig. 2 (Right). The case $m = 0$ corresponds to the isothermal case: $c_A = 1$. For thin thermal boundary layers c_A increases with mh; for ticker layers c_A decreases to a minimal value $c_A^{(min)} \lesssim 1$. (This minimal value is reached for the purely linear vertical profile: $m = 1$).

Let us consider typical temperature values in EAIS: the bed is at $0C^\circ$ and the surface at $-40C^\circ$. These values correspond to $A_a \approx 10^{-26}$ therefore $k \approx 1000$, see Fig. 2 (Right). (The value $k \approx 10$ would correspond to typical inland Greenland cases, see [39]). Assuming a boundary layer corresponding to $m \in [0.1, 0.5]$, it follows: $c_A \in (3.11, 4.64)$ (Fig. 2). Finally it follows from (2) that: $\gamma \approx [1-(0.78\pm 0.15)R_s]$.

This rough uncertainty analysis based on typical values in the targeted areas shows that the uncertainty on γ comes similarly from the vertical thermal profile uncertainty (represented by the term $(c_A/(q+2))$) and the slip ratio R_s. In the targeted regimes and EAIS areas with the vertical profile (6), this corresponds to γ varying within the interval $]0, \approx 0.7]$. If relaxing the assumption on the vertical profile as defined in (6), one may estimate the upper bound of γ by setting $c_A = 1$ and $R_s = 0.5$ which gives: $\gamma \in]0, \approx 0.9]$. (Recall that in fast plug-like flows, γ is close to 1).
In the forthcoming numerical results, the estimations of γ by NNRK algorithm are within intervals $]0, \approx 0.9]$, see e.g. figs. 6 and 9(Up)(Right).

2.3 The inversion method

The inversion method to estimate the ice thickness h is developed in three steps. Step 1) and Step 3) are physical-based inversions: the RU-SIA equation (1) is inverted by VDA. Step 2) is based on ANN algorithm; it is a purely data-driven inversion. Step 1) and Step 3) are similar than those proposed in [39], while Step 2) is different.

2.3.1 Sketch of the complete inversion method

The estimations are performed in three steps as follows.

`Step 1)` Estimation of the effective diffusivity $\eta = (\gamma h)$ in RU-SIA equation (1) by VDA.
Given the surface measurements H and $|\mathbf{u}_H|$, the effective diffusivity $\eta = (\gamma h)$ in (1) is infered by solving the following optimal control problem:

$$\min_{\mathbf{k}} g(\mathbf{k}) \quad \text{with } g(\mathbf{k}) = g_{obs}(\mathbf{k}) + \alpha \, g_{reg}(\mathbf{k}) \tag{7}$$

with $\mathbf{k} \equiv \eta = (\gamma h)$ and

$$g_{obs}(\cdot) = \frac{1}{2} \int_{\Omega} |H(\cdot)(\mathbf{x}) - H^{obs}(\mathbf{x})|^2 \chi_{tr}(\mathbf{x}) d\mathbf{x}, \tag{8}$$

χ_{tr} is the spatial restriction operator to the flight tracks, $g_{reg}(\cdot)$ a Tykhonov's regularization term, see e.g. [28]. In this step, it is defined as: $g_{reg}(\eta) = \frac{1}{2} \int_{\Omega} |\nabla \eta(\mathbf{x})|^2 d\mathbf{x}$. The weigh coefficient α is empirically set "at best" (see [39] for details). The surface elevation $H(\eta)$ corresponds to the solution of the RU-SIA equation (1) (with Dirichlet boundary conditions) with η given. The gradient of the cost functional is computed by introducing the adjoint equation. The minimisation algorithm is a quasi-Newton method (the L-BFGS algorithm of the Python routine scipy.optimize.minimize). We refer to [39] for more details.
The iterative minimization process is performed until convergence. Numerous numerical experiments have demonstrated robust convergences; see [39] for a thoroughly assessment of the method. In particular the optimal solution does not significantly depend on the smoothing length scale of the surface data (done in the present $Ant-p$ test cases at ≈ 24 km, see next Section), nor on the first guess (set to h_b the Bedmap2 value in EAIS).
Past this computational VDA step providing the optimal value η^*, the value of γ along the flights tracks where depth measurements h_b are available are straightforwardly deduced: $\gamma_{tr}^* = \dfrac{\eta^*}{h_b} \chi_{tr}(\mathbf{x})$. These values are inputs of the next algorithm, Step 2).
These obtained values are representative (and accurate, see [39]) at the flow model scale, that is at $10\bar{h} \approx 25$ km minimal wave length (with $\approx$ 2km mesh cells).

`Step 2)` Extension of γ in the whole domain by NNRK.
Given γ_{tr}^* along the flights tracks (result of Step 1)), a NNRK algorithm ([11,30]) is applied to extend values of γ to the whole area. This statistical learning algorithm is done in two steps: 1) an ANN estimator (deep learning) is buit up; 2) an ordinary Kriging of the residuals is added. Details are presented in next paragraph.

`Step 3)` Estimation of the pair $(h, \dot{a})$ in RU-SIA equation (1) by VDA.
Given γ all over the domain (result of Step 2)), the thickness h is infered simultaneously with the RHS $\dot{a}$ in (1) by another VDA process. The equations and the complete VDA algorithm are presented and assessed in detail in [39]. Let us recall them briefly.
Similarly to Step 1), the pair $(h, \dot{a})$ in (1) is infered by solving the optimal control problem (7) with g_{obs} defined by (8) but minimizing with respect to $\mathbf{k} = (h, \dot{a})$ (and not w.r.t. $\eta = \gamma h$ like in Step 1)). In this VDA process, the regularization term reads:

$$g_{reg}(h, \dot{a}) = \frac{1}{2} \|(h - h_b)\|_{C_h^{-1}} + \frac{1}{2} \|\dot{a} - \dot{a}_b\|_{C_a^{-1}} \tag{9}$$

with C_h^{-1} and C_a^{-1} covariance operators defining metrics. The latter are classically defined as the second order auto-regressive correlation matrices with length scale respecting a balance between the regularisation and the preconditioning effects of the VDA algorithm, see [39] for details and [20] for an analysis in another physical context. Next following [33,20], a change of the control variable is made. The numerous numerical experiments have demonstrated that this choice of covariance operators combined with the change of variable improves greatly the robustness and the convergence speed of the VDA algorithm, see [39].

In (7) the weight coefficient α is defined as a decreasing sequence following an iterative regularisation strategy, see [39] for details and [28] for an analysis. This iterative regularisation strategy improves the convergence speed of the VDA algorithm too.

Numerous assessment experiments are presented in [39], in particular the sensitivity of the present inversions with respect to: i) the uncertainties on γ; ii) the density of flights tracks (by removing some of them); iii) the smoothing length scale of the surface data (altimetry, InSar) from ≈ 24 to 48 km; iv) the first guess (chosen here as the Bedmap 2 value h_b).

Remark 1 *Using the explicit expression (2) of γ one can compute a-posteriori estimations of the (spatially distributed) slip ratio value R_s. This is an interesting feature to analyse a-posteriori the degree of the RU-SIA flow model consistency, see [39] for details.*

Remark 2 *Based on a-priori vertical thermal profile(s) (e.g. the one defined by (6)), the RU-SIA equation (1) provides a-posteriori estimations of the effective thermal boundary layer thickness $(B-b)$, see Fig. 2 (Left). Such a-posteriori estimations may be interesting for various analyses. Moreover the vertical profiles could be adjusted by constraining them with (the very few) in-situ measurements.*

2.3.2 The Neural Network Residual Kriging (NNRK) algorithm (Step 2))

The employed NNRK algorithm is decomposed in three steps as follows.

`Step 2a)` Considering the surface data $(H, \mathscr{S}, |\mathbf{u}_H|, \dot{a}_b)$ at all in-situ measurements locations (e.g. along the flights tracks of all areas Ant$p$) plus the values of γ^*_{tr} computed at Step 1), an estimator of γ is built up by training an Artificial Neural Network (ANN). This estimator is denoted by $\bar{\gamma}$.

The training dataset is denoted by $\mathscr{D}$; it contains "examples" (I_i, O_i), $i = 1, \cdots, N_{ft}$, where $I_i = (H, \mathscr{S}, |\mathbf{u}_H|, \dot{a}_b)(\mathbf{x}_i)$ is the i-th input and $O_i = \gamma^*_{tr}(\mathbf{x}_i)$ is the corresponding output. $\mathbf{x}_i$, $i = 1, \cdots, N_{ft}$ denote the i-th in-situ measurements coordinates (e.g. along the flight tracks).

The estimator $\bar{\gamma}$ is computed as the minimizer of the mean square misfit: $\frac{1}{N_{ft}} \sum_{i=1}^{N_{ft}} (O_i - \bar{\gamma}(I_i))^2$. This misfit function (also called loss function) reads:

$$j(\mathscr{D}; \cdot) = \frac{1}{N_{ft}} \sum_{i=1}^{N_{ft}} [\gamma^*_{tr}(\mathbf{x}_i) - \bar{\gamma}((H, \mathscr{S}, |\mathbf{u}_H|, \dot{a}_b)(\mathbf{x}_i); \cdot)]^2 \quad \text{for } \mathbf{x}_i \in \Gamma_{tr} \tag{10}$$

To solve this large dimensional data-based optimisation problem, the currently most efficient methods are ANN with few hidden layers (deep learning). Here, 5 layers are considered, see Fig. 3; each of the 4 hidden layers contains 50 neurones. The most efficient activation function is chosen: the rectified linear unit (ReLU) function, see e.g. [16,32]. The ANN is determined by its architecture and the weight parameters $(W_1, \cdots, W_5)$, Fig. 3.

The training step consists to identify the optimal values of these parameters W_j, $j = 1, .., 5$. Each W_j is a matrix of dimension $n_{out} \times n_{in}$. Here, W_1 has $5 \times 50 = 250$ parameters, W_j for $j = 2, 3, 4$ have $50 \times 50 = 2500$ parameters each, W_5 has $50 \times 1 = 50$ parameters. The ANN has been coded in Python using the PyTorch and Mpi4Py libraries [10].

To minimize $j(\mathscr{D}; (W_1, \cdots, W_5))$ with respect to $\{W_j\}_j$, the classical Adam method [31], a first-order gradient-based stochastic optimization, is employed. The learning rate (the gradient descent step size) is classically adjusted during the optimization procedure.

The input variables are heterogeneous and of different order of magnitude e.g. the elevation H (m) and the slope $\mathscr{S}$ (radian). Therefore each input variable v, v an element of $\{H, \mathscr{S}, |\mathbf{u}_H|, \dot{a}_b\}$, are reduced centred as follows: $\bar{v}_i = (v_i - mean(v)/\sigma(v))$, for all i, $1 \leq i \leq N_{ft}$. The normalisation is applied to mini-batches in hidden layers; this technique is supposed to improve the stability of the model; see e.g. [30] for more details and know-hows on ANN and NNRK algorithms.

Also to avoid overfitting, the dropout method [50] is adopted. (This technique may help to prevent overfitting). As usual, the hyper-parameters of the algorithm (learning rate, decay rate, dropout probability) are experimentally chosen; the selected values are those providing the minimal value of j.

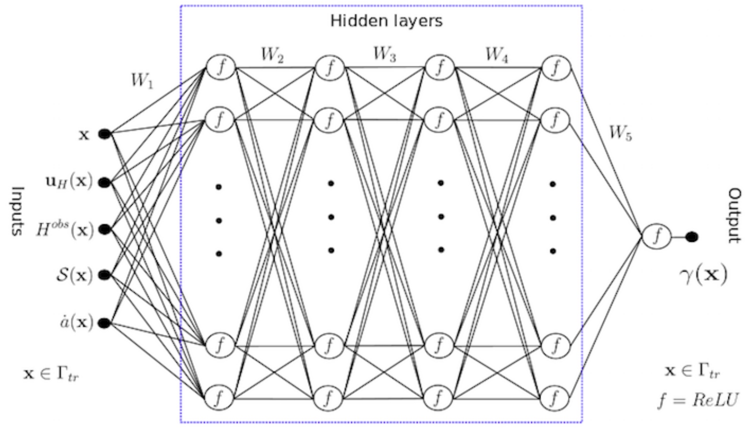

Fig. 3 The Artificial Neural Network (ANN) with four hidden layers. First step of the NNRK algorithm.

The ANN has been trained by including h_b as an input parameter or not. Both estimators (considering h_b as an input or not) turned out to have similar accuracies; thus confirming the strategy to predict the dimensionless parameter γ of the flow model from the surface data only.

`Step 2b)` The K-fold cross-validation method, see e.g. [1], is employed to assess the ANN accuracy and to confirm if the ANN can be used as a predictor. Let us recall that K-fold cross-validation method is as follows, see e.g. [1]:

- Divide randomly the original training data set $\mathscr{D}$ into K (roughly equal) subsets;
- For each subset $\mathscr{D}_k$, $k = 1 \cdots K$, the ANN is trained from the other $(K-1)$ subsets $\mathscr{D}_i,\ i \neq k$.
  We denote by $\mathscr{D}i_{test} = \mathscr{D}_i$ and $\mathscr{D}i_{train} = \cup_{j \neq i} \mathscr{D}_j$, $i = 1, \cdots, K$.
- Compute the loss function $j(\mathscr{D}i_{test})$ for each case.
  Finally, choose the best ANN i.e. those providing the smallest total loss function $j(\mathscr{D}i_{test}) + j(\mathscr{D}i_{train})$.

`Step 2c)` The residual along at the measurements locations is computed: $\varepsilon_\gamma = (\gamma^*_{tr} - \bar{\gamma})$ with $\bar{\gamma}$ computed by the (best) ANN. Next an ordinary Kriging (with a spherical semi-variogram model) is used to extend ε_γ all over the domain. The obtained estimator is denoted by $\hat{\varepsilon}_\gamma$. By construction this residual satisfies: $\mathbb{E}(\varepsilon_\gamma) \approx 0$. Moreover the correlation between two points depends on the distance between them and not on their location. Performing an ordinary Kriging on the residual after ANN is known to be particularly efficient, see e.g. [30] Chapter 3.

The final estimator in the whole domain is denoted by $\hat{\gamma}$. It is obtained as the sum of the ANN estimator and the ordinary Kriging estimator of residuals:

$$\hat{\gamma}(\mathbf{x}) = \bar{\gamma}(\mathbf{x}) + \hat{\varepsilon}_\gamma(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega \tag{11}$$

The forthcoming numerical results show that the estimator $\hat{\gamma}(\mathbf{x})$ provides (surprisingly) very accurate values of the parameter γ from the surface data only (altimetry, InSAR and $\dot{a}$).

2.3.3 On the linked uncertainty between γ and h

In the inversion algorithm previously described, after Step 1), one has to separate the effects of the two unknown fields: the physical-based dimensionless parameter γ and the ice thickness$h$. The accuracy and robustness of each VDA process has been demonstrated in [39]. It will be demonstrated in next section that the NNRK algorithm is robust and accurate too. Then it can be assumed that: $\gamma^t h^t \approx \gamma^* h^*$; where the superscript * denotes the optimal computed values while the superscript ^{t} denotes the (effective) true value. Let us denote: $e_\varphi = (\varphi^* - \varphi^t)/\varphi$. At order 1, one has:

$$e_h \approx -e_\gamma \text{ (where } \gamma \text{ does not tend to 0)} \tag{12}$$

In other words, Step 2) and Step 3) of the inversion algorithm propagates the error made on γ to h in the same order of magnitudes (in %).

Remark 3 *It would be straightforward to apply the same NNRK algorithm to directly estimate the thickness h all over the domain. However it seems definitively more consistent to estimate a dimensionless parameter of a flow model able to represent accurately the surface data, than to estimate the thickness data partially responsible only of the employed surface data. Following this idea of purely data-driven estimations, [9] had proposed an ANN trained and assessed on synthetic data generated by an ice flow model and geomorphic premises to estimate the bedrock elevation of four mountain glaciers.*

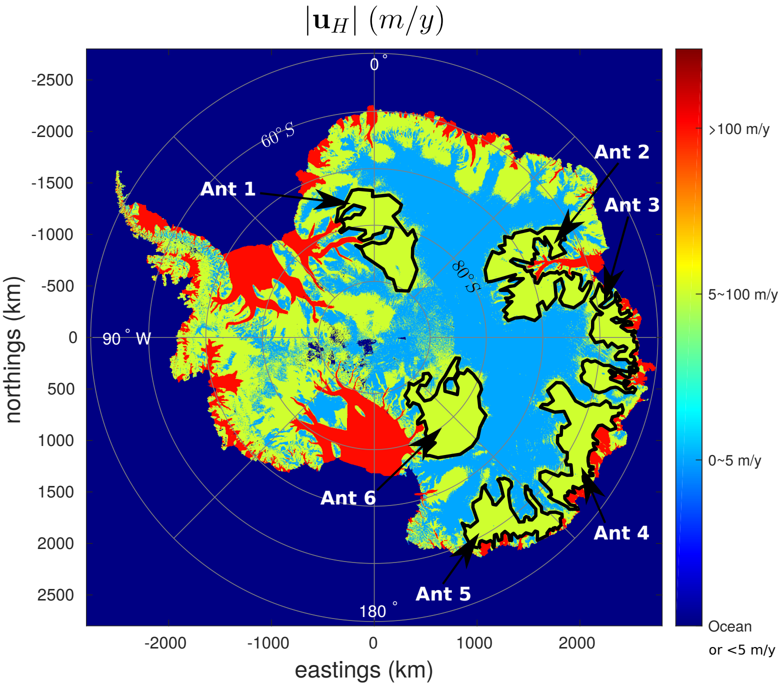

Fig. 4 Location of the 6 test areas Ant$p$ (east Antarctica) with InSAR-based surface velocity values in m/y (from [43]).

3 Data pre-processing

In the next sections, the algorithm is applied to 6 large areas in EAIS (ranging from 250268 to 439045 km^2); they are denoted by Ant$p$, $p = 1,..,6$, see Fig. 4. These areas include the upstream areas of major ice-sheds; all of them respect the flow model domain of validity. The mean thickness value of Bedmap2 ([14]) ranges from 1822 to 2745 m, see tables 2-11 for details. The coordinates of the characteristic points defining each area ($\approx 100-150$ points per area, see Fig. 4) are available on the open source computational software DassFlow webpage[2].
Estimating more accurately the bed elevation in these areas may be interesting since global warming may threaten EAIS stability as mentioned e.g. in [13].

The correlations between the given variables (H^{obs}, $\|\mathbf{u}_H\|$, $\mathscr{S}$, $\dot{a}$, h_b) for all areas have been computed. It turns out that no significant linear correlation between the variables have been observed.

In this section, the method to smooth the surface data accordingly with the flow model domain of validity is presented; the definition of adequate numerical grids follows. Moreover since Bedmap2 values are considered as the reference values, the method to obtain these values is recalled.

3.1 Minimal wave length, surface data smoothing and numerical grids

The surface data $|\mathbf{u}_H|$ and H have to be defined at an adequate scale to be consistent with the shallow (long wave assumption) flow model (1); next providing a minimal wave length L^* (km) of the inversions. The RU-SIA equation is accurate as soon as $\varepsilon = \frac{[H]}{[L]} \lesssim 0.1$. (In other respect it is shown in [54] that the ice-sheet areas presenting surface velocity ranging in $\approx [5-100]$ m/y are accurately modelled by the SIA model as soon as the minimal wave length equals $\approx 10-12$ km in mean).
The mean value of the Bedmap2 ice thickness (denoted by $\bar{h}_b$) [14] in the 6 Ant$p$ areas equals ≈ 2.7 km. In the considered regimes, the velocity field is co-linear to the slopes $\mathscr{S}$; therefore to be fully consistent, the smoothing of surface data should be done non-isotropically by defining a streamline minimal wave length and a cross-line one. For a sake of simplicity, here an isotropic smoothing is performed. To do so, a Gaussian with standard deviation $\sigma = 4$ km is convoluted with each given surface field: elevation and velocity norm. Then the smoothing effects are sensitive in disks of diameter

---

[2] Open-source computational software DassFlow: Data Assimilation for Free Surface Flows. Python version for 2D shallow generalised Newtonian fluids. INSA, University of Toulouse, CNES. *www.math.univ-toulouse.fr/DassFlow*.

$\approx 2 \times (3\sigma) \approx 10\bar{h}_b$ km.

Next a finite element mesh is built up using Gmsh software [15] with a grid size $\delta x \approx 3$ km. Indeed $\delta x = 3$ km provides ≈ 10 points per minimal wave length L^*, therefore respecting the minimal number of points to properly approximate all fields. The lineic flight tracks are meshed with cells of $\delta x \approx 2$ km. The given thickness measurements (provided in the Bedmap2 database [14]) are interpolated along the flight tracks .

3.2 Recalls of the origins of Bedmap2 values

Bedmap2 values h_b are considered as reference values; moreover they are employed to set the first guesses of the VDA processes at Step 1) and Step 3) (see paragraph 2.3.2). h_b and its a-priori uncertainty as derived in [14] are plotted for each test area Ant$p$ in figures 6, 9, 12, 13, 14 and 15 (Middle)(Left). In Bedmap2 database [14], the interpolation - extrapolation of airborne measurements are performed throughout the domain by the ArcGIS Topogrid routine (ESRI Ltd, ArcGIS 9); the latter is based on the ANUDEM algorithm [26]. This algorithm uses an iterative finite difference interpolation technique which is essentially a thin plate spline technique [53]. Next, empirical uncertainty values are stated as follows, see [14]. The thickness measurements are split into two datasets $(D1)$ and $(D2)$. Dataset $(D1)$ is used to build up an interpolation including at Dataset $(D2)$ location points; values of $(D2)$ being not used at this stage. Next, Dataset $(D2)$ is used to quantify the misfit with the "predicted" - interpolated values; and basic statistics on the results with dependence on the distance to data are deduced. For cells located between 5 and $20\,km$ from any data, [14] suggests that the interpolation error is an increasing function of distance from the closest data; beyond it would be not correlated. (Observe that this distance corresponds approximatively to the minimal wave length of the RU-SIA model). For cells that are more than 50 km from airborne measurements, the thickness estimation is based on gravity-field inversion (gravity-derived thickness); the proposed related uncertainty equals $\pm 1000\,m$. This is how the uncertainty values on Bedmap2 values h_b are defined; see figures 6, 9, 12, 13, 14 and 15 (Middle)(Right).

4 Analysis of each inversion algorithm step

4.1 Step 1: Estimation of η by VDA

The effective diffusivity η defined in (1) is estimated in each area Ant$p$ by VDA following Step 1) described in Paragraph 2.3.1. The convergence of this iterative VDA process is very slow (a few hundreds of iterations) but very robust in particular with respect to the first guess value; it has been thoroughly assessed in [39]. The stopping criteria is the stationarity of $\|\eta\|$. Here the RHS $\dot{a}$ provided by [52] is supposed to be exact. After convergence of the VDA process, given the ice thickness along the flights tracks, the computed optimal value γ^* is saved for Step 2), that is:

$$\gamma^*_{tr}(\mathbf{x}) = \frac{\eta^*}{h_b}(\mathbf{x}) \;\text{ for } \mathbf{x} \in \Gamma_{tr}$$

4.2 Step 2: Estimation of γ by NNRK

The ANN algorithm input values, see Fig. 3, are datasets along all flight tracks in all test areas Ant$p$, $p = 1,..,6$ plus the values of γ^*_{tr} computed at Step 1). This dataset is denoted by $\mathscr{D}$. Following Step 2b) (see paragraph 2.3.2), the K-fold cross-validations are performed with $K = 6$. The results are presented in Fig. 5. (The value $K = 6$ is completely independent of the total test cases number). Here, the training sets $\mathscr{D}i_{train}$ contains 16774 examples and a test set $\mathscr{D}i_{test}$ contains 3354 examples; each example being an (input, output) pair of ANN.

It can be read on Fig. 5 (Left) that all ANN models trained from the different data sets $\mathscr{D}i_{train}$ provide very close cost values $j(\mathscr{D}i_{train})$ (see (10)). Moreover, the cost values of all test sets $j(\mathscr{D}i_{test})$ are almost all equal. This shows that all ANN have very close prediction capability, all being excellent. Indeed after optimisation, $j(\mathscr{D}i_{test}) \approx 5\ 10^{-3}$ (Fig. 5 (Right)); this corresponding to $\approx 1\%$ of the mean value of $\bar{\gamma}$. The ANN trained from $\mathscr{D}2_{train}$ is selected since having a slightly smaller misfit value $(j(\mathscr{D}i_{train}) + j(\mathscr{D}i_{test}))$. These tests of prediction capabilities demonstrate the robustness and accuracy of the trained ANN.

Next, values of $\bar{\gamma}$ are predicted in the whole domains Ω_p, $p = 1, \cdots, 6$ by performing the selected ANN.

Next by performing the Kriging step (Step 2c) described in Paragraph 2.3.2), the predictor $\hat{\gamma}$ is obtained.

Following the a-priori estimation derived in Section 2.2, the upper bound 0.9 is imposed to the estimation; this upper bound is active at very few locations only; moreover it is in great majority where the uncertainty on h_b is low.

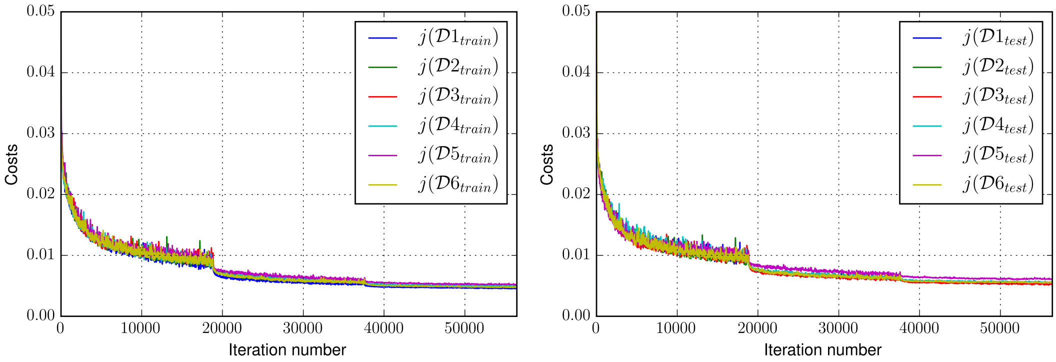

Fig. 5 Misfit functional $j(\mathscr{D})$ defined by (10), vs minimization iterations for: (Left) different train sets $\mathscr{D}i_{train}$; (Right) different test sets $\mathscr{D}i_{tests}$. The learning rate (gradient descent step size) is decayed by 0.2 every 30 epochs (1 epoch = 1 forward pass and 1 backward pass of all the training examples), see e.g. [30] for details on this classical method.

Table 1 Difference between values of H the response of RU-SIA equation (1) if changing the RHS value: $\dot{a}_b$ the Racmo2 value, $\dot{a}^*$ the infered value at Step 3). Domain Ω_p=Ant1.

Difference of H	Median	Mean	Max
$\lvert H(\dot{a}_b) - H(\dot{a}^*)\rvert$	$3.3m$	$6.0m$	$26.3m$

For each test area, the predicted values of γ are plotted; see figures 6, 9, 12, 13, 14 and 15 (Up) (R).

4.3 Step 3: Estimation of h (and adjustment of $\dot{a}$) by VDA

After Step 2), the dimensionless parameter γ in RU-SIA equation (1) is given. Then the thickness h is infered by VDA following the method presented in paragraph 2.3.1. The optimisation algorithm converges in $\sim 20-50$ iterations depending on the test case. Its convergence is robust: it has been thoroughly analysed in [39].
At this step, h is simultaneously infered with $\dot{a}$. Indeed this enables to adjust the value of $\dot{a}$ within its uncertainty range which is here $\sim \pm 20\%$, see [52]. It can be noticed in tables 2, 5, 8, 9, 10 and 11 that the corrections made remains in great part within this a-priori uncertainty range. Indeed the upper bound is active at few locations only. In other words, this adjustment based on the physical-based model RU-SIA is consistent with the (totally independent) estimations derived in [52].
Of course, given the surface data, any change of $\dot{a}$ in RU-SIA equation (1) has an effect on the infered value of h. As an example let us compute the response H of RU-SIA equation (1) in Ant1 area with: i) the RACMO2 value $\dot{a}_b$ in the RHS (providing $H(\dot{a}_b)$); ii) the infered value $\dot{a}^*$ by VDA in the RHS (providing $H(\dot{a}^*)$). The obtained difference are the ones indicated in Tab. 1. Therefore the correction made on the RHS $\dot{a} = (\partial_t h - a)$ implies a correction on the ice thickness h negligible ($\sim 1\%$) compared to the one obtained by the complete inversioin method, see Tab. 2 in next section. This remark holds for all the domains Ω_p.

The assessment analysis presented here is complimentary to those proposed in [39]. Moreover in [39], Step 2) is based on a more basic ordinary Kriging algorithm; also a single test area only is considered. The assessment analysis in [39] focuses on: i) the uncertainty on γ; ii) the proximity or not of flight tracks; iii) the smoothing length scale of the surface data; iv) the first guess value (chosen here as Bedmap 2 h_b).

4.4 On the RU-SIA model accuracy

For each test area Ant$p$, domain information and basic statistics on the results are presented, see tables 2, 8, 5, 9, 10, 11. Statistics on the computed surface elevation H, output of RU-SIA model, are indicated. It can be noticed that the RU-SIA equation solved from Bedmap2 value h_b and the (purely data-driven) estimation of γ obtained at Step 2) already fits very well the measured surface elevation H^{obs}, see "Direct model validation" lines in the tables. This very good accuracy (based on the Bedmap2 bed elevation h_b i.e. without any additional calibration of h) demonstrates the validity and the relevance of the RU-SIA model. After Step 3) of the inversion algorithm i.e. after the identification of h and $\dot{a}$ by VDA, of course the RU-SIA model fits even better the measured surface elevation H^{obs}, see "$\lvert H(h^*) - H^{obs}\rvert$ (after h-inversion)" in the tables.

5 Results and sensitivity tests (Ant1 and Ant3 areas)

In this section, the bed elevation b (equivalently the ice thickness h) is infered by the inverse method described in Section 2.3 for the two areas Ant1 and Ant3, see Fig. 4, tables 2 and 5. Different estimations of h are compared, depending if:

a) isolated flight tracks (hence locally highly constraining) are considered or not;
b) the learning method at Step 2) is the NNRK algorithm described in Paragraph 2.3.2, or an ordinary Kriging method as in [39].

These comparisons aim at analysing the robustness and accuracy of the present inverse method.
Ant1 is a 370809 km^2 area north-east upstream of Bailey, Slessor and Recovery ice-streams; Ant3 is a 250268 km^2 area in Wilhelm and Queen Mary lands, upstream of Shackleton ice shelf and Davis sea.
Among the considered six areas, Ant1 and Ant3 are those presenting the largest uncovered parts during airborne campaigns. As a consequence they contain large areas where Bedmap2 estimation h_b is based on gravity field inversions, therefore presenting very large uncertainties.
For each case, the domain information and basic statistics on the numerical results are presented in tables 2 and 5. For each case, the most relevant fields are plotted, see figures 6 and 9: the surface velocity norm $|\mathbf{u}_H|$ and the flight tracks locations (Up)(L), the NNRK estimation $\hat{\gamma}$ defined by (11) (Up) (R), the Bedmap2 value h_b (Middle)(L) with its uncertainty as presented in [14] (Middle)(R), the present thickness estimation h^* (Down)(L) and its difference with h_b (Down)(R).

5.1 Results for Ant1 area

This domain presents large unexplored areas during the airborne campaigns therefore huge uncertainty on h_b values, see Fig. 6 (Middle).

5.1.1 The ice thickness estimation h^*

Recall that $\hat{\gamma}$ is the NNRK estimation of the dimensionless parameter γ defined by (2).No correlation is observed between h and γ; the only clearly observed correlation is : γ is small where $|\mathbf{u}_H|$ is small, see Fig. 6 (Up). This observation is fully consistent with the a-priori analysis done in Section 2.2, see (2) and Fig. 2.
Recall that h_b values are thin plate spline based estimations (see [14] and paragraph 3.1) hence intrinsically smooth, Fig. 6 (Middle)(Left). The present estimation h^* is the optimal solution of a data-driven model combined with a physical-based model; no additional smoothing has been added, then the obtained value h^* is much less smooth than h_b. It is worth to notice that a non isotropic smoothing of the surface data (see the previous section) would have provided bed elevation patterns more correlated to the surface streamlines.
The difference between h^* and h_b is not correlated to the distance to the nearest measurement (flight track), on contrary to the empirically stated uncertainty in [14]. Indeed large corrections of h_b (up to 1500 m) are obtained close to flight tracks; close meaning at $1-2$ minimal wave lengths of the RU-SIA model that is $\sim 20-40$ km, see e.g. in Fig. 6 (Down)(R) the areas around coordinates (2750,1950)(2900,1550)(3050,2050). At the opposite, h^* may remain very close to h_b in areas relatively far from any flight tracks.
Recall that the flight tracks are meshed as 1d lines; along these segments ($\approx$ 2 km long), the infered depth value can vary of $\pm$150m around the measured values (inequality constraints imposed in the VDA processes). Therefore in the adjacent triangles (which are nearly equilateral with $\approx$ 2km edges), the plotted mean depth values may be already much different than the measured ones. This large scale smoothing may explain the potential great differences between the two estimations even relatively close to the flight tracks.
The basic statistics presented in Tab. 2 show that after the VDA processes, the RU-SIA equation fits extremely well the measured surface elevation. The correction made on $\dot{a}$ is relatively consequent, 17% in mean (it is the highest correction made among the 6 test cases Ant$p$). However it remains lower than the maximal authorised correction: $\pm$20%.
Finally the correction made on h_b is non negligible: 13.3% (356.8 m) in mean, with a 1.6% (1.6 $10^4\,km^3$) of volume change only (for 370809 km^2).

5.1.2 Ant1: if removing some flight tracks

In this paragraph, a new ice thickness estimation is computed. It differs from h^* since the flight tracks indicated in Fig.6 (Up)(L) are not considered anymore. The original complete set of flight tracks is denoted by Γ^{all}; the partial one

Table 2 Domain Ω_p=Ant1, information and results.

Domain Ω_p & mesh information			
Surface $\lvert\Omega_p\rvert$	$370809\,km^2$		
Mean ice thickness of h_b (Bedmap2)	$2696.2\,m$		
# mesh vertices: in Ω_p / on flight tracks	57661 / 2152		
RU-SIA model output (with $\hat{\gamma}$)	Median	Mean	Max
$\lvert H(h_b) - H^{obs}\rvert$ (before h-inversion)	$7.0\,m$	$10.3\,m$	$53.0\,m$
$\lvert H(h^*) - H^{obs}\rvert$ (after h-inversion)	$2.6\,m$	$3.7\,m$	$44.6\,m$
Infered RHS $\dot{a}$			
$\lvert\dot{a}^* - \dot{a}_b\rvert$	$0.7\,cm/y$	$0.7\,cm/y$	$3.6\,cm/y$
$\lvert\dot{a}^* - \dot{a}_b\rvert / \lvert\dot{a}_b\rvert$	19.2%	17.0%	20.0%
Infered thickness h			
$\lvert h^* - h_b\rvert$	$275.2\,m$	$356.8\,m$	$1953.5\,m$
$\lvert h^* - h_b\rvert / \lvert h_b\rvert$	10.4%	13.3%	65.7%
Ice volume change in km^3 / in %	$1.6\ 10^4\,km^3$ / 1.6%		

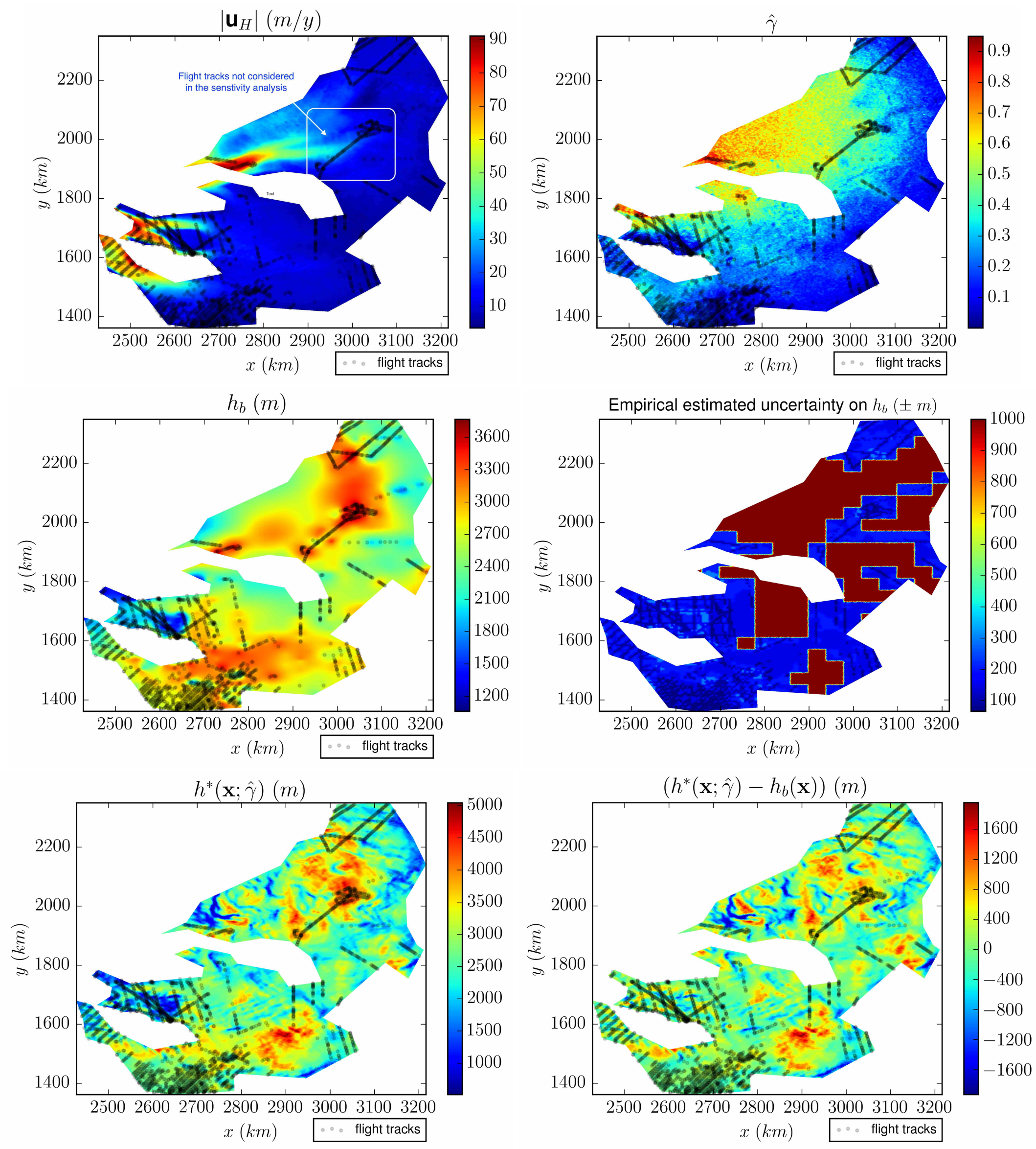

Fig. 6 Domain Ω_p=Ant1 (the plotted coordinates equal the Eastings-Northings plus $(2800, 2800)$km): (Up)(L) Surface velocity module $|\mathbf{u}_H|$ and flight tracks (R) $\hat{\gamma}$ computed by NNRK, see (11). (Middle)(L) Thickness h_b from Bedmap2 [14] (R) Empirical uncertainty on h_b from [14]. (Down)(L) Infered thickness with $\hat{\gamma}$: $h^*(\hat{\gamma})$ (R) Difference $(h^*(\hat{\gamma}) - h_b)$.

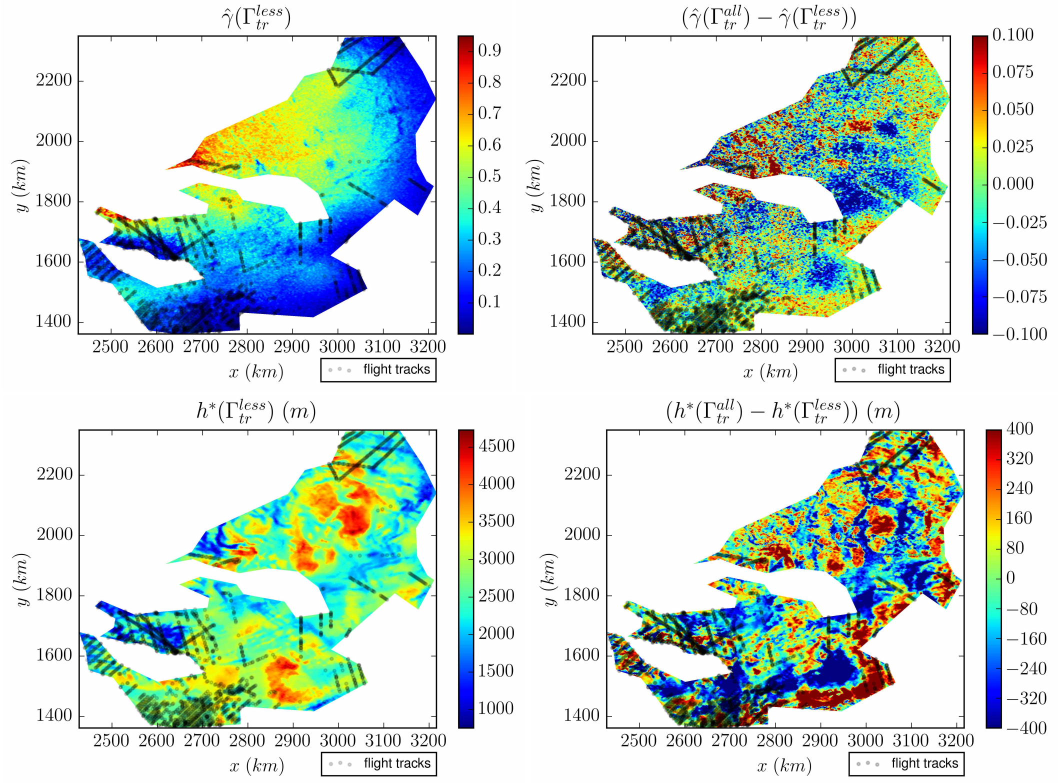

Fig. 7 Domain Ω_p=Ant1: comparison if not considering the flights tracks indicated in Fig.6 (Up)(L). (Up)(L) Field $\hat{\gamma}(\Gamma_{tr}^{less})$ (i.e. without the flights tracks indicated in Fig.6 (Up)(L)). (R) Difference $(\hat{\gamma}(\Gamma_{tr}^{all}) - \hat{\gamma}(\Gamma_{tr}^{less}))$. (Down)(L) Infered thickness $h^*(\Gamma_{tr}^{less})$ (R) Difference between the two estimations: $(h^*(\Gamma_{tr}^{all}) - h^*(\Gamma_{tr}^{less}))$.

Table 3 Domain Ω_p=Ant1. Comparison of the estimations if considering or not the flights tracks indicated in Fig.6 (Up)(L).

Infered thickness difference	Median	Mean	Max
$\lvert h^*(\Gamma_{tr}^{all}) - h^*(\Gamma_{tr}^{less})\rvert$	$151.3m$	$196.9m$	$1524.5m$
$\lvert h^*(\Gamma_{tr}^{all}) - h^*(\Gamma_{tr}^{less})\rvert / \lvert h^*(\Gamma_{tr}^{less})\rvert$	5.6%	6.6%	80.0%

is denoted by Γ^{less}. In Γ^{less} case, the mesh of the entire area has been re-builded (since in each mesh, the lineaic flight tracks are meshed differently). The inverse problem based on Γ^{less} is less constrained in particular the two VDA processes (Steps 1 and 3). Indeed the removed flight tracks are isolated, see Fig. 6(Up)(L); no constraint is imposed in the vicinity of these removed flight tracks anymore.

The statistical learning at Step 2) is unchanged, therefore values of $\bar{\gamma}$ are unchanged too. However the estimation $\hat{\gamma}$ defined by (11) is not the same since the Kriging step is changed. Indeed the latter is based on less flight tracks data. The difference between the two estimations $(\bar{\gamma}(\Gamma^{all}) - \hat{\gamma}(\Gamma^{less}))$ is plotted in Fig. 3 (Up)(R). It can be noticed that $\hat{\gamma}$ is changed all over the domain and not particularly in the vicinity of the missing flight track. Indeed, the Kriging method (Step 2c) in Paragraph 2.3.2) aims at computing the minimal variance in norm 2 (least square) and not point-wise; hence the global change of $\hat{\gamma}$.

Next the infered thickness h^* is different for two reasons: 1) values of γ are different; 2) the VDA process of Step 3) is not locally constrained at the missing flight tracks locations. The difference between the two thickness estimations is plotted in Fig. 3 (Down)(R). For a sake of readability, the legend in Fig. 7(Down)(R) has been bounded at ±400m (very few locations were greater than this bound). Basic statistics on the difference are presented in Tab. 3. Differences of 300 m correspond to $\approx 10-15\%$ of change. As expected, see (12), the variations of h are correlated to the variations of γ: compare Fig. 7 (Up)(R) to (Down)(R). Since the global change of γ, the change of h is not particularly important around the missing flight tracks.

Finally, it is worth to mention that the present inversion method is relatively global with local constraints ("in-situ" measurements along the flight tracks); it is not purely local inversions. In the present experiment, the obtained variations of h^* are roughly half than the ones obtained from h_b, see tables 2 and 3: difference of 6.6% in mean vs 13.3%, and 5.6% vs 10.4% for median values.

Table 4 Domain Ω_p=Ant1. Comparison the original thickness estimation (obtained using NNRK) to the one obtained using ordinary Kriging at Step2)

Infered thickness difference	Median	Mean	Max
$\lvert h^*(\hat{\gamma}) - h^*(\gamma_{krig})\rvert$	$145.5m$	$183.5m$	$1264.2m$
$\lvert h^*(\hat{\gamma}) - h^*(\gamma_{krig})\rvert / \lvert h^*(\hat{\gamma})\rvert$	5.3%	7.2%	117.2%

5.1.3 Ant1: with a different statistical learning method at Step 2)

It has been previously shown that uncertainties on γ generates uncertainties on h of same order of magnitude, see (12). In this paragraph, the influence of the statistical estimator considered at Step 2) is investigated. To do so, first we compare:

a) $\bar{\gamma}$ obtained by ANN algorithm at Step 2a) to $\hat{\gamma}$ obtained by the complete NNRK algorithm (see (11);
b) γ_{krig} obtained by an ordinary Kriging of values infered along the flights tracks to $\hat{\gamma}$.

It may be a-priori guessed that a deep learning method (as the present NNRK algorithm) is more reliable than a Kriging inter/extrapolation of its values between the flight tracks.

Second, we compare the infered thickness h^* obtained from $\hat{\gamma}$ (that is the original estimation plotted in Fig. 6 (Down)(L)) to the one obtained from γ_{krig}.

As expected, the difference between $\bar{\gamma}$ and $\hat{\gamma}$ (i.e. before and after the Kriging Step 2c) are localised in the vicinity of the flight tracks. In Ant1 case, these differences may be up to $\sim \pm 20\%$, see Fig. 8(Up)(R). More interestingly and as expected too, the differences between $\hat{\gamma}$ and γ_{krig} are not clearly correlated to the distance to the nearest flight track. The observed difference in Ant1 case may be consequent: $\sim \pm 40\%$, see Fig. 8(Middle)(R).
Next, like in the previous sensitivity test (and for the same reasons), the variations of h are correlated to the variations of γ: compare Fig. 8 (Middle)(R) to (Down)(R). Some statistics on the differences are presented in Tab. 4. Again the obtained variations in h are roughly half than the ones obtained from h_b, see tables 2 and 4: difference of 7.2% in mean vs 13.3%, and 5.3% vs 10.4% for median values.

5.2 Results for Ant3 area

Like Ant1, Ant3 area presents large uncovered areas during the airborne campaigns, corresponding to huge uncertainty on h_b, see Fig. 9.

5.2.1 The ice thickness estimation h^*

Like in Ant1 case, the only observed correlation is: γ is small if $\mathbf{u}_H$ is small, see Fig. 9 (Up). Again the difference between h^* and h_b is uncorrelated to the distance to the nearest flight track (on contrary to the empirically established uncertainty in [14]). Large corrections of h_b (up to 700 m) are found close to flight tracks see e.g. the area around coordinates (5550,2800) in Fig. 6 (Down)(R); at the opposite, h^* may remain very close to h_b in areas relatively far from any flight tracks, see e.g. the area around coordinates (5250,2800).
The few statistics presented in Tab. 5 show that again after the VDA processes, the RU-SIA equation fits extremely well the measured surface elevation. The correction made on $\dot{a}$ is much lower than the authorised maximal variation: 11.2% in mean. In Ant3, the global correction made on h_b is relatively low: 6.6% in mean (3.5% median) with a 0.5% of volume change only. However in the most uncertain areas, the corrections made can be either low, see e.g. the areas around coordinates (5000,2650) (5200,3250), or important ($\pm \sim 700$m), see e.g. the area around coordinates (5000,2650).

5.2.2 Ant3: if removing some flight tracks

The ice thickness obtained if not considering the flight tracks indicated in Fig.9 (Up)(L) is compared to the original estimation h^* (the one plotted in Fig.9 (Down)(L)). For the same reason as in Ant1 case, both $\hat{\gamma}$ and h are changed all over the domain and in the vicinity of the missing flight track only. Largest changes are obtained in areas far to the missing tracks; also it may close to (assimilated) flight tracks, see e.g. the area around coordinates (5100,3200). The difference between the two estimations are plotted in Fig. 6 (Up)(R) and (Down)(R). (Again for a sake of readability, the legend in Fig. 10(Down)(R) has been bounded at ±400m; very few values being greater than this bound). Basic statistics on the difference are presented in Tab. 6. (A difference of 200m corresponds to $\approx 10\%$ of change). Again, the obtained variations on h are roughly half than the ones obtained from h_b, see tables 5 and 6: difference of 2.8% in mean vs 6.6%, and 2.2% vs 3.5% for median values.

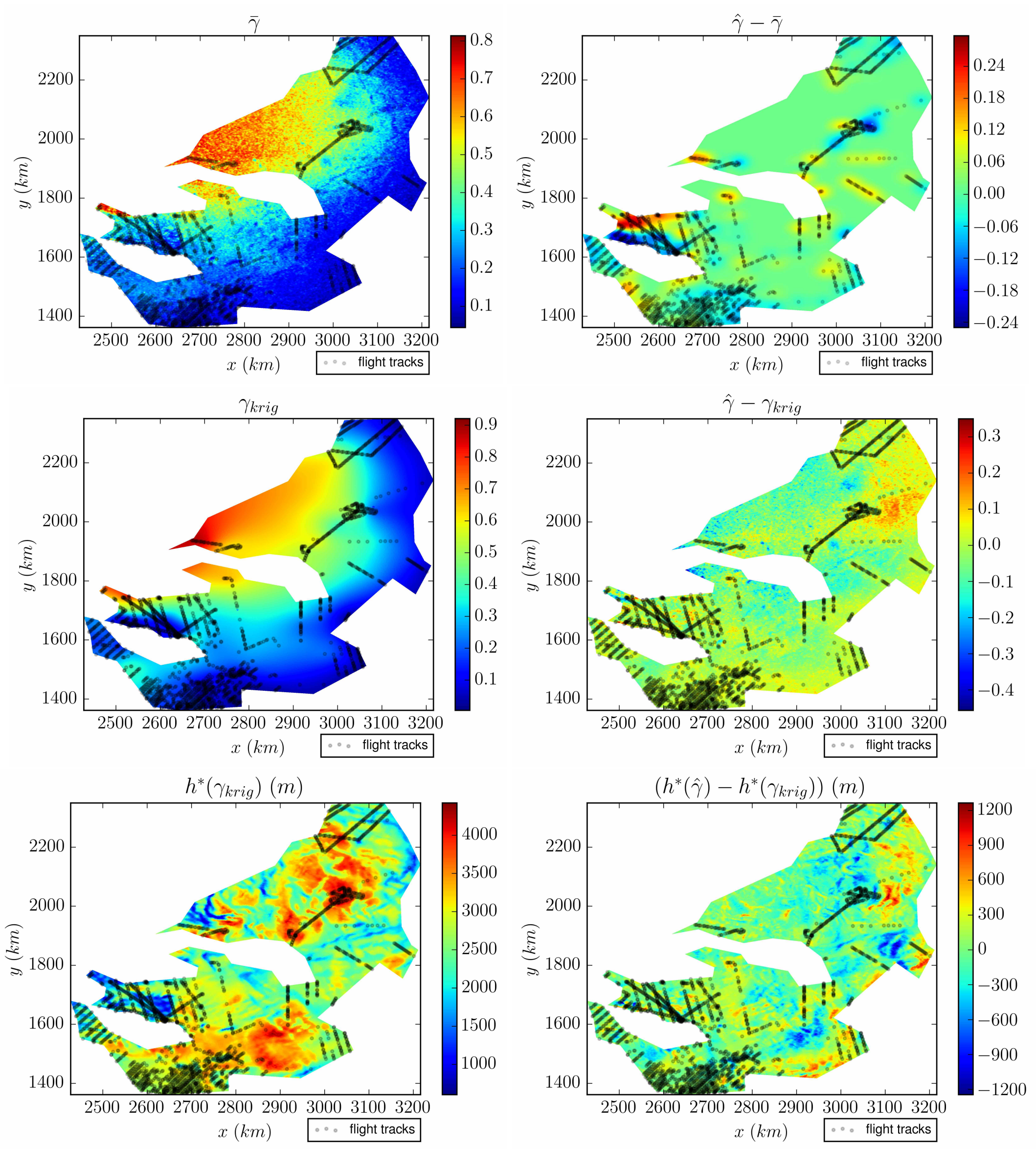

Fig. 8 Domain Ω_p=Ant1: comparison between different statistical learning methods at Step 2). (Up)(L) $\bar{\gamma}$ computed by ANN only. (R) Difference $(\hat{\gamma}-\bar{\gamma})$. (Middle)(L) γ_{krig} computed by ordinary Kriging. (R) Difference $(\hat{\gamma}-\gamma_{krig})$. (Down)(L) Infered thickness $h^*(\gamma_{krig})$. (R) Difference $(h^*(\hat{\gamma})-h^*(\gamma_{krig}))$.

Table 5 Domain Ω_p=Ant3, information and results.

Domain Ω_p & mesh information			
Surface $\lvert\Omega_p\rvert$	$250268\,km^2$		
Mean ice thickness of h_b (Bedmap2)	$1822.8\,m$		
# mesh vertices: in Ω_p / on flight tracks	$42881/2443$		
RU-SIA model output (with $\hat{\gamma}$)	Median	Mean	Max
$\lvert H(h_b)-H^{obs}\rvert$ (before h-inversion)	$7.8\,m$	$12.7\,m$	$274.0\,m$
$\lvert H(h^*)-H^{obs}\rvert$ (after h-inversion)	$2.8\,m$	$4.0\,m$	$110.6\,m$
Infered RHS $\dot{a}$			
$\lvert\dot{a}^*-\dot{a}_b\rvert$	$2.2\,cm/y$	$2.5\,cm/y$	$22.1\,cm/y$
$\lvert\dot{a}^*-\dot{a}_b\rvert/\lvert\dot{a}_b\rvert$	11.2%	11.4%	20%
Infered thickness h			
$\lvert h^*-h_b\rvert$	$70.0\,m$	$124.5\,m$	$862.2\,m$
$\lvert h^*-h_b\rvert/\lvert h_b\rvert$	3.5%	6.6%	63.5%
Ice volume change in km^3 / in %	$3.0\ 10^3\,km^3$ / 0.5%		

Table 6 Domain Ω_p=Ant3. Comparison if considering or not the flights tracks indicated in Fig.9 (Up)(L).

Infered thickness difference	Median	Mean	Max
$\lvert h^*(\Gamma_{tr}^{all})-h^*(\Gamma_{tr}^{less})\rvert$	$42.7\,m$	$56.3\,m$	$904.6\,m$
$\lvert h^*(\Gamma_{tr}^{all})-h^*(\Gamma_{tr}^{less})\rvert/\lvert h^*(\Gamma_{tr}^{less})\rvert$	2.2%	2.8%	78.7%

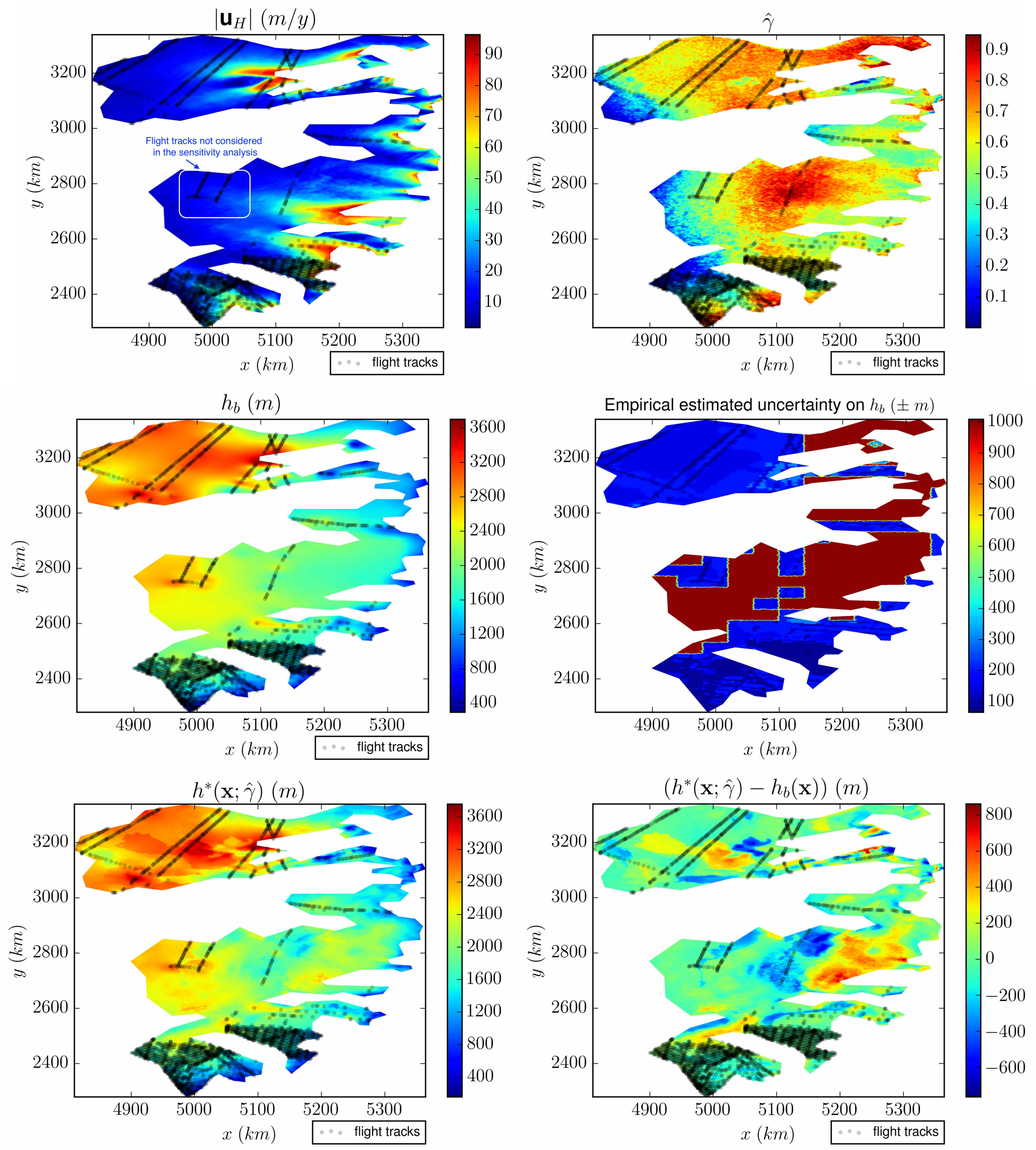

Fig. 9 Domain Ω_p=Ant3 (the plotted coordinates equal the Eastings-Northings plus $(2800, 2800)$km): (Up)(L) Surface velocity module $|\mathbf{u}_H|$ and flight tracks (R) $\hat{\gamma}$ computed by NNRK, see (11). (Middle)(L) Thickness h_b from Bedmap2 [14] (R) Empirical uncertainty on h_b from [14]. (Down)(L) Infered thickness with $\hat{\gamma}$: $h^*(\hat{\gamma})$ (R) Difference $(h^*(\hat{\gamma}) - h_b)$.

Table 7 Domain Ω_p=Ant3. Comparison of the original thickness estimation (obtained using NNRK) to the one obtained using ordinary Kriging at Step2).

Infered thickness difference	Median	Mean	Max
$\lvert h^*(\hat{\gamma}) - h^*(\gamma_{krig})\rvert$	$58.3m$	$84.9m$	$1418.7m$
$\lvert h^*(\hat{\gamma}) - h^*(\gamma_{krig})\rvert / \lvert h^*(\hat{\gamma})\rvert$	2.9%	4.3%	124.5%

5.2.3 Ant3: with a different statistical learning method at Step 2)

Similarly to the Ant1 case, the difference between $\bar{\gamma}$ and $\hat{\gamma}$ (i.e. before and after the Kriging Step 2c)) are mainly localised in the vicinity of the flight tracks with amplitudes up to $\sim \pm 0.20$, Fig. 11 (Up)(R). Again too, the differences between $\hat{\gamma}$ and γ_{krig} are not in the vicinity of the tracks only; they may be anywhere. The observed difference is up to $\sim \pm 0.45$, see Fig. 11(Middle)(R). Statistics on the differences on the corresponding estimated thicknesses are presented in Tab. 7. The obtained differences on h are about one third lower than the ones obtained from h_b, see tables 2 and 7: difference of 4.3% in mean vs 6.6%, and 2.9% vs 3.5% for median values.

In summary, these empirical sensitivity analyses highlight the robustness and the reliability of the inversion method. If not considering some in-situ measurements (along some flight tracks), see figures 6 and 9 (Up)(L), the obtained differences on h are roughly half than the ones obtained between h^* and h_b.

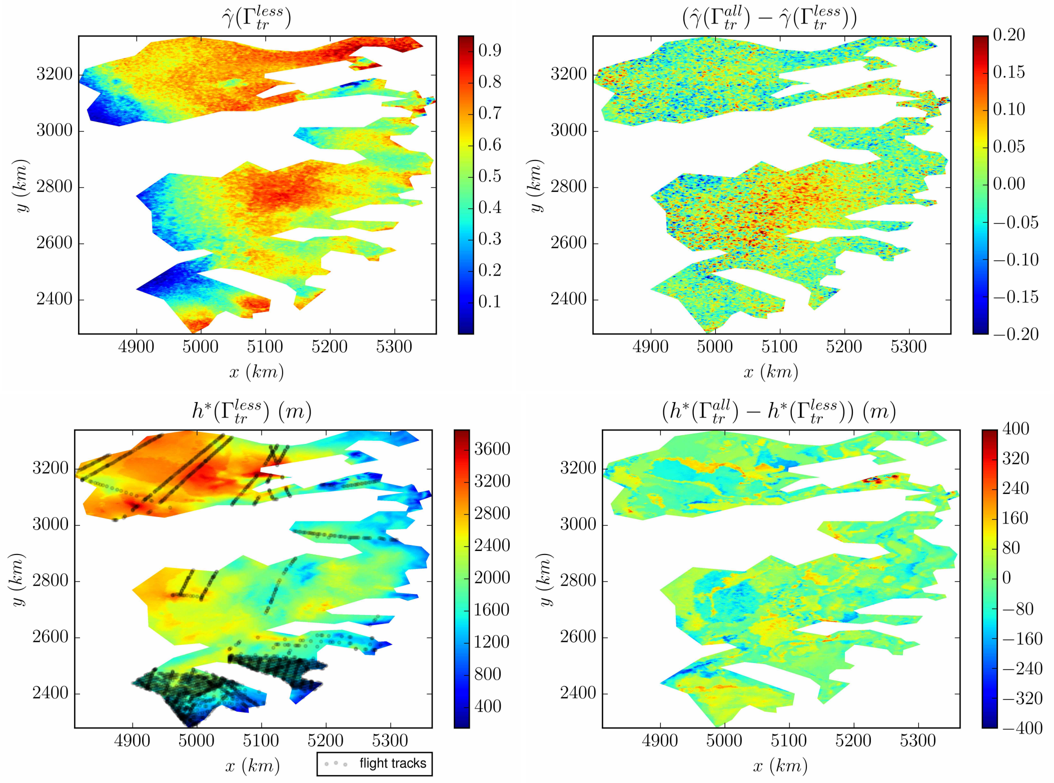

Fig. 10 Domain Ω_p=Ant3: comparison if not considering the flights tracks indicated in Fig.9 (Up)(L). (Up)(L) Field $\hat{\gamma}(\Gamma_{tr}^{less})$ (i.e. without the flights tracks indicated in Fig.9 (Up)(L)). (R) Difference $(\hat{\gamma}(\Gamma_{tr}^{all}) - \hat{\gamma}(\Gamma_{tr}^{less}))$. (Down)(L) Infered thickness $h^*(\Gamma_{tr}^{less})$. (R) Difference between the two estimations: $(h^*(\Gamma_{tr}^{all}) - h^*(\Gamma_{tr}^{less}))$.

If considering a simple Kriging method to estimate γ instead of the NNRK algorithm, the obtained differences on h in Ant1 case (resp. in Ant3 case) are roughly $1/2$ (resp. $2/3$) the ones obtained between h^* and h_b.
Therefore in all investigated cases, the obtained variations are lower than the ones obtained from h_b, see sections 5.1.1 and 5.2.1.

6 Conclusion

In this study, a method to infer the bedrock topography beneath glaciers at wave length $\sim 10\bar{h}$, $\bar{h}$ a characteristic thickness value (with a resolution at $\sim \bar{h}$) is developed. The key ingredients of this inversion method are the following: a) a dedicated Reduced-Uncertainty flow model (RU-SIA) taking into account a relatively complete physics (including depth-varying internal deformation) while presenting a single dimensionless parameter γ; b) two advanced VDA processes; c) an Artificial Neural Network (ANN) estimating γ from the surface data and in-situ measurements (acquired during airborne campaigns). A guiding idea of the method is to consider at each step robust (somehow well-posed) inverse problems, or at least not trivially ill-posed ones. All steps of the algorithm have been thoroughly assessed, partly in [39] next in the present study, in particular the estimation sensitivity with respect to the presence or not of local in-situ measurements (airborne campaings). The ANN capabilities to infer the single dimensionless parameter γ turned out to be excellent for all considered datasets in EAIS.

Observe that it is straightforward to apply the same ANN to estimate the ice thickness h (instead of γ) with the ANN trained from the in-situ datasets acquired along the flight tracks. We have performed such estimations; they turned out to be too much dependent on the training datasets. In particular they are sensitive to the presence or not of local in-situ measurements. In other words these purely data driven estimations turned out to be not robust enough. It seems definitively more consistent to estimate a single dimensionless parameter of the flow model which accurately represent the measured surface data.
The present hybrid physically-constrained data-driven method enables to estimate the bed elevation between in-situ measurements, in poorly covered areas and even in large uncovered areas e.g. in EAIS where the current estimations are gravimetry-based (hence highly uncertain). Indeed the method is relatively global (with local constraints along the flight tracks) and not purely local. It has been shown to be relatively robust. The estimations in large uncovered areas remain accurate as elsewhere since based on the surface signature (measured by altimetry and InSAR). In the considered six large EAIS areas, the corrections made to Bedmap2 may be large (up to 1000m) even at 2 wave lengths ($\approx$ 50km) away

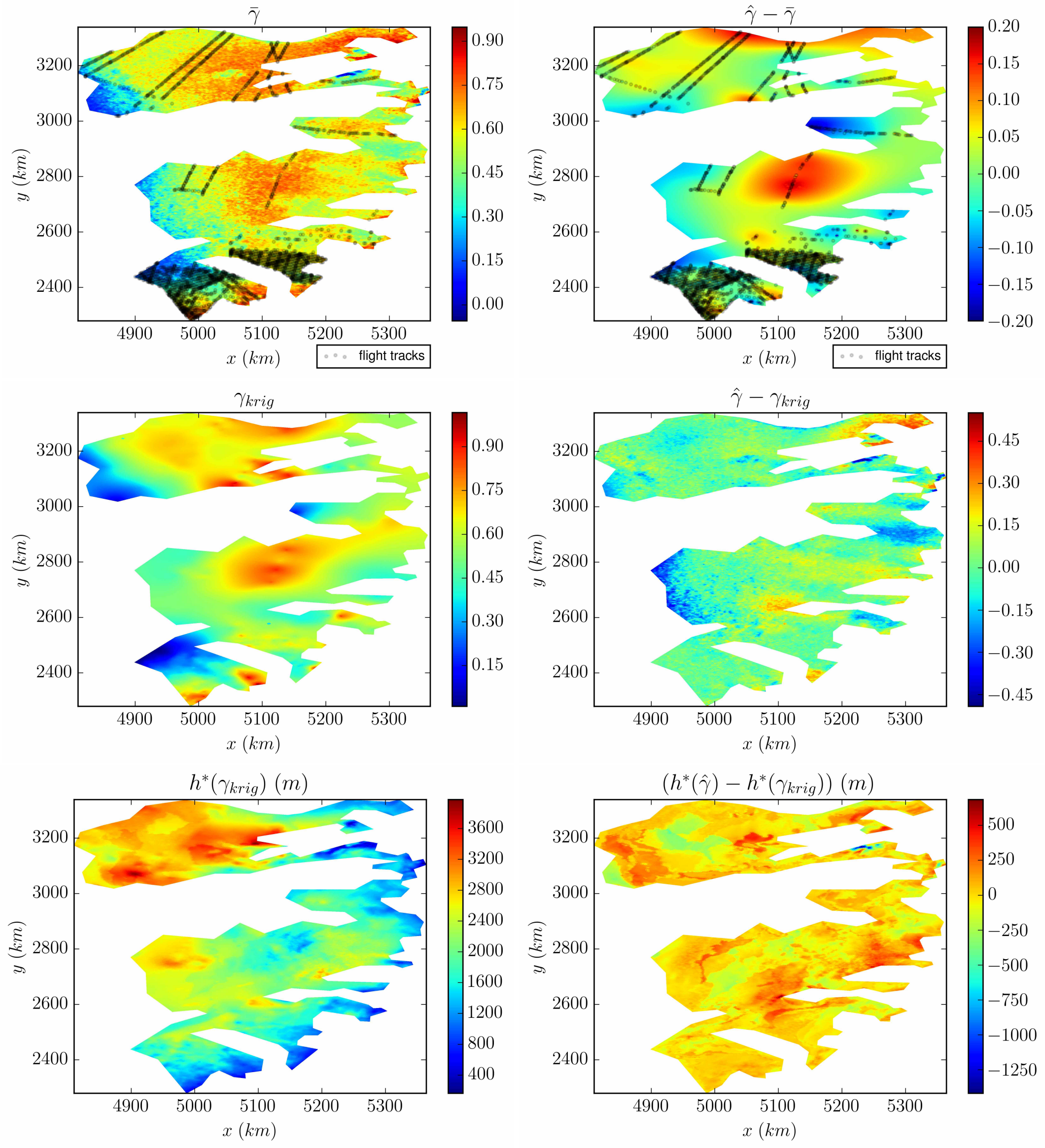

Fig. 11 Domain Ω_p=Ant3: comparison between different statistical learning methods at Step 2). (Up)(L) $\bar{\gamma}$ computed by ANN only. (R) Difference $(\hat{\gamma} - \bar{\gamma})$. (Middle)(L) γ_{krig} computed by ordinary Kriging. (R) Difference $(\hat{\gamma} - \gamma_{krig})$. (Down)(L) Infered thickness $h^*(\gamma_{krig})$. (R) Difference $(h^*(\hat{\gamma}) - h^*(\gamma_{krig}))$.

to "in-situ" measurements (flight tracks). Away from the flight tracks, the correction obtained to Bedmap2 may be significant or not, depending on the surface signature. The obtained corrections led to a total ice volume change of $0.5 - 5.6\%$, depending on the area.

In view to edit new bedrock maps, these bed elevation estimations should be more investigated by adding additional prior information, in particular regularisation length scales of h. In a mathematical point of view, this can be easily done by adding extra regularisation terms in the optimisation formulation(s). However the difficulty is to quantify such terms. This prior information may be introduced for example from geomorphologic analyses. Moreover the smoothing of the observational surface term $|U_H|/\mathscr{S}$ may be defined as non-isotropic to distinguish the streamline minimal wave length from the cross-line one.

This bed estimation method may be applied to any ice-sheets or ice-caps, as soon as the method domain of validity is compatible with the observed flows, that is highly to moderately sheared flows (i.e. not highly slipping flows). The method may be extended to unsteady flows if the surface observations (elevation and velocity) are given in time and assuming that the initial condition is either not important at the considered time scale or approximatively known.

Acknowledgements

Dr J. Zhu has been funded by a CNES TOSCA grant (oct. 2017-sept. 2018) during her post-doctoral stay at IMT - INSA Toulouse.
The first author acknowledges M. Morlighem (Univ. California of Irvine) and H. Seroussi (JPL/Caltech-NASA) for numerous and crucial discussions he had with them on real-world ice flows and datasets. The authors acknowledge for their comments which have helped to improve the presentation.

References

1. S. Arlot, A. Celisse, et al. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.
2. D. B. Bahr, W. T. Pfeffer, and G. Kaser. Glacier volume estimation as an ill-posed inversion. *Journal of Glaciology*, 60(223):922–934, 2014.
3. J. Bamber, J. Griggs, R. Hurkmans, J. Dowdeswell, S. Gogineni, I. Howat, J. Mouginot, J. Paden, S. Palmer, E. Rignot, et al. A new bed elevation dataset for greenland. *The Cryosphere*, 7(2):499–510, 2013.
4. JL Bamber, JL Gomez-Dans, and JA Griggs. A new 1 km digital elevation model of the antarctic derived from combined satellite radar and laser data–part 1: Data and methods. *The Cryosphere*, 3(1):101–111, 2009.
5. Andrew F Bennett. *Inverse modeling of the ocean and atmosphere*. Cambridge University Press, 2005.
6. M. Boutounet, J. Monnier, and J.-P. Vila. Multi-regime shallow free surface laminar flow models for quasi-newtonian fluids. *European Journal of Mechanics-B/Fluids*, 55:182–206, 2016.
7. D. Brinkerhoff, A. Aschwanden, and M. Truffer. Bayesian inference of subglacial topography using mass conservation. *Frontiers in Earth Science*, 4:8, 2016.
8. W.F. Budd and I.F. Allison. An empirical scheme for estimating the dynamics of unmeasured glaciers. In *Proceedings of the Moscow Symposium Snow and Ice*, pages 246–256, 1975.
9. G. Clarke, E. Berthier, C. Schoof, and A. Jarosch. Neural networks applied to estimating subglacial topography and glacier volume. *Journal of Climate*, 22(8):2146–2160, 2009.
10. L. Dalcín, R. Paz, and M. Storti. Mpi for python. *Journal of Parallel and Distributed Computing*, 65(9):1108–1115, 2005.
11. V Demyanov, M Kanevsky, S Chernov, E Savelieva, and V Timonin. Neural network residual kriging application for climatic data. *Journal of Geographic Information and Decision Analysis*, 2(2):215–232, 1998.
12. D. Farinotti, D. Brinkerhoff, D. Clarke, J. Fürst, H. Frey, P. Gantayat, F. Gillet-Chaulet, C. Girard, M. Huss, P. Leclercq, et al. How accurate are estimates of glacier ice thickness ? results from itmix, the ice thickness models intercomparison experiment. *The Cryosphere*, 11(2):949–970, 2017.
13. Christopher J Fogwill, Christian SM Turney, Katrin J Meissner, Nicholas R Golledge, Paul Spence, Jason L Roberts, Mathew H England, Richard T Jones, and Lionel Carter. Testing the sensitivity of the east antarctic ice sheet to southern ocean dynamics: past changes and future implications. *Journal of Quaternary Science*, 29(1):91–98, 2014.
14. P. Fretwell, H. Pritchard, D. Vaughan, J. Bamber, N. Barrand, R. Bell, C. Bianchi, R. Bingham, D. Blankenship, G. Casassa, et al. Bedmap2: improved ice bed, surface and thickness datasets for antarctica. *The Cryosphere*, 7(1), 2013.
15. C. Geuzaine and J.-F. Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331, 2009.
16. X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
17. R. Greve and H. Blater. *Dynamics of Ice Sheets and Glaciers*. Advances in Geophysical and Environmental Mechanics and Mathematics. Springer-Verlag, 2009.
18. G. Gudmundsson. Transmission of basal variability to a glacier surface. *Journal of Geophysical Research: Solid Earth*, 108(B5), 2003.
19. G. Gudmundsson. Analytical solutions for the surface response to small amplitude perturbations in boundary data in the shallow-ice-stream approximation. *The Cryosphere*, 2(2):77–93, 2008.
20. S.A. Haben, A.S. Lawless, and N.K. Nichols. Conditioning of incremental variational data assimilation, with application to the met office system. *Tellus A: Dynamic Meteorology and Oceanography*, 63(4):782–792, 2011.
21. C. Heining and M. Sellier. Direct reconstruction of three-dimensional glacier bedrock and surface elevation from free surface velocity. *AIMS Geosciences*, 2:45–63, 2016.
22. V. Helm, A. Humbert, and H. Miller. Elevation and elevation change of greenland and antarctica derived from cryosat-2. *The Cryosphere*, 8(4):1539–1559, 2014.
23. R. Hindmarsh. A numerical comparison of approximations to the stokes equations in ice sheet and glacier modeling. *J. geophysical research*, 109, 2004.
24. R. Hindmarsh. Stress gradient damping of thermoviscous ice flow instabilities. *Journal of Geophysical Research: Solid Earth*, 111(B12), 2006.
25. M. Huss and D. Farinotti. A high-resolution bedrock map for the antarctic peninsula. *The Cryosphere*, 8(4):1261–1273, 2014.
26. M.F. Hutchinson. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of hydrology*, 106(3-4):211–232, 1989.
27. D. Jansen, H. Sandhäger, and W. Rack. Model experiments on large tabular iceberg evolution: ablation and strain thinning. *Journal of glaciology*, 51(174):363–372, 2005.
28. B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6. Walter de Gruyter, 2008.
29. B. Kamb and K. Echelmeyer. Stress-gradient coupling in glacier flow: I. longitudinal averaging of the influence of ice thickness and surface slope. *Journal of Glaciology*, 32(111):267–284, 1986.
30. M. Kanevski, V. Timonin, and A. Pozdnukhov. *Machine learning for spatial environmental data: theory, applications, and software*. EPFL press, 2009.
31. D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
32. Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.
33. A. Lorenc. Optimal nonlinear objective analysis. *Quarterly Journal of the Royal Meteorological Society*, 114(479):205–240, 1988.
34. N. Martin and J. Monnier. Adjoint accuracy for the full-stokes ice flow model: limits to the transmission of basal friction variability to the surface. *The Cryosphere*, 8:721–741, 2014.
35. N. Martin and J. Monnier. Inverse rheometry and basal properties inference for pseudoplastic geophysical flows. *European Journal of Mechanics-B/Fluids*, 50:110–126, 2015.

36. L. Michel, M. Picasso, D. Farinotti, A. Bauder, M. Funk, and H. Blatter. Estimating the ice thickness of mountain glaciers with an inverse approach using surface topography and mass-balance. *Inverse Problems*, 29(3):035002, 2013.
37. L. Michel, M. Picasso, D. Farinotti, M. Funk, and H. Blatter. Estimating the ice thickness of shallow glaciers from surface topography and mass-balance data with a shape optimization algorithm. *Computers & Geosciences*, 66:182–199, 2014.
38. J. Monnier and P.-E. des Boscs. Inference of the bottom properties in shallow ice approximation models. *Inverse Problems*, 33(11):115001, 2017.
39. J. Monnier and J. Zhu. Inference of the bottom topography in anisothermal mildly-sheared shallow ice flows. *Computer Methods in Applied Mechanics and Engineering*, 348:954–977, 2019.
40. M. Morlighem, E. Rignot, J. Mouginot, H. Seroussi, and E. Larour. High-resolution ice-thickness mapping in south greenland. *Annals of Glaciology*, 55(67):64–70, 2014.
41. M. Morlighem, E. Rignot, H. Seroussi, E. Larour, H. Ben Dhia, and D. Aubry. A mass conservation approach for mapping glacier ice thickness. *Geophysical Research Letters*, 38(19), 2011.
42. M. Morlighem, C. Williams, E. Rignot, L. An, J.E. Arndt, J. Bamber, G. Catania, N. Chauché, J. Dowdeswell, B. Dorschel, et al. Bedmachine v3: Complete bed topography and ocean bathymetry mapping of greenland from multibeam echo sounding combined with mass conservation. *Geophysical research letters*, 44(21), 2017.
43. J. Mouginot, E. Rignot, B. Scheuchl, and R. Millan. Comprehensive annual ice sheet velocity mapping using landsat-8, sentinel-1, and radarsat-2 data. *Remote Sensing*, 9(4):364, 2017.
44. P.B. Price, O. Nagornov, R. Bay, D. Chirkin, Y. He, P. Miocinovic, A. Richards, K. Woschnagg, B. Koci, and V. Zagorodnov. Temperature profile for glacial ice at the south pole: Implications for life in a nearby subglacial lake. *Proceedings of the National Academy of Sciences*, 99(12):7844–7847, 2002.
45. U. Radok, D. Jenssen, and W. Budd. Steady-state temperature profiles in ice sheets. *Bull. Int. Assoc. Scient. Hydrol*, 8(1):36, 1970.
46. L. Rasmussen. Bed topography and mass-balance distribution of columbia glacier, alaska, usa, determined from sequential aerial photography. *Journal of Glaciology*, 34(117):208–216, 1988.
47. C. Schoof and C. Hindmarsh. Thin-film flows with wall slip: an asymptotic analysis of higher order glacier flow models. *Q. J. Mech. Appl. Math.*, 63(1):73–114, 2010.
48. M. Sellier. Inverse problems in free surface flows: a review. *Acta Mechanica*, 227(3):913–935, 2016.
49. H. Seroussi, M. Morlighem, E. Rignot, A. Khazendar, E. Larour, and J. Mouginot. Dependence of century-scale projections of the greenland ice sheet on its thermal regime. *Int. Glacio. Soc.*, 59:1024–1034, 2013.
50. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
51. W. Van Pelt, J. Oerlemans, C. Reijmer, R. Pettersson, V. Pohjola, E. Isaksson, D. Divine, et al. An iterative inverse method to estimate basal topography and initialize ice flow models. *The Cryosphere*, 7:987–1006, 2013.
52. J.M. van Wessem, W. Jan Van De Berg, B. Noël, E. Van Meijgaard, C. Amory, G. Birnbaum, C. Jakobs, K. Krüger, J. Lenaerts, S. Lhermitte, et al. Modelling the climate and surface mass balance of polar ice sheets using racmo2: Part 2: Antarctica (1979-2016). *The Cryosphere*, 12(4):1479–1498, 2018.
53. G. Wahba. *Spline models for observational data*, volume 59. SIAM, 1990.
54. R. Williams, R. Hindmarsh, and R. Arthern. Calculating balance velocities with a membrane stress correction. *Journal of Glaciology*, 60(220):294–304, 2014.
55. R. Winkelmann, M.A. Martin, M. Haseloff, T. Albrecht, E. Bueler, C. Khroulev, and A. Levermann. The potsdam parallel ice sheet model (pism-pik) part 1: Model description. *The Cryosphere*, 5(3):715–726, 2011.

A Results for other EAIS areas

In this section, the estimation of ice thickness is performed in the four other areas indicated in Fig. 4: Ant2, Ant4, Ant5 and Ant6. Ant2 is located upstream of Fisher and Merllor ice-streams (upstream Amery ice shelf). Ant4 is located upstream Totten ice-streams (Wilkes land and Terre Adlie), one of the largest discharger of ice in EAIS. Ant5 is located upstream Ninnis and Mertz ice-streams in Terre Adlie and George V land. Ant6 is located upstream Byrd ice-streams (east of Ross Ice shelf), see Fig. 4.

These areas have been relatively well covered during the airborne campaigns excepted the north-east part of Ant2, see Fig. 12. Distances between flight tracks are relatively low, therefore the empirical uncertainty assigned to Bedmap2 estimations h_b is low too: $\approx \pm[100-250]$m, see figures 12, 13, 14, 15 (Middle).

For each area, domain information and statistics on the numerical results are presented, see tables 8-11.

As already noticed, RU-SIA model already fits well with the surface topography after the data-driven model (Step 2) only; that is with $\hat{\gamma}$ and h_b as parameters in (1): see lines "($|H(h_b)-H_{obs}|$ (before h-inversion)" in the tables. Next RU-SIA model fits accurately the surface topography after the re-calibration / estimation of h. Indeed, misfit values range within $\approx [4-5]$m in mean; see lines "($|H(h_b)-H_{obs}|$ (after h-inversion)" in tables.

The corrections made on $\dot{a}$ are $\sim [10-17]\%$ in mean, that is within the a-priori uncertainty range indicated in [52].

The estimated thickness h^* is plotted in each case, see figures 12-15 (Down)(L); its difference with h_b is plotted (Down)(R). The corrections made to h_b are non negligible: they are ranging within $[7.0-15.9]\%$ in mean (corresponding to $[185-373]$ m, see the lines "$|h^*-h_b|$" in tables). These corrections lead to changes of total ice volume by $[1.5, 5.6]\%$. Again, the obtained difference with h_b is independent of the distance to the closest flight track. Maximum values of correction to h_b can be locally high. In the uncovered north-east Ant2 area, corrections are up to ≈ 2000m. At ≈ 50km from the nearest flight track (that is ~ 2 minimal wave lengths of the model) correction may reach 1000m (even in area surrounded by well covered areas), see e.g. figures 14 and 15 (Down)(R). Beyond ≈ 100km from the nearest data, the correction of h_b may be significant or not, depending on the surface signature, see e.g. figures 14 and 15 (Down)(R).

Table 8 Domain Ω_p=Ant2, information and results.

Domain Ω_p & mesh information			
Surface $\lvert\Omega_p\rvert$	$431860\,km^2$		
Mean ice thickness of h_b (Bedmap2)	$2144.4\,m$		
# mesh vertices: in Ω_p / on flight tracks	$65123/5194$		
RU-SIA model output (with $\hat{\gamma}$)	Median	Mean	Max
$\lvert H(h_b)-H^{obs}\rvert$ (before h-inversion)	$8.7\,m$	$17.3\,m$	$149.9\,m$
$\lvert H(h^*)-H^{obs}\rvert$ (after h-inversion)	$3.9\,m$	$5.1\,m$	$49.2\,m$
Infered RHS $\dot{a}$			
$\lvert\dot{a}^*-\dot{a}_b\rvert$	$0.6\,cm/y$	$0.8\,cm/y$	$4.0\,cm/y$
$\lvert\dot{a}^*-\dot{a}_b\rvert/\lvert\dot{a}_b\rvert$	14.8%	13.2%	20%
Infered thickness h			
$\lvert h^*-h_b\rvert$	$171.1\,m$	$302.4\,m$	$2025.6\,m$
$\lvert h^*-h_b\rvert/\lvert h_b\rvert$	8.3%	14.2%	80.0%
Ice volume change in km^3 / in %	5.2 $10^4\,km^3$ / 5.6%		

Table 9 Domain Ω_p=Ant4, information and results.

Domain Ω_p & mesh information			
Surface $\lvert\Omega_p\rvert$	$439045\,km^2$		
Mean ice thickness of h_b (Bedmap2)	$2745.4\,m$		
# mesh vertices: in Ω_p / on flight tracks	$61219/4977$		
RU-SIA model output (with $\hat{\gamma}$)	Median	Mean	Max
$\lvert H(h_b)-H^{obs}\rvert$ (before h-inversion)	$6.3\,m$	$8.4\,m$	$66.9\,m$
$\lvert H(h^*)-H^{obs}\rvert$ (after h-inversion)	$3.0\,m$	$4.1\,m$	$45.3\,m$
Infered RHS $\dot{a}$			
$\lvert\dot{a}^*-\dot{a}_b\rvert$	$2.0\,cm/y$	$2.7\,cm/y$	$15.0\,cm/y$
$\lvert\dot{a}^*-\dot{a}_b\rvert/\lvert\dot{a}_b\rvert$	8.7%	10.1%	20%
Infered thickness h			
$\lvert h^*-h_b\rvert$	$147.6\,m$	$185.0\,m$	$1241.6\,m$
$\lvert h^*-h_b\rvert/\lvert h_b\rvert$	5.7%	7.0%	50.4%
Ice volume change in km^3 / in %	1.5 $10^4\,km^3$ / 1.5%		

Table 10 Ant 5, information and results.

Domain Ω_p & mesh information			
Surface $\lvert\Omega_p\rvert$	$362019\,km^2$		
Mean ice thickness of h_b (Bedmap2)	$2415.3\,m$		
# mesh vertices: in Ω_p / on flight tracks	$41597/2351$		
RU-SIA model output (with $\hat{\gamma}$)	Median	Mean	Max
$\lvert H(h_b)-H^{obs}\rvert$ (before h-inversion)	$18.3\,m$	$23.9\,m$	$173.3\,m$
$\lvert H(h^*)-H^{obs}\rvert$ (after h-inversion)	$3.9\,m$	$5.1\,m$	$43.1\,m$
Infered RHS $\dot{a}$			
$\lvert\dot{a}^*-\dot{a}_b\rvert$	$3.9\,cm/y$	$4.4\,cm/y$	$14.5\,cm/y$
$\lvert\dot{a}^*-\dot{a}_b\rvert/\lvert\dot{a}_b\rvert$	18.8%	15.4%	20%
Infered thickness h			
$\lvert h^*-h_b\rvert$	$275.1\,m$	$373.1\,m$	$1989.2\,m$
$\lvert h^*-h_b\rvert/\lvert h_b\rvert$	12.1%	15.9%	68.8%
Ice volume change in km^3 / in %	3.8 $10^4\,km^3$ / 5.6%		

Table 11 Domain Ω_p=Ant6, information and results.

Domain Ω_p & mesh information			
Surface $\lvert\Omega_p\rvert$	$406388\,km^2$		
Mean ice thickness of h_b (Bedmap2)	$2672.9\,m$		
# mesh vertices: in Ω_p / on flight tracks	$63981/3012$		
RU-SIA model output (with $\hat{\gamma}$)	Median	Mean	Max
$\lvert H(h_b)-H^{obs}\rvert$ (before h-inversion)	$8.3\,m$	$11.0\,m$	$46.8\,m$
$\lvert H(h^*)-H^{obs}\rvert$ (after h-inversion)	$2.7\,m$	$3.4\,m$	$21.3\,m$
Infered RHS $\dot{a}$			
$\lvert\dot{a}^*-\dot{a}_b\rvert$	$0.4\,cm/y$	$0.5\,cm/y$	$1.8\,cm/y$
$\lvert\dot{a}^*-\dot{a}_b\rvert/\lvert\dot{a}_b\rvert$	15.8%	14.2%	20%
Infered thickness h			
$\lvert h^*-h_b\rvert$	$218.6\,m$	$313.9\,m$	$1777.3\,m$
$\lvert h^*-h_b\rvert/\lvert h_b\rvert$	8.1%	11.6%	63.8%
Ice volume change in km^3 / in %	2.6 $10^3\,km^3$ / 1.7%		

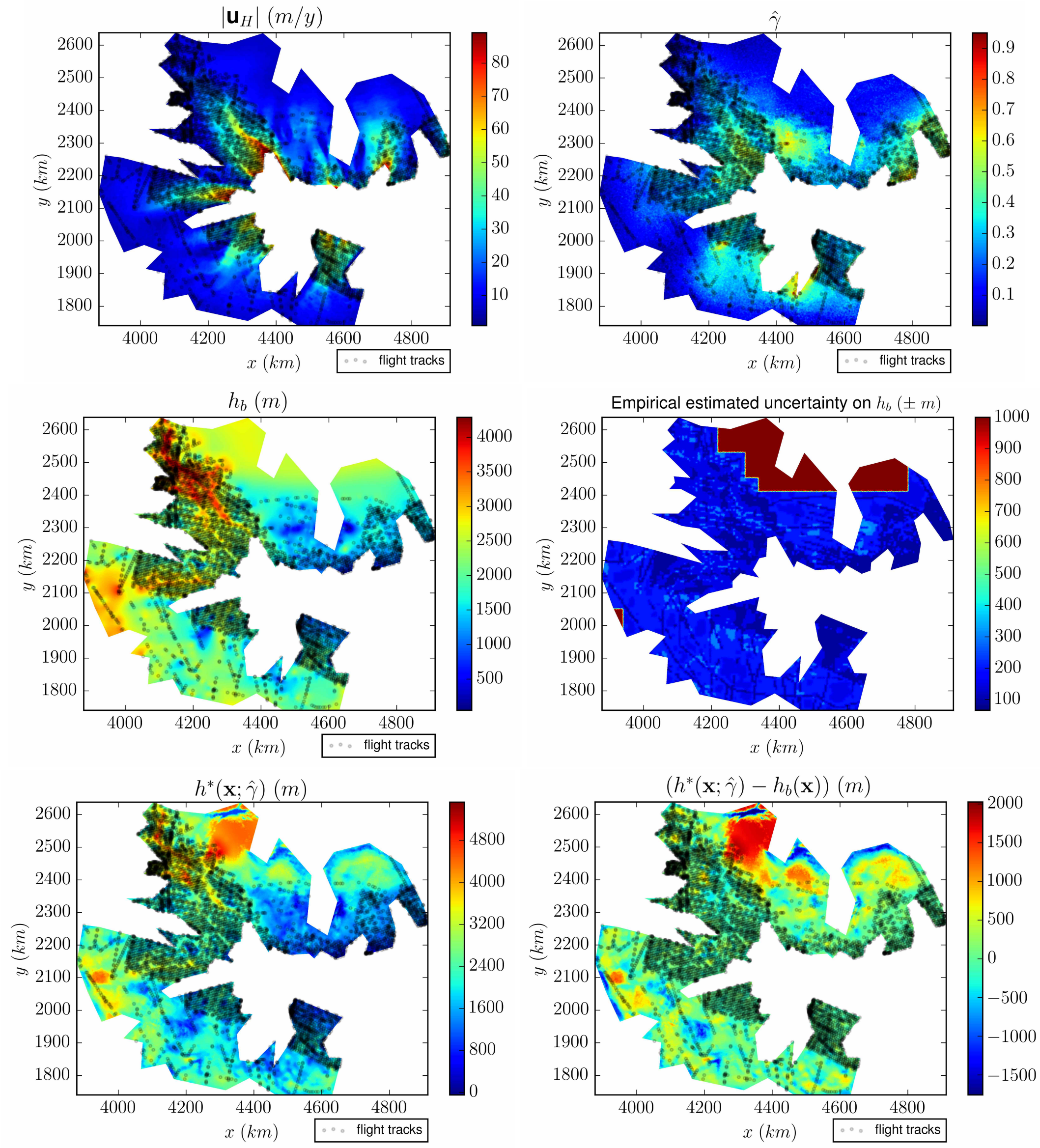

Fig. 12 Domain Ω_p=Ant2 (the plotted coordinates equal the Eastings-Northings plus $(2800,2800)$km): (Up)(L) Surface velocity module $|\mathbf{u}_H|$ and flight tracks. (R) $\hat{\gamma}$ computed by NNRK, see (11). (Middle)(L) Thickness h_b from Bedmap2 [14]. (R) Empirical uncertainty on h_b from [14]. (Down)(L) Infered thickness with $\hat{\gamma}$: $h^*(\hat{\gamma})$. (R) Difference $(h^*(\hat{\gamma}) - h_b)$.

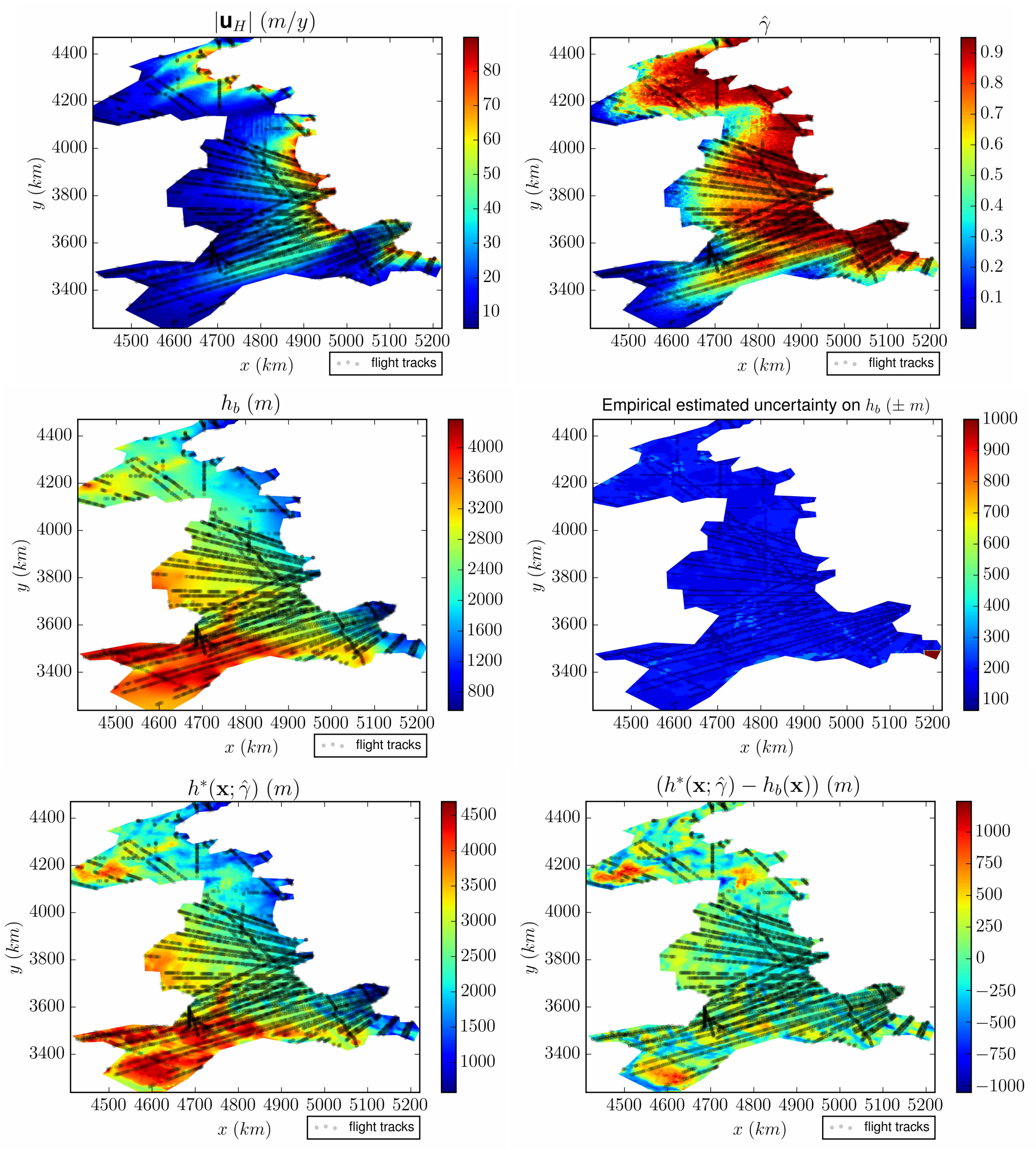

Fig. 13 Domain Ω_p=Ant4 (the plotted coordinates equal the Eastings-Northings plus $(2800, 2800)$km): (Up)(L) Surface velocity module $|\mathbf{u}_H|$ and flight tracks. (R) $\hat{\gamma}$ computed by NNRK, see (11). (Middle)(L) Thickness h_b from Bedmap2 [14]. (R) Empirical uncertainty on h_b from [14]. (Down)(L) Infered thickness with $\hat{\gamma}$: $h^*(\hat{\gamma})$. (R) Difference $(h^*(\hat{\gamma}) - h_b)$.

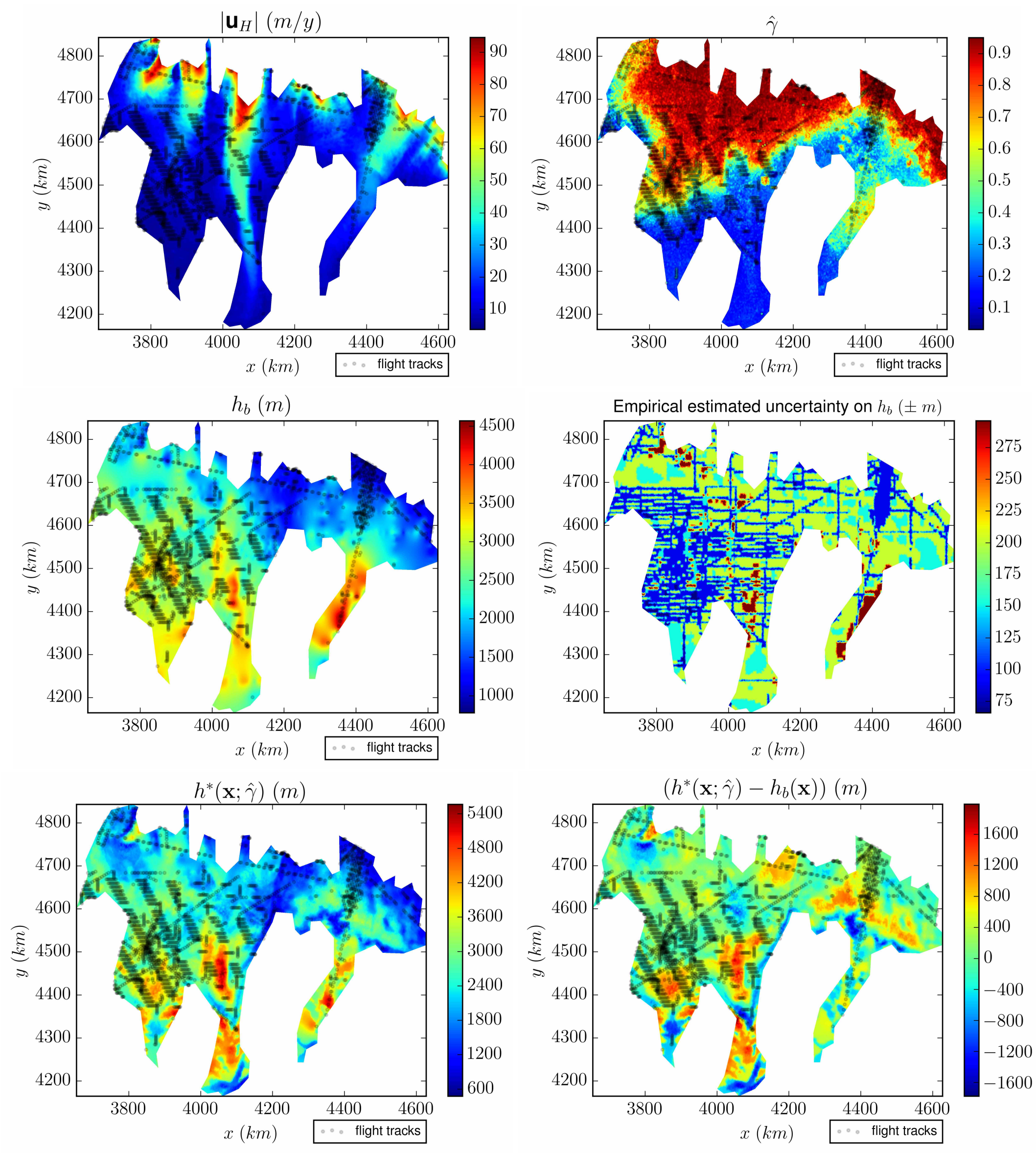

Fig. 14 Domain Ω_p=Ant5 (the plotted coordinates equal the Eastings-Northings plus $(2800,2800)$km): (Up)(L) Surface velocity module $|\mathbf{u}_H|$ and flight tracks. (R) $\hat{\gamma}$ computed by NNRK, see (11). (Middle)(L) Thickness h_b from Bedmap2 [14]. (R) Empirical uncertainty on h_b from [14]. (Down)(L) Infered thickness with $\hat{\gamma}$: $h^*(\hat{\gamma})$. (R) Difference $(h^*(\hat{\gamma}) - h_b)$.

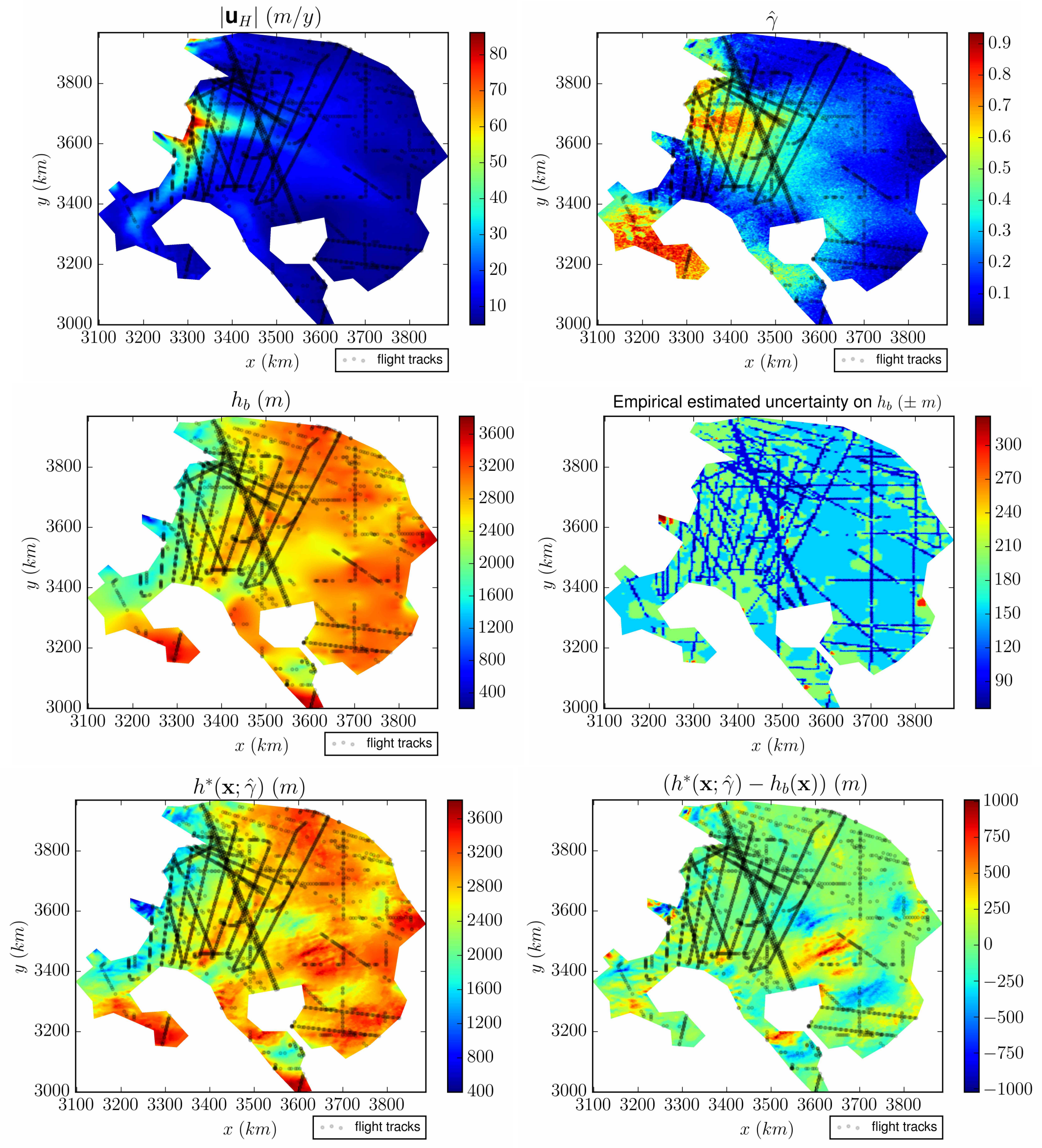

Fig. 15 Domain Ω_p=Ant6 (the plotted coordinates equal the Eastings-Northings plus $(2800,2800)$km): (Up)(L) Surface velocity module $|\mathbf{u}_H|$ and flight tracks. (R) $\hat{\gamma}$ computed by NNRK, see (11). (Middle)(L) Thickness h_b from Bedmap2 [14]. (R) Empirical uncertainty on h_b from [14]. (Down)(L) Infered thickness with $\hat{\gamma}$: $h^*(\hat{\gamma})$. (R) Difference $(h^*(\hat{\gamma}) - h_b)$.