Dear Editor, dear reviewers,

I really want to thank the two reviewers for their comments that I found very complementary and helpful to better explain the aims and results of this paper.

Both reviewer agreed that this paper has the potential of providing a good introduction of ensemble DA methods for the ice sheet modelling community.

To introduce the sequential DA algorithm and illustrate the difference with variational DA, I have added a new figure 1 (Fig. 1), that provides a schematic representation of the results expected with both methods.

Following the suggestions from the reviewers, I have mostly rewritten section 2.2 where I describe the assimilation setup. There is an abundant literature in geophysics on DA methods and it is not always easy to understand all the subtleties and differences between different methods. I hope that I have been successful now in giving a self-contained and clear introduction to the method.

To show that the synthetic experiment is realistic I have added a new Figure 3 (Fig. 2) where I compare the values for b and C used in this study to those obtained by Brondex et al. (2018) for Thwaites Glacier in Antarctica.

For the interpretation of the results I have performed additional experiments where I study the DA performance in retrieving the basal conditions as a function of the noise level in the surface observations (Figs. 8 and 9).

You will find below my point by point answer to the reviewer comments, with the original comments in black and my answers in red.

The manuscript with highlighted differences is at the end of the document.

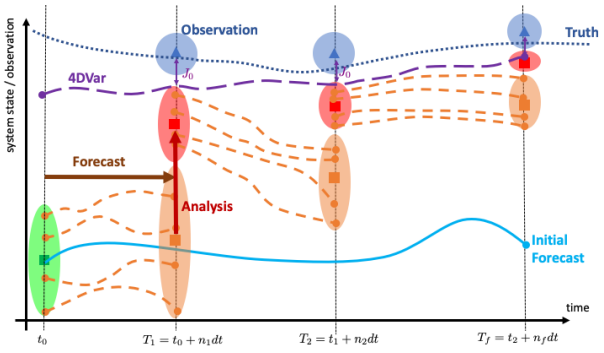

Figure 1. NEW FIGURE 1. Principle of data assimilation (Adapted from Carrassi et al. (2018)). Having a physical model able to forecast the evolution of a system from time $t = t_0$ to time $t = T_f$ (cyan curve), the aim of DA is to use available observations (blue triangles) to correct the model projections and get closer to the (unknown) truth (dotted line). In EnKFs, the initial system state and its uncertainty (green square and ellipsoid) is represented by N_e members. The members are propagated forward in time during n_1 model time steps dt to $t = T_1$ where observations are available (Forecast phase, orange dashed lines). At $T = t_1$ the analysis uses the observations and their uncertainty (blue triangle and ellipsoid) to produce a new system state that is closer to the observations and with a lower uncertainty (red square and ellipsoid). A new forecast is issued from the analysed state and this procedure is repeated until the end of the assimilation window at $t = T_f$. The model state should get closer to the truth and with lower uncertainty as more observations are assimilated. Time dependent variational methods (4D-Var) iterate over the assimilation window to find the trajectory that minimizes the misfit (J_0) between the model and all observations available from t_0 to T_f (violet curve). For linear dynamics, Gaussian errors and infinite ensemble sizes, the states produced at the end of the assimilation window by the two methods should be equivalent (Li and Navon, 2001).

REVIEWER 1, Dan Goldberg

This study is one of the first to apply Ensemble Kalman Filtering methods to an ice-sheet model with a nontrivial stress balance (attempts have been made with Shallow- ice models). Such methods, rather than using deterministic means to optimise a cost/misfit function in order to infer hidden properties from error-prone observations, essentially generate an ensemble meant

to encompass a probability distribution, and repeatedly apply model dynamics and bayesian inference on this ensemble in order to refine the statistical properties of an unknown state and parameter set. The paper implements a variant of the EnKF known as the Ensemble Subspace Transform Kalman Filter, which is simply a particular choice and one which is meant to avoid costly computation of an ensemble covariance (which i do question – please see specific comments). The methodology presented is really just one step in a long long road toward operational ice-sheet forecasting (compare with over a half-century of development in numerical weather prediction) but an important one – especially when considering there is at least one person with a question about filtering methods every time a talk is presented in state and parameter estimation for ice sheets (as the authors have pointed out, there are already others working toward the ice-sheet version of the other main tool of weather prediction, 4Dvar). Thus i feel it is a worthwhile study which should be worthy of publication as a methodological investigation (and the authors frame it as such, at least in the conclusions section. However I do think the manuscript needs work before this can happen. On the basis of the extent of the comments below I choose "major revisions" – but there is no formal definition of what this means, and the editor may choose to ignore this classification. I am not suggesting modification of the algorithm and/or results, simply clarity of text.

I thanks Dan Goldberg for the time spend on trying to understand the description of the method. His comments where very helpful to point parts that where unclear. I totally agree that there is still along road toward operational ice-sheet forecasting, and that will require to *i)* develop/adapt new methodologies but also *(ii)* improve the data bases and the characterisation of the observation uncertainties as pointed in my discussion. Indeed, I see this paper as a methodological investigation to see and present the potential of the method.

Indeed there is several deterministic variants of the Ensemble Kalman Filter. All are based on the Kalman update equation for the analysis. Keeping the covariance matrix in a square root factorisation form, allows to derive an expression for the transformation of the forecast ensemble to the analysed ensemble. As different ensembles could have the same mean and covariance matrix, different implementations will lead to different analysed ensembles and thus different solutions. As mentioned in the text, ESTKF is closely related to the SEIK and ETKF filters that are widely used in oceanography and meteorology. In the original paper that derives the ESTKF, the implementation of ESTKF is estimated to be slightly more efficient that the ETKF, while it shares the same interesting properties of minimal transformation. However the two filters should lead to very similar solutions. I hope that the description of the filter has been clarified in the new version.

GENERAL COMMENTS

For one thing upon reading I had significant detail understanding what was done. There were a number of points on which i felt clarity was needed, and most of these are addressed below in line-by-line comments so I will not list them here. Note that these specific questions compose the bulk of my review – and this is because without having a better idea of what was actually implemented, it is difficult to critique the results further!

Thanks. See my point by point answer below.

However something I will state in the general comments is that despite similarities, ice-sheet models are distinct from e.g. atmospheric models in that the unknown parameters most sought cannot generally be observed directly, in contrast to models in which the initial conditions in a forecast/analysis cycle represent the parameters of greatest interest. This is exemplified by the fact that the “state” vector contains non-dynamic variables (friction and bed elevation) and the fact that the observation operator, rather than being a simple averaging or restriction, encapsulates a fully nonlinear solve of an elliptic partial differential equation. I think this is something that should be made very clear to members of the climate and meteorological community who read this work.

It is relatively classical that not all the variables are observed, even for dynamic state variables; and EnKFs are also widely used for state and parameter estimation. I have clarified the fact that the observation operator is the force balance equation and thus highly non-linear..

One overall comment is regarding the distribution of the ensemble. As I understand it, even if the initial ensemble is evenly distributed given the prior, it is difficult to know a pri-ori whether the projection of the ensemble will represent a favorable distribution of the projected space. That is, what if the forecast “clusters”, underrepresenting important regions of state space? As I ask below, it is unclear whether there is a “reinitialisation” of the ensemble at every step. Clearly this topic has already

been considered in the NWP literature, for example Song et al (2013) makes use of a time-dependent adjoint (a tool the authors state the work here is meant to circumvent the need for) in order to generate a more representative ensemble.
The algorithm leads to a transformation of the forecast ensemble to the new analysed ensemble, that has the mean and covariance matrix given by the original Kalman update equation. Yes with small ensembles, because the covariances matrices are given by a "small" ensemble that can not represent the full error-space, the analysis tends to be over-confident resulting in ensembles with a spread that becomes to small. On the long run the analyses may start to ignore the observations, causing an ensemble that will diverge from the truth. Localisation and inflation are pragmatic remedies to this undersampling issue. Indeed combining ensemble and variational methods will lead to more robust estimation but at an higher cost. I have clarified the presentation of the localisation and inflation methods, and state in the introduction that the tendency is to combine variational and ensemble methods to take the advantages of both. The objective of this first study is to use tools that are widely use in other domains to test their applicability for ice-sheet modelling. Complexity can be added in future studies when the main weaknesses of the method will be clearly identified; so I'm not aiming to give here a detailed discussion on hybrid ensemble/var methods.
The new section 2.2.2 now reads:

2.2.2 Filter stabilisation: inflation and localisation

In practice for large scale problems, EnKFs as Monte-Carlo methods can suffer from under-sampling issues. First, because of the rank deficiency of the covariance matrix $\mathbf{P}^f$*, the analysis adjust the model state only in the error subspace, ignoring error directions not accounted for by the ensemble (Hunt et al., 2007). This can result in an analysis that is overconfident and underestimates the true variances. On the long run, the ensemble spread will become too small and the analysis will give to much weight on the forecast finally disregarding the observations and diverging from the true trajectory. A common simple* ad-hoc *remedy is to inflate the forecast covariance matrix with a multiplicative factor (Pham et al., 1998; Anderson and Anderson, 20 199). Here, inflation has been introduced in Eq. (**??**) using the forgetting factor* $\rho \in [0,1]$ *with* $\rho = 1$ *corresponding to no inflation (Pham et al., 1998). It is the inverse of the inflation factor used by Bonan et al. (2014).*
Second, the rank deficiency of $\mathbf{P}^f$ *leads to the appearance of spurious correlations between parts of the system that are far away. As these correlations are usually small, a common remedy is to damp these correlations with a procedure called localisation. In covariance localisation, localisation is applied by using an ensemble covariance matrix that results from the Schur product of* $\mathbf{P}^f$ *with an* ad-hoc *correlation matrix that drops long range correlations (Hamill et al., 2001; Houtekamer and Mitchell, 2001). However, this localisation technique is not practical for square-root filters where* $\mathbf{P}^f$ *is never explicitly computed. Here, as in Bonan et al. (2014), we use a localisation algorithm based on domain localisation and observation localisation (Ott et al., 2004; Hunt et al., 2007). Both methods are illustrated in Sakov and Bertino (2011) who conclude that they should yield to similar results. Domain localisation assumes that observations far from a given location have negligible influence. In practice, the state vector in each single mesh node is updated independently during a loop through the nodes that can easily be parallelized for numerical efficiency. For each local analysis, only the observations within a given radius* r *from the current node are used. In addition to avoid an abrupt cut-off, the observation error covariance matrix* $\mathbf{R}$ *is modified so that the inverse observation variance decreases to zero with the distance from the node using a fifth-order polynomial function which mimics a Gaussian function but has compact support (Gaspari and Cohn, 1999). Because it drops spurious long-range correlations and allows the local analyses to choose different linear combinations of the ensemble members in different regions, localisation implicitly increases the rank of the covariance matrix, leading to a larger dimension of the error subspace, implicitly increasing the effective ensemble size and the filter stability(Nerger et al., 2006; Hunt et al., 2007). However, it has been reported that localisation could produce imbalanced solutions (Mitchell et al., 2002). Here, because the force balance are non-inertial and the SSA assumes that the ice-shelves are in hydrostatic equilibrium, this shouldn't be an issue. Another disadvantage is that, when long-range correlations truly exist, the analysis will ignore useful informations that could have been used from distant observations.*
Here, the forgetting factor ρ *and the localisation radius* r *will be used as tuning parameters of the filter. Improving the theoretical understanding of these* ad hoc *procedures and developing adaptive scheme is an active research area and interested readers can refer to review articles ((e.g. Bannister, 2017; Carrassi et al., 2018; Vetra-Carvalho et al., 2018)*

I also question whether the "toy problem" proposed by the authors truly tests all of the difficulties a filtering approach might encounter. I bring this up in more detail below but some aspects of the approach seem to hang on the "locality" of the problem (a technique called "localisation" is employed to ignore long-distance correlations of the state). I wonder if this only works because the problem is one-dimensional with no buttressing involved, so essentially to a strong degree (though not completely) the velocities depend locally on basal friction and geometry? Would this still be a good approach in a 2D domain with an expansive embayed ice shelf (such as the Ross or FRIS), or very weak basal traction over a large part of the domain (such as Pine Island)?

As explained in the new version, localisation ignores observations that are far away for the local analyses as, often, long range correlation, are poorly estimated due to the sampling. In the case where true long range correlations may exist it will simply ignore this so the analysis may not use the full informations contained in the observations. The lower computational cost due to a small ensemble comes at a price. However I don't think that the 1D experiment ignore this problem. On the contrary, because the shelf is unbuttressed we may expect that the velocities on the shelf are extremely correlated with the basal conditions at the GL. Localisation is now better explained but I agree that more work will be required in the future to see if this can be improved. Following the suggestions from reviewer 2, I know compare the values used in this study with those given by Bonan et al. (2014):

We have used inflation and localisation to stabilise the filter. The inflation giving the best results in Bonan et al. (2014) ($\rho = 0.87 - 1.02$) is similar to the values tested in this study. For the localisation radius r we have used values between 4 and 16 km, while it ranges from 80 to 120 km in Bonan et al. (2014). While this seems counter-intuitive as the velocities depends only on the local conditions with the shallow ice approximation used by Bonan et al. (2014), in fact, because we use a different grid size ($dx = 0.2$km compared to $dx = 5$km in Bonan et al. (2014)), for each node we assimilate twice as much observations. Our results are in agreement with the adaptive localisation radius proposed by Kirchgessner et al. (2014) . Using three different models, Kirchgessner et al. (2014) have shown that good performances are obtained when r is such that the effective local observation dimension, defined as the sum of the weights attributed to each observation during the local assimilation, is equal to the ensemble size. Here, the value $r = 8$km used for the 50 members-ensemble corresponds to an effective observation dimension of 56. Future studies should investigate if this result can be transposed to realistic 2D simulations with unstructured meshes.

The methodology essentially uses a whole "family" of geometries and velocities to infer hidden parameters of the system. This somewhat bears similarity to a different paper led by the author, "Assimilation of surface velocities acquired between 1996 and 2010 to constrain the form of the basal friction law under Pine Island Glacier" – aside from the statistical formality, and the introduction of consistency between these geometries by way of the continuity equation (which is actually not so consistent if the analysis updates do not conserve mass!!!) – I wonder if the author would consider comparing and contrasting these approaches.

In the paper you mention, this was a collection of "snapshot" inversions where we assumed a constant sliding coefficient between consecutive observed geometries and indeed we didn't use the continuity equation. It would be interesting to compare both method for the reconstruction part, i.e. the performances for retrieving constant basal properties, but I think this would make a full paper. The advantage of the method used here is that we have, at the end of the assimilation window, a transient model that has been initialised using the available observations and we can make projections. So I prefer to discuss this approach in the context of initialising a transient ice flow model, similarly to 4D-Var, and I think a discussion between these approaches will add more confusion. Concerning the mass conservation, yes the analysis does not conserve mass as the ice-sheet volume is uncertain because the geometry is uncertain. However we expect that the uncertainty is reduced as observations are assimilated and we hope that the volume we have at the end of the assimilation window is a good estimator of the true volume, obtained by the combination of the model and the observations. However, you are right that if we are interested by a re-analysis of the volume change during the assimilation window a 4D-Var or a smoother might be more appropriate. I have tried to clarify this point and hope that the new figure 1 clarifies the concept of sequential DA.

I have added a new paragraph to discuss this point in section 3.3 Assimilation set-up:

Because both z_s and b are included in the state vector, the analysis does not conserve the ice sheet volume, neither for the ensemble mean and the individual members. However, the estimation of the ice-sheet volume is improved at each analysis as more data are assimilated, and the final volume is the best estimation provided by the filter knowing the model, all the observations during the assimilation window and their uncertainties. As mentioned in the introduction, if the main interest is

an analysis of past volume changes, as smoother might be more appropriate and the smoother extension of the ESTKF can be found in Nerger et al. (2014).

The distinction between EnKF, 4DVar and smoother is introduced in the introduction as:

EnKF approximates the state and the error covariance matrix of a system using an ensemble that is propagated forward in time with the model, avoiding the computation of the covariance matrices and the use of linearised or adjoint models. Contrary to time-dependent variational methods where the objective is to find the model trajectory that minimizes the difference with all the observations within an assimilation window, EnKF assimilates the observations sequentially in time as they become available using the analysis step of the Kalman Filter, as illustrated in Fig. 1. The model trajectory is then discontinuous and, at a given analysis, the model is only informed by past and present observations. For the retrospective analysis of a time period in the past, i.e. a reanalysis, ensemble filters can easily be extended to smoothers to provide analyses that are informed by all past, present and future observations (Evensen and van Leeuwen , 2000; Li and Navon, 2001; Cosme et al., 2012; Nerger et al., 2014)

Finally, i point out that, despite the divide between filtering and adjoint-based methods, there is a growing sentiment in NWP to take what is "best" from the various approaches and form more hybrid schemes (for instance the Song paper referenced above, see also Kalnay 2010). Therefore I urge the author to reflect on such innovations and how they might be useful in further developments for filtering of ice-sheet models.

I now mention this in the introduction, but I think first we have to test the performances of individual methods to identify the weak points, so going to far in a discussion of what we might expect by using hybrid methods is premature.

Line by Line comments

- p2.l3: would be good to state this is a point when only resolving 1 horizontal dimension

  this has been changed to:
  Improving SLR estimates requires, amongst others, to correctly model the dynamics of the grounding line (GL), i.e. the location where the ice detaches from its underlying bed and goes afloat on the ocean, implying that this is a line in 2D-plane view and a point in 1D..

- p3.10 variational

  done

- p3.13 "use of linearised or adjoint models" this assumes a trivial mapping from model vars to observations – see my comments below.

  This point has been clarified. If it is true that the linearised observation operator is used in the KF update equation, in practice we do not need to compute it as it always act as an operator to project the ensemble members in the observation space, and we make the classical linear approximation

  $$\mathbf{Y}^f = \mathbf{H}\mathbf{X}^f \tag{1}$$

  with $\mathbf{X}^f = (\boldsymbol{x}_1^f, ..., \boldsymbol{x}_{N_e}^f) \in \mathbb{R}^{N_x \times N_e}$ the forecast ensemble matrix and $\mathbf{Y}^f = (\boldsymbol{y}_1^f, ..., \boldsymbol{y}_{N_e}^f) \in \mathbb{R}^{N_y \times N_e}$ its equivalent in the observation space where $\boldsymbol{y}_i^f = \mathcal{H}(\boldsymbol{x}_1^f)$, $i = 1, ..., N_e$. This point has been clarified in section 2.2

- p3.15 rewrited?

  yes thanks.

- p3.35 – would be good to explain as soon as possible i.e. here what you mean by a twin experiment, or give a reference, as this is jargony

  The meaning of twin experiment is now introduced in the introduction:
  In the context of ice-sheet modelling, encouraging results have been obtained by **?** *for the estimation of the state and basal conditions of an ice-sheet model using the Ensemble Transform Kalman Filter (ETKF,* **??***). They study the performance*

of the method using idealised twin experiments where perturbed observations generated from a model run are used in the DA framework to retrieve the true model states and parameters.

- p 5 thru eq (10): this is a well written explanation of the EnKF. However I have a few questions which might be due to my lack of familiarity with filtering methods, but I think this might be true of many readers of this paper. This is also important as, though it is not the algorithm used, the one used is far more complex so this is a chance to explain your methods to the reader.

  (a) is $P\hat{}f = P_k$?

  Yes I mention that I omit the time index k as all the analysis is done at a given time t_k. However, ESTK rewrites P^f as a function of X^f and Ω as explained in the new version.

  (b) you do not say how the individual ensemble members ($x_i\hat{}k, a$) arise/are updated, only the state vector (which looks like the mean of the analysed ensemble)?

  it is now clear that the update equation for the mean and covariance is rewritten to give a unique equation for the transformation of the forecast members x_i^f to the analysed members x_i^a

  (c) is the posterior/analysis covariance used at all in subsequent time/filtering steps, as from eq 7 the covariance is always formed from the present ensemble – so i am struggling to grasp what is done in the algorithm in a multi-time step (k>1) framework.

  I now clearly mention that the analysed ensemble is used as the initial ensemble for the next forecast, and so on until the end of the assimilation window. I also think that the new figure 1 helps to clarify this point.

  (d) For each new forecast/analysis cycle, is the ensemble generated anew from the analysis-generate ensemble statistics?

  See reply above, the analysed ensemble is directly obtained for the analysis equation and can be used as the initial ensemble for the next forecast.

  (e) the formula given assumes normality of the ensemble does it not?

  As mentioned in the introduction the KF analysis is optimal for Gaussian distributions and it only uses the first moments of the distribution, mean and covariance.

- P5 eq(8): M_k is trivially the identity on the time-invariant components of x, i.e. b and C, correct?

  Yes, I now explicitly mention that I assume persistence for the parameters during the forecast step, in sec 3.3:
  The state vector is augmented by the two parameters to be estimated, the bedrock topography b and the basal friction coefficient C. For the parameters we assume a persistence model, i.e. no time evolution, during the forecast step (Eq. 8).

- p6.4. I am struggling to see why $P\hat{}f$ need be formed, as it is a tensor product of X with itself (subject to (a) above). For instance, the last term in 9(a) is written

$$X(X\hat{}TH\hat{}T)((HX)(X\hat{}TH\hat{}T)+R)\hat{}-1 \tag{2}$$

  so the largest matrix that need be formed is HX, and no matrix of (Nx x Nx) need be formed. Perhaps I do not understand where and how P^a is actually used however.

  Yes this is the basis of the implementation, because the covariance matrices can not be formed, we keep the square-root factorisation. This allows to obtained an expression for the transformation of the forecast ensemble to the analysed ensemble that exactly has the covariance matrix P^a. So we don't have to explicitely compute and store P^a.

- P6.5 I don't feel the concept of "error subspace" is ever suitably explained as i read the paper still wondering about this. Ω as defined in eq 11 simply seems to be a "mixing" matrix that slightly changes the ensemble members – how is this an "error subspace"? (Assumiing that $X \in R\hat{}NxxNe$ – i take it this is the case for eq 11 to make sense...)

This is now explained as follow:
Moreover, the sample covariance matrix approximated with an ensemble of size N_e (Eq. 7) is only a low-rank approximation of the true covariance matrix and its rank is at most $N_e - 1$. ESTKF uses this property to write the analysis in a $(Ne-1)$-dimensional subspace spanned by the ensemble and referred to as the error subspace (Nerger et al., 2005a) The forecast covariance matrix $\mathbf{P}^f$ is then rewritten as

$$\mathbf{P}^f = \frac{1}{N_e - 1}\mathbf{L}\mathbf{L}^T \tag{3}$$

where $\mathbf{L} \in \mathbb{R}^{N_x \times N_e - 1}$ is given by

$$\mathbf{L} = \mathbf{X}^f \mathbf{\Omega} \tag{4}$$

The matrix $\mathbf{\Omega} \in \mathbb{R}^{N_e \times N_e - 1}$ defined as

$$\Omega_{ij} = \begin{cases} 1 - \frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i = j,\ i < N_e \\ -\frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i \neq j,\ i < N_e \\ -\frac{1}{\sqrt{N_e}} & \text{for } i = N_e \end{cases} \tag{5}$$

projects the ensemble matrix $\mathbf{X}^f$ onto the error subspace. The multiplication with $\mathbf{X}^f$ subtracts the ensemble mean and a fraction of the last column of the ensemble perturbation matrix $\mathbf{X}'^f$ from all other columns.

- Eqs 11-16: in contrast to the discussion of the EnKF this is very nonintuitive. You state (P6 line 7) that you approximate the covariance matrix by a low-rank matrix, which seems intuitive, but where is the equation describing this low-rank approximation and how is it done? (for what it is worth, low-rank approximations of covariance generally involve eigenvalue decompositions to retain the leading order covariance structure, but i do not see this here...

  See reply above. It is a low rank matrix as it is approximated by an ensemble of size N_e, so the rank of the sample covariance matrix is at most $N_e - 1$, while, in principle, the true covariance matrix could be full-rank.

- P6.23-25: can you give a more intuitive description of inflation? Why do you need it and what does it achieve? As it is I am not even sure if inflation corresponds to lower or higher ρ.

  Following suggestions from reviewer 2, I now give better explanations of inflation and localisation in section 2.2.2. See reply above.

- P7, first paragraph. I'm sorry but I am struggling to follow this paragraph. For instance, how does the non-linear observation operator applied to x_i lead to the product $HX\hat{}f$? i imagine they are related, as H is a linearisation of $\mathcal{H}$ (which, by the way, clashes with the symbol for ice thickness) but this is not explained.

  See reply above. This is now better explained and I use $\mathcal{H}$ for the non-linear observation operator.

- P7.8. This assumption, i imagine, is valid in many NWP settings given the hyperbolicity of the equations. Are you confident it is a good approach for marine ice-sheet modelling?

  Also in NWP you may have long-range correlations. Localisation is better explained and presented as a pragmatic way to counteract under-sampling issues. As you mention this paper is a first step toward operational DA and more work will be required on these aspects.

- P8.26. I was surprised by your suggestion that annual DEMs would be available over a multiple decadal period, as i do not know of such products for antarctica. The best i have seen is decadal or semidecadal with MUCH lower spatial

resolution (e.g. Konrad et al 2017, GRL). Having skimmed the ArcticDEM website i do see mention of the spatial resolution, but not the temporal. Unless you can argue that such spatiotemporal resolution is reasonable and available, i suggest caveating this discussion by saying it is an idealised experiment and this is the type of spatiotemporal resolution to which the community should aspire.

I agree that we don't have this kind of product yet for decadal periods, but it should become available quickly and you can now find DEMs for the Greenland Ice sheet with a with a 3-month temporal resolution (https://nsidc.org/data/nsidc-0715). This has been updated.

- P9.6: prognostic ice sheet models generally step forward the ice thickness, not surface elevation (as shown in your eq 4). In the analysis step you are updating z_s. Is there a simple mapping from your model state X to thickness?

  Yes I now clearly mention that the floatation equation is used for the mapping between the ice thickness and the free surface elevation. As mentioned in the discussion, with a full-stokes model that solves the two free surfaces you will have to put both free surfaces as state variables.

- P9.6: as mentioned in the previous comment you are updating z_s in each analysis step, which i am inferring then maps on to an update in thickness (tell me if i am wrong). Is this update at all volume conserving? If not should this be a concern?

  See reply above. Because this is a sequential algorithm the analysis does not conserve volume because the volume is uncertain, i.e. both the free surface and bed elevation are in the state vector. This is not a concern as this is the aim of DA to improve the estimation of the system state by using the observations. Again, for a re-analysis it might be more appropriate to use a smoother instead of a filter to interpret past volume changes.

- P9.21-28: Lots of jargony language in this paragraph, likely not to be understood by the target audience. What is a sill and a nugget? You talk of the prediction obtained by kriging – is this something you have calculated? Is there any way to evaluate whether the ensemble does converge to it? Is this a way of evaluating whether the ensemble is large enough?

  It is better explained now. The method directly draw realisations, i.e. members, from the distribution that would be obtained from kriging. This is not an exact sampling method, i.e. the mean and covariance of the ensemble will not exactly match the mean and covariance of the kriging prediction. As the spatial correlation is already given by a model, i.e. the analytical variograms, we want an initial ensemble that is representative of what we think is the initial uncertainty given the available a-priori or observations, and this is what we get.

- Section 4.1: Upon reading this, I realised that (a) i am unsure what time step you used, and (b) more importantly, whether each time step is a forecast/analysis step, as M_k in eq 8 could easily encompass multiple time steps – is this the case?

  I now mention that the time step is 0.005 a and the time interval between two analysis is 1 year. So yes it is clear now that we can have multiple model time steps between two analyses.

- P10.15: again, these factors seem very important, and as mentioned above not overly well explained.

  see reply above

- Section 4.2 – section headings should be capitalised

  Done

- P12.25: Two comments about this paragraph: (a) Code that is continuously being up- dated and new algorithms developed might be an issue for *analytically* derived adjoint models, but not as much for automatic differentiation, which is specifically designed to generate new adjoint code when the "primal" code is changed; and (b) I return to the my confusion over the first paragraph on P7. Your observation matrix H contains, at the very least, a linearisation of the stress balance equation mapping geometry and basal friction onto velocity. It is not clear how you are finding this operator if not through some sort of forward model linearisation.

ok I have removed this sentence. I agree that with automatic code differentiation, in principle it should be relatively straightforward; however, I think most people will still agree that maintaining an adjoint code remains difficult in practice for research codes. As explained above there is no linearisation required in the method used here.

0.1 References:

Song et al, 2013: An adjointbased adaptive ensemble Kalman filter. Mon. Wea. Rev., 141, 3343–3359, https://doi.org/10.1175/MWR-D-12-00244.1.
Kalnay, E. (2010), Ensemble Kalman Filter: Current status and potential, in Data Assimilation: Making Sense of Observations, edited by W. Lahoz, B. Khattatov, and R. Menard, 24 pp., Springer, New York

ANONYMOUS REVIEWER 2

First of all, I would like to apologise to the author and the editor for providing my review so late. I consider this paper treats an important subject in an innovative manner and, as such, deserves a careful review. I hope you will find my review insightful.
Yes many thanks for this insightful review. Please find below my point by point answer.

OVERVIEW

The paper aims to adapt an Ensemble Kalman Filter (EnKF) to estimate jointly the surface elevation, the bedrock topography and the basal friction coefficient in the case of a flowline marine ice sheet. Time-dependant ensemble data assimilation (DA) approaches are relatively new for ice sheet initialisation. The paper focuses on the case of a grounding line retreat for unstable glaciers, which is a hot topic in ice sheet modelling and climate change. It also studies the influence of DA on forecasts of grounding line retreat.

GENERAL COMMENTS

The paper targets an important and timely question: how to initialise and estimate basal parameters to forecast the evolution of marine ice sheets and glaciers especially in the case of an unstable retreat? and proposes to use an EnKF (here an ESTKF) in this context. EnKFs have shown how efficient they can be in a wide range of applications (not exclusively meteorology and oceanography but also hydrology, crop modelling, oil extraction, pollutant dispersion, . . .), are a good alternative to adjoint-based methods and are the basis of hybrid methods that are now popular in DA (see e.g. Bannister, 2017). The paper shows clearly that EnKFs are a good toolbox for DA in glaciology. The experiment shows clearly how beneficial this approach could be with adequate figures and a very insightful analysis. Overall I am convinced that the paper in its final form will be a very good introduction of ensemble DA methods in ice sheet modelling.
Nevertheless, there are few points that prevent me to publish the paper as is. I list them below:

- I have some reservation about the experimental setup mainly on how the reference run is designed and on how the basal friction coefficient C is estimated.
  - About the reference run: The long-term objective of the study is to forecast accurately grounding line retreat for Antarctic hemispheric glaciers in the context of global change. However, in the reference run of the experiment, the grounding line retreat is triggered by the abrupt change of ice rigidity B. I wonder if the simulated grounding line retreat is “realistic” compared to one triggered by climate change. Also I have the same question about the sinusoidal basal friction parameter C. How realistic is it compared to real cases? I know we are in flowline cases using SSA equations. But it would strengthen the paper if the author could reflect on that subject in the section 3.1. Part of my comment may be due to my lack of knowledge in glaciology, so please accept my apologies if my comment is irrelevant.
    I now compare the synthetic values for b and C with values obtained by Brondex et al. (2018) in Thawaites Glacier Antarctica (Fig. 2). This shows that the values are realistic. Note however that the mesh resolution in Brondex et al.

(2018) varies from $\approx$ 200m close to the GL to 10km 100km upstream, so that the comparison for the wavelengths are really meaningful only in the first tens of kilometres. For the initial retreat I have added the following discussion: *In Jenkins et al. (2018), observed ice-flow accelerations in the Amundsen sea sector have been attributed to the decadal oceanic variability, where warm phases associated with increased basal melt induce a thinning of the ice shelves reducing their buttressing effect initiating short lived periods of unstable retreat of the most vulnerable GLs. In a flow line experiment the ice shelf do not exert any buttressing effect. Using a suite of melting and calving perturbation experiments for Pine Island Glacier, Favier et al. (2014) have shown that, when initiated, the dynamics of the unstable retreat is fairly independent of the type and magnitude of the perturbation. Here, to trigger the initial acceleration, we instantaneously decrease the ice rigidity to* $B = 0.3$ *MPa* a^{-3} *at* $t = 0$*, keeping all the other parameters constant.*

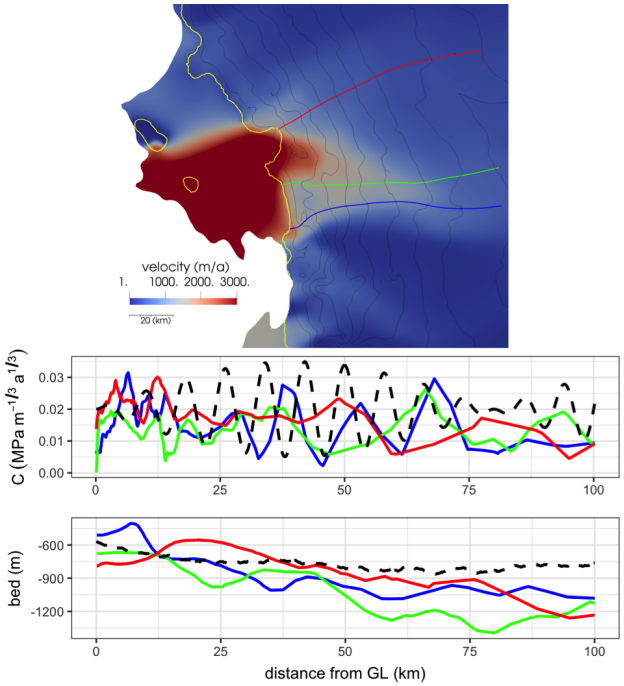

Figure 2. NEW FIGURE 4. Thwaites Glacier (Antarctica). Model results from **?**: model velocities (top) and friction coefficient C and bed elevation b extracted along three streamlines (same color code). Synthetic values used in this study are shown with black dashed lines. Note that the mesh resolution varies from $\backsim$ 200m close to the GL, shown in yellow in the top panel, to $\backsim$ 10km at the upstream end of the streamlines.

– About the assimilation setup: The basal friction coefficient C must be positive. To ensure such thing, you use the following change of variable $C = \alpha^2$. But by doing so, you do not ensure the unicity of the estimation as α and $-\alpha$ would lead to the same C. Also I thought that the C parameter could be of different order of magnitude. To counteract those potential issues, Bonan et al. (2014) chose the following change of variable $C = 10^{\alpha}$ ensuring the positivity of C, the unicity of the change of variables and mimicking behaviours with different order of magnitude. Could you explain why the change of variable you used is more appropriate in your context?

There was no particular reason, only that the reference only span one order of magnitude. To my knowledge I was one of the first to introduce the change of variable $C = 10^{\alpha}$. I have performed a new experiment with this change of variable to see the differences. The results a given in Figure 3. It can be seen that the performances are extremely similar. However, for C the RMSE is a little higher (0.0045, instead of $0.004\,\mathrm{Mpa\,m^{-1/3}\,a^{1/3}}$), while

there is nearly no difference for the velocities. I think this can be explained by the fact that the sensitivity of the model velocities to the remaining uncertainty in C is lower that the observation uncertainty so the DA can not really discriminate the two reconstructions. To see this effect I present in the new version of the paper, additional experiments where I study the error on the retrieved values as a function of the observation noise. see answers below.

The part describing the change of variable as been updated as follow:

Because the velocities are insensitive to the basal conditions where ice is floating, these two parameters are included in the state vector only for the nodes where at least one member is grounded. In addition, to insure that C remains positive, we use the following change of variable for the assimilation $C = \alpha^2$. Although it does not insure uniqueness of the estimation as α and $-\alpha$ would lead to the same C, this change of variable is classical (Mac Ayeal (993) and was chosen as the reference friction coefficient spans only one order of magnitude. Similar performances where found using the other classical change of variable $C = 10^{\alpha}$ as in Gillet-Chaulet et al. (2014).

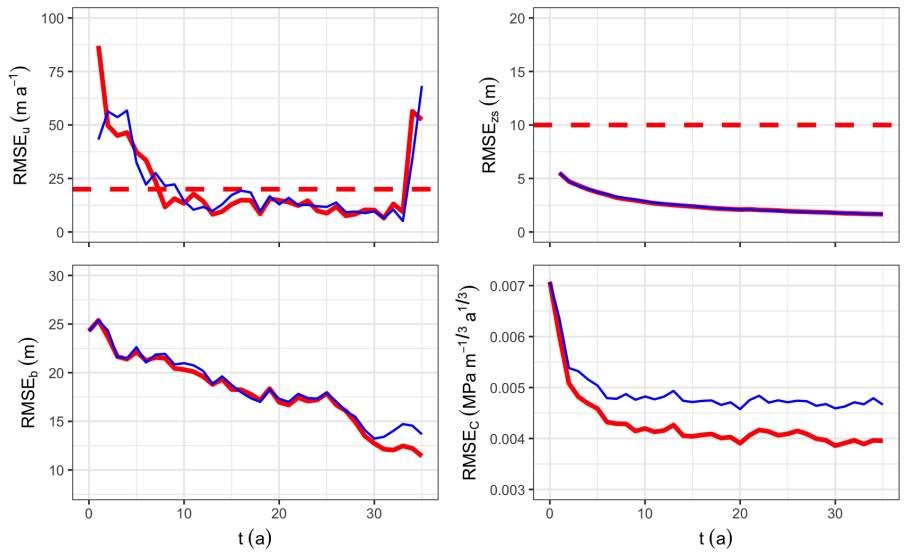

Figure 3. RMSEs after analysis for the assimilation up to $t = 35$a. Results shown in the paper with he change of variable $C = \alpha^2$ are in red; results with the change of variable $C = 10^{\alpha}$ are in blue.

- The paper is full of interesting results but sometimes they deserve a more thorough analysis.
  - The assimilation window is between 1 yr and 35 yr (you run forecasts from analysed states at t = 20 yr and t = 35 yr). But the grounding line is almost steady between t = 13 yr and t = 32 yr. We also see that after the first 10 years of assimilation, RMSD for C remains stable. I wonder if those two points are correlated meaning is it easier or more difficult to estimate C in the case of a retreating grounding line (hence a more dynamic ice sheet) or an almost steady state? As the behaviour of grounding line is different from one marine glacier to another, it would be beneficial to the community if you could push your study in that direction in the revised version of the manuscript (for example continuing DA after t = 35 yr for the next ten years and study what happens).

    We don't see more improvement if the assimilation is pursued up to $t = 50$a. My understanding is that the sensitivity of the observations to the remaining uncertainties is already below the noise level, especially for the velocities, so we still see an improvement for the few kilometres upstream of the grounding line but this do not reflects on the values for the RMSE for C and b, that are computed from $x = 300km$ up to the position where at least one member is grounded (Fig. 4 and 5). However, we still see an improvement for the forecast as shown in Figs. 6 and 7

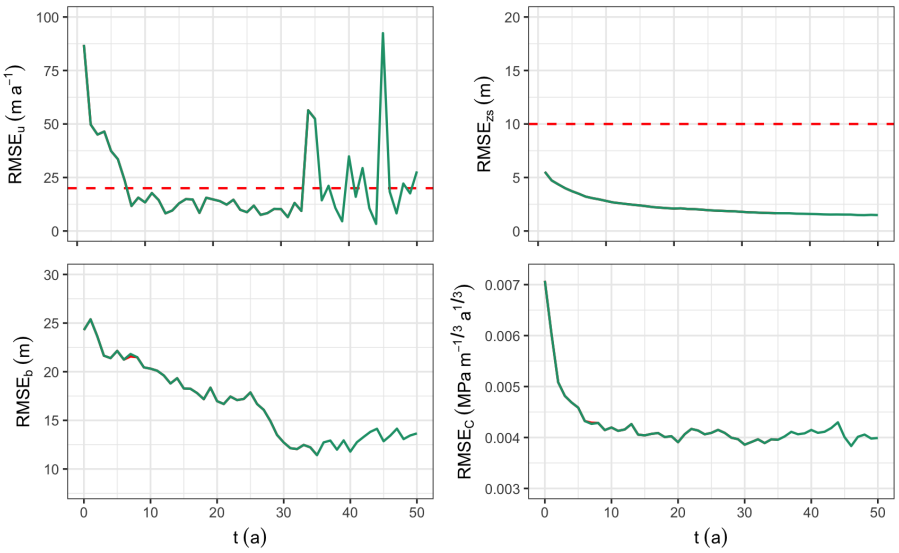

Figure 4. RMSEs after analyses for an assimilation up to $t = 50$a.

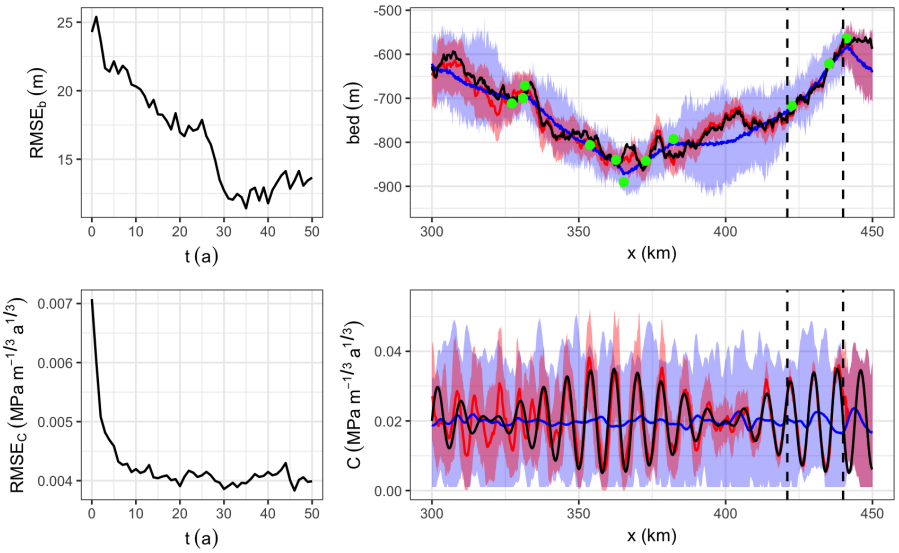

Figure 5. Same as Fig. 3 but or an assimilation up to $t = 50$a.

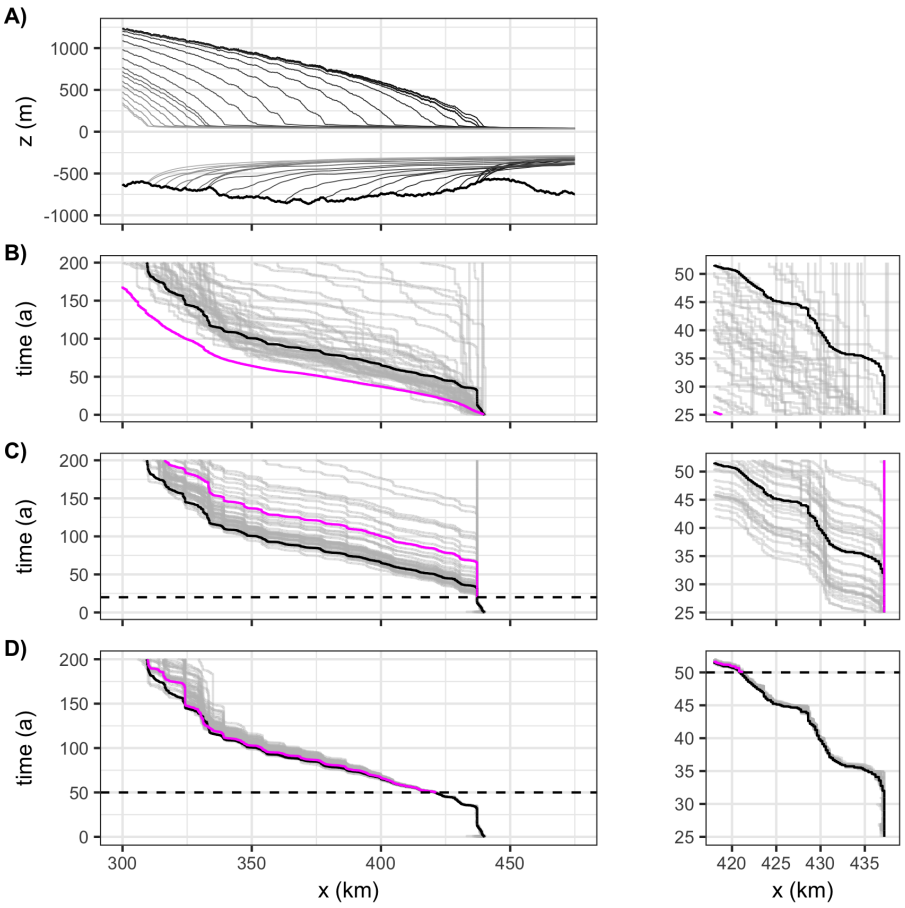

Figure 6. Same as Fig.2 but for an assimilation up to $t = 50$a.

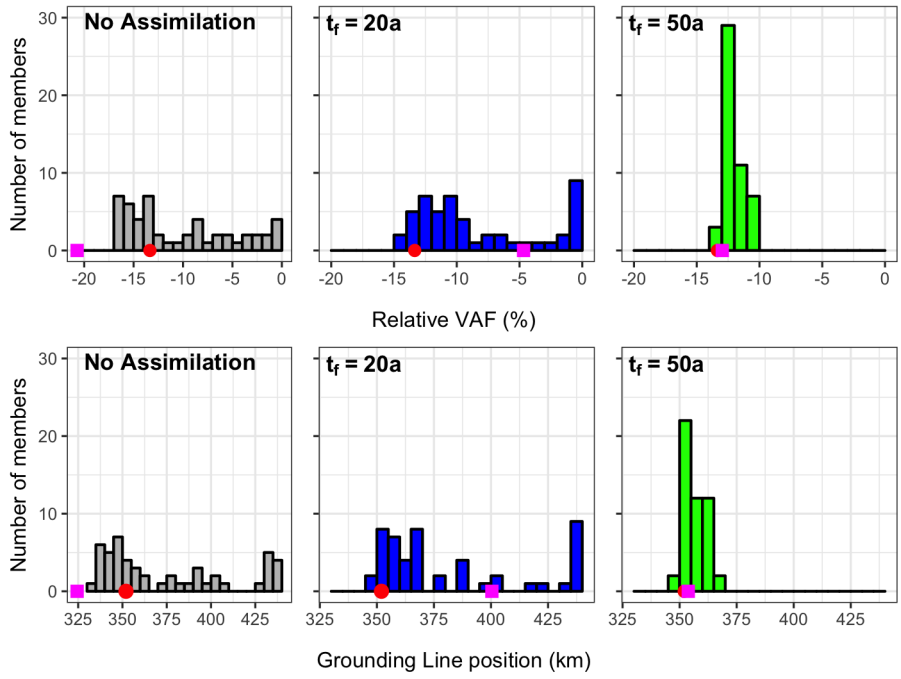

Figure 7. Same as Fig. 12 but or an assimilation up to $t = 50$a.

I have added two additional experiments where I vary the noise level for the observed velocities and the surface elevation (Figures 8 and 9). This shows that decreasing the uncertainty on the observed velocities improves the RMSE for b and C, but the results for the reconstructed velocities are not significant, so that 2 ensembles with slightly different RMSE have very similar differences for the velocities and the reconstruction stagnates. Changing the noise for the observed surface elevation has a small effect and in fact $RMSE_C$ increases for the lowest noise levels but this do not reflect on the velocities.

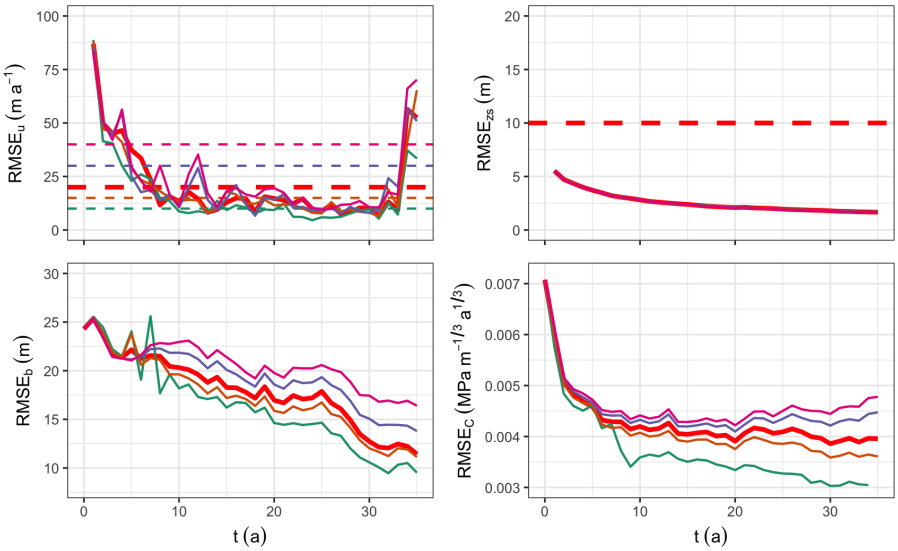

Figure 8. NEW FIGURE 10.Sensitivity to the surface velocities observation error σ_u^{obs} : RMSEs after each analysis, computed only for $x \geq 300km$ for b and C.. The thick red lines correspond to the results with $\sigma_u^{obs} = 20$ ma^{-1} and $\sigma_{z_s^{obs}} = 10$ m shown in Figs. **??** and **??**. The horizontal dashed lines correspond to the observation errors σ_u^{obs} and the results are presented with solid lines using the same color code. $\sigma_{z_s^{obs}} = 10$ m for all the experiments.

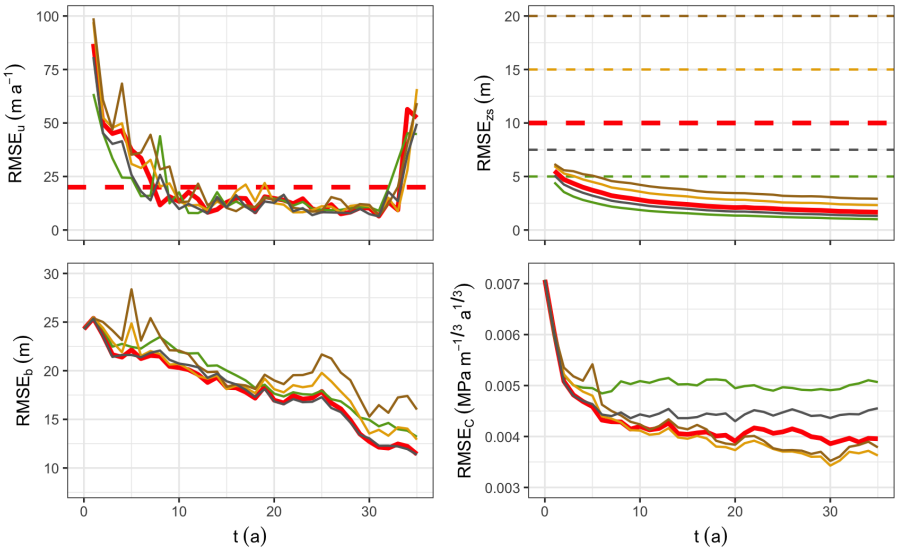

Figure 9. NEW FIGURE 11. Sensitivity to the surface elevation observation error $\sigma_{z_s^{obs}}$: RMSEs after each analysis, computed only for $x \geq 300km$ for b and C. The thick red lines correspond to the results with $\sigma_u^{obs} = 20$ ma^{-1} and $\sigma_{z_s^{obs}} = 10$ m shown in Figs. **??** and **??**. The horizontal dashed lines correspond to the observation errors $\sigma_{z_s^{obs}}$ and the results are presented with solid lines using the same color code. $\sigma_u^{obs} = 20$ ma^{-1} for all the experiments.

– Figure 5 shows interesting results about the performance of ESTKF with inflation and localisation and varying sizes of the ensemble. Bonan et al. (2014) has performed the same kind of studies for grounded ice sheet using an ETKF. Do you obtain similar optimal parameters for inflation and localisation in your experiment or are they different from Bonan et al. (2014)? It would be interesting to reflect on how the physics of the model influences such parameters.

I have added the following discussion:

We have used inflation and localisation to stabilise the filter. The inflation giving the best results in Bonan et al. (2014) ($\rho = 0.87 - 1.02$) is similar to the values tested in this study. For the localisation radius r we have used values between 4 and 16 km, while it ranges from 80 to 120 km in Bonan et al. (2014). While this seems counter-intuitive as the velocities depends only on the local conditions with the shallow ice approximation used by Bonan et al. (2014), in fact, because we use a different grid size (dx = 0.2km compared to dx = 5km in Bonan et al. (2014)), for each node we assimilate twice as much observations. Our results are in agreement with the adaptive localisation radius proposed by Kirchgessner et al. (2014). Using three different models, Kirchgessner et al. (2014) have shown that good performances are obtained when r is such that the effective local observation dimension, defined as the sum of the weights attributed to each observation during the local assimilation, is equal to the ensemble size. Here, the value $r = 8$km used for the 50 members-ensemble corresponds to an effective observation dimension of 56. Future studies should investigate if this result can be transposed to realistic 2D simulations with unstructured meshes.

– You only show results after 300 km (nothing between 0 km and 300 km). Does it mean that what happens between 0 km and 300 km does not have an influence on the grounding line retreat? Does it mean DA is pointless in those areas (in that case, that would make DA more affordable as less grid points need to be treated)? If so, please state it more clearly and if not, provide more results for that area.

Yes as shown by Durand et al. (2011), we expect that uncertainties in the ice-sheet interior should not affect short-term forecast of the coastal regions. However for completeness, I have added a figure that shows the results in the

first 300km. Because the noise level on the observed velocity is close to 100% the is only little improvement for b. It is better for C. Also the ensemble spread is only slightly reduced. My understanding is that the sensitivity of the observations to the initial uncertainty is smaller than the noise level.

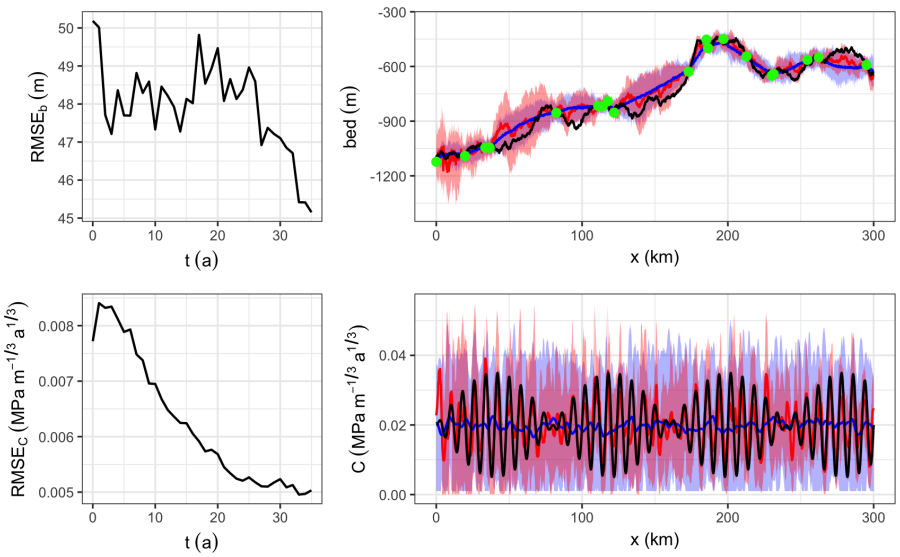

Figure 10. NEW FIGURE 8. Same as Fig.3. but with the RMSE computed for $x \in [0, 300]$ km.

– I am worried by some of the results shown for ESTKF for small ensembles (Ne = 30) as RMSD reduction is only 20% for the bedrock topography and 30% for the basal friction coefficient. For 2D real cases (either using Full-Stokes or SSA), we need an ensemble approach working for small ensembles due to the cost of the experiment and 30 members might be too expensive. Could you reassure me and provide the reader that the approach would be beneficial even in 2D real cases?

Results are similar to Bonan et al. (2014), the performances start to deteriorate with ensemble sizes ≤ 50. However, I'm optimistic that we shoul be able to run 2D application, at least at the scale of a drainage bassin with ensemble sizes N_e at the order of 50 to 100. See the following sentences in the discussion:
Good results have been obtained with relatively small ensembles (50 to 100 members) for a state vector of size $N_x \approx 8400$ and $N_y = 4002$ observations. Similarly to Bonan et al. (2014), we still see an improvement with a 30-members ensemble but the performances to retrieve the basal conditions are not as good. Running 2D plane view simulations with such ensemble sizes is largely possible as demonstrated by Ritz et al. (2015) who, using an hybrid shallow ice-shallow shelf model, have run a 200 years ensemble forecast of the whole Antarctic Ice Sheet using 3000 members.

– About the forecast experiments, Figure 1 shows clearly how beneficial an ensemble forecast can be compared to a deterministic forecast. It also shows that the distribution of the grounding line position is not Gaussian. Could you provide more information (maybe using histograms) on how grounding lines are distributed in the ensemble?

Thanks for the suggestion; I now give the following figure that shows the histograms from the relative volume above floatation (VAF; i.e. what matters when looking at the contribution of ice sheets to sea level rise) and GL position.

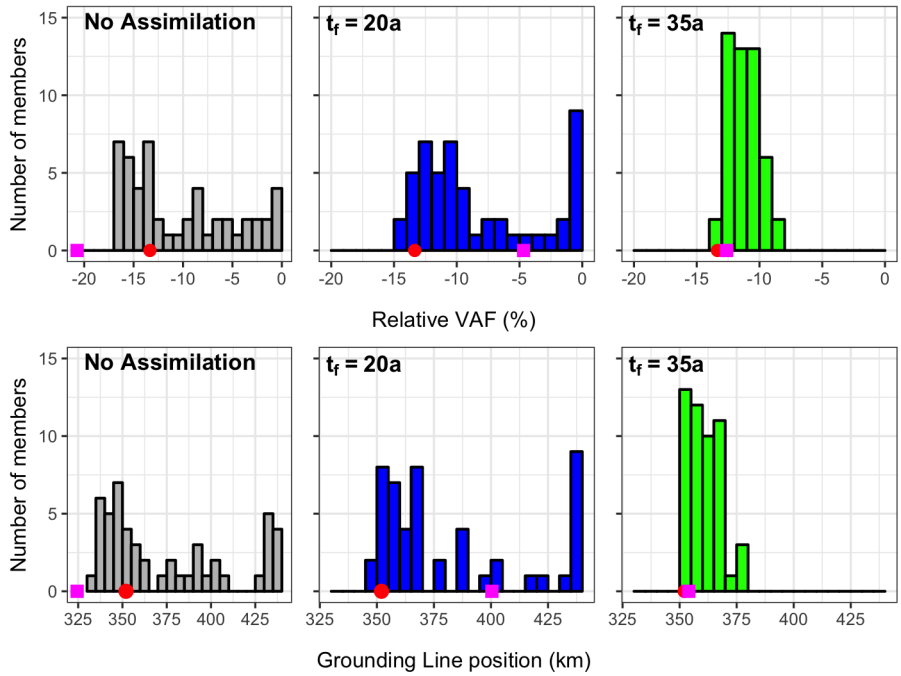

Figure 11. Ensemble forecast at $t = 100$ a: (top) relative change of volume above floatation (VAF) and (bottom) GL position with (left) no assimilation, (center) assimilation up to $t = 20$ a and (right) assimilation up to $t = 35$ a. The red circle correspond to the reference run and the magenta square to the deterministic forecast.

- While on average, the paper is perfectly readable. There are few parts that are hardly accessible to the reader. As I think the target audience of this paper is the ice sheet community and more generally the glaciology community, the paper would benefit from a careful editing. I detail those parts in the Specific comments section.

  Thanks for these specific comments; see the specific replies below.

Overall I consider the paper as a highly valuable addition to data assimilation for ice sheets. But it deserves mostly some rewriting. Therefore I recommend a minor revision as almost all the science is already here!

SPECIFIC COMMENTS

Overall I think it would be nice for the reader to have two tables summarising the various variables used in this paper, one for the ice flow model and one for data assimilation.

Good suggestion; I now have the following two tables in appendix

Table 1. NEW APPENDIX TABLE A1: Notations and values used in this study associated with the ice flow model

Prognostic variables:		
$H = z_s - z_b$	m	Thickness
z_s	m	top surface elevation
z_s	m	bottom surface elevation
Diagnostic variable:		
u	$\mathrm{m\,a^{-1}}$	horizontal velocity
Parameters:		
$a_b = 0.0$	$\mathrm{m\,a^{-1}}$	basal melting
$a_s = 0.5$	$\mathrm{m\,a^{-1}}$	surface accumulation
b	m	bed elevation
$B = 0.4$	$\mathrm{Mpa\,a^{1/3}}$	ice rigidity
C	m	basal friction coefficient
$m = 1/3$		friction law exponent
$n = 3$		Glen's creep exponent
$\rho_i = 900$	$\mathrm{kg\,m^3}$	ice density
$\rho_w = 1000$	$\mathrm{kg\,m^3}$	sea water density
Numerical parameters:		
$dt = 5\,10^{-3}$	a	model time step
$dx = 200$	m	mesh resolution

Table 2. NEW APPENDIX TABLE A2: Notations and values used in this study associated with the ensemble filter

Variables:		
$\boldsymbol{x} = (z_s, b, C)$		state vector
$\mathbf{P}$		covariance matrix
Stabilisation parameters:		
r	m	localisation radius
ρ		forgetting factor
Sizes:		
N_e		ensemble size
N_x		state vector size
N_y		observation vector size
Others:		
$\Delta t = 1$	a	time interval between two analyses

About Ensemble Kalman Filters:

- I feel localisation and inflation should be better explained in the paper (either in the introduction and in the DA section). I agree they are both used to counteracts the effects of undersampling. But the reader would benefit from having more information. Undersampling causes underestimated variances (counteracted by inflation) and spurious correlations (counteracted by localisation in the case of long-range spatial spurious correlations). Could you add few lines on the subject. Also few references are missing. For inflation:

  Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, doi: 10.1175/1520- 0493(1999)127<2741/AMCIOT>2.0.CO;, 1999.

  For localisation, the first one is for local analysis (the one you use), the other two are for covariance localisation (the historical one):

  Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, A.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus A, 56, 415–428, doi: 10.1111/j.1600-0870.2004.00076.x, 2004.

  Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent fil- tering of background error covariance estimates in an ensemble Kalman filter, Mon. Weather Rev., 129, 2776–2790, doi: 10.1175/1520- 0493(2001)129<2776/DDFOBE>2.0.CO;, 2001.

  Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, doi: 10.1175/1520-0493(2001)129<0123/ASEKFF>2.0.CO;, 2001.

  I now have a section *2.2.2 Filter stabilisation: inflation and localisation* where inflation and localisation are better described using the given references. See reply to reviewer 1.

- Also about inflation, the term "forgetting factor" introduced by Pham et al. (1996) is, unfortunately, very uncommon in the EnKF community. Could you state somewhere that this is just the inverse of the traditional inflation parameter known widely in the EnKF community?

  I now state that the forgetting factor ρ is the inverse of the inflation factor used by Bonan et al. (2014).

- p. 6, l. 23: There is no unicity of the symmetric square root matrix C. It is known that the choice of C can have a significant impact on results (see e.g. Livings et al., 2008). Could you detail how C is calculated in PDAF?

  Livings, D. M., Dance, S. L., and Nichols, N. K.: Unbiased ensemble square root filters, Physica D, 237, 1021–1028, doi: 10.1016/j.physd.2008.01.005, 2008.

  I now mention that $\mathbf{C}$ is the symmetric square root of $\mathbf{A}$ obtained by singular value decomposition. I have added a remark to say that $\mathbf{C}$ could also be computed from a Cholesky decomposition.

- p. 7, first paragraph on how to use the ESTKF with a nonlinear observation operator. It is a very good point you raise especially in the case of assimilating surface ice velocities (highly nonlinear observation operator). However, I find it difficult to see where the nonlinearity of the observation operator intervenes. Could you rewrite the whole section 2.2 and consider directly the case when the observation operator is nonlinear? That would avoid confusion for readers.

  Section 2.2 has been updated. I now use $\mathbf{Y}^f = (\boldsymbol{y}_1^f, ..., \boldsymbol{y}_{N_e}^f) \in \mathbb{R}^{N_y \times N_e}$ for the ensemble projected in the observation space with $\boldsymbol{y}_i^f = \mathcal{H}(\boldsymbol{x}_1^f)$, $i = 1, ..., N_e$. The formulas are now given directly using $\mathbf{Y}^f$, however for clarity I still use the linearised observation operator $\mathbf{H}$ for the formula of the Kalman filter update:

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \tag{6}$$

  Here, $\mathbf{H}$ is the linearised observation operator at the forecast mean. However, in practice $\mathbf{H}$ does not need do be computed as it always acts as an operator to project the ensemble members in the observation space. Defining the

forecast ensemble projected in the observation space by $\boldsymbol{y}_i^f = \mathcal{H}(\boldsymbol{x}_1^f)$, $i = 1, ..., N_e$ with $\bar{\boldsymbol{y}}^f$ the ensemble mean, we make the linear approximation

$$\mathbf{Y}^f = \mathbf{H}\mathbf{X}^f \tag{7}$$

with $\mathbf{X}^f = (\boldsymbol{x}_1^f, ..., \boldsymbol{x}_{N_e}^f) \in \mathbb{R}^{N_x \times N_e}$ the forecast ensemble matrix and $\mathbf{Y}^f = (\boldsymbol{y}_1^f, ..., \boldsymbol{y}_{N_e}^f) \in \mathbb{R}^{N_y \times N_e}$ its equivalent in the observation space.

About the experiment:

- p. 8, l. 1-2: About the roughness signal br ,could you detail, in annex for example, how do you simulate the roughness (which equation?) because I do not know this approach.

  I have added a reference to Fournier et al. (1982) where the original algorithm can be found and added the following information in the main text:
  This is a classical algorithm for artificial landscape generation. In 1D, the algorithm recursively subdivide a segment and a random value drawn from a normal distribution $\mathcal{N}(0, \sigma^2)$ is added to the elevation of the midpoint. The standard deviation σ is decreased by a factor $2h$ between two recursions. Here we have used 12 recursions using an initial standard deviation $\sigma = 500$ m and a roughness $h = 0.7$.

- p. 9, l. 21-22: you mention the term "variogram" which is not well known for readers. Could you provide more details how do you generate the ensemble of initial bedrock topographies, in annex for example?

- p. 9, l. 30-31: same comment as previous.

  For the two points above, I now introduce the definition of a variogram and give the formulas for the variograms used in this study:
  Following previous studies (Gudmundsson and Raymond, 2008; Pralong and Gudmundsson, 2011; Bonan et al., 2014; Brinkerhoff et al., 2016), we assume that the initial distributions for b and C are Gaussian with a given mean and a prescribed covariance model. Furthermore we assume no cross-correlation between the initial b, C and z_s and we draw the initial ensembles independently. For b and C, the initial samples are drawn using the R package gstat (Pebesma and Wesseling, 1998). As classical in geostatistics, the covariance model is prescribed using a variogram $\gamma(d)$ that is half the variance of the difference between field values as a function of their separation d. It is usually defined by two parameters, the sill s that defines the semi-variance at large distances and the range r_a which, for asymptotic functions, is defined as the distance where the $\gamma(r_a) = 0.95s$. The package gstat allows directly to draw simulations, i.e. random realisations of the field, from the prescribed spatial moments (Pebesma and Wesseling, 1998).

About the discussion:

- p. 12, l. 25-29: you seem to oppose ensemble and variational methods, but more and more, the tendency is to develop hybrid methods as detailed in Bannister (2017) and Vetra-Carvalho et al. (2018). The main tendency is to use variational approaches in which the adjoint is replaced by ensembles making those adjoint- free approaches. Could you modify your paragraph to reflect this tendency?

  I have removed this part from the conclusion, but added a sentence in the introduction:
  Ensemble DA methods, based on the ensemble Kalman filter (EnKF), have been successful in solving DA problems with large and non-linear geophysical models. Comparative discussions of the performances and advantages of variational and ensemble DA methods can be found in, e.g. Kalnay et al. (2007), Bannister (2017) and Carrassi et al. (2018). As they aim at solving similar problems, a recent tendency is to combine both methods to benefit from their respective advantages.

- p. 13, l. 4-13: There is a now long range of DA literature on how to estimate model bias. One good reference is the following:

  Dee, D. P. Bias and data assimilation, Q. J. Roy. Meteor. Soc., 131, 3323–3343, doi: 10.1256/qj.05.137, 2005.

  Could you reflect on that possibility in your discussion?

– p. 13, l. 14-20: Same comment as before on estimating observation error covariances matrices. A good review paper:

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M. and Zhen, Y.: Joint Estimation of Model and Observation Error Covariance Matri- ces in Data Assimilation: a Review. Mon. Weather Rev., submitted, available at: https://arxiv.org/abs/1807.11221v2, 2018.

Most approaches are based on Desroziers diagnostics, see:

Desroziers, G., Berre, L., Chapnik, B. and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. Roy. Meteor. Soc., 131, 3385–3396, doi: 10.1256/qj.05.108, 2005.

For the two points above, I have slightly reformulated the discussion and added the references :
In a review paper Tandeo et al. (2018) illustrate the impacts of badly calibrated observation and model error covariance matrices in a sequential DA framework and discuss available methods and challenges for their joint estimation. For the question of the impact of systematic errors, i.e. bias, either in the model and in the observations, and their correction by augmenting the system state in variational and ensemble DA, interested readers are referred to Dee (2005).

MINOR COMMENTS AND TYPOS

– p. 1, l. 5: “starting FROM this initial state . . .”

Done

– p. 3, l. 15: “the Kalman filter analysis is REWRITTEN and . . .”

Done

– p. 3, l. 16: The references you mention are all about deterministic versions of the EnKF (Pham et al., 1998, SEIK filter; Bishop et al., 2001, ETKF filter: Nerger et al., 2012, ESTKF filter). But not every EnKF has a deterministic analysis, the stochastic EnKF has also been an important part of EnKF history. Could you add the following references to make your point broader?

Burgers, G., van Leeuwen, P. J., and Evensen, G. Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, doi: 10.1175/1520- 0493(1998)126<1719:ASITEK>2.0.CO;2, 1998.

Houtekamer, P. L., and Mitchell, H. L. Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., 126, 796-811, doi: 10.1175/1520- 0493(1998)126<0796%3ADAUAEK>2.0.CO%3B2, 1998.

References added. I do not introduce in the paper this distinction between stochastic and deterministic filters, as I think it would add more confusion for the reader. As suggested in the sentence just after, interested readers may refer to the review paper by Vetra-Carvalho et al. (2018).

– p. 3, l. 23: “However the many applications in meteorology and oceanography show . . .” While EnKFs have been primarily developed for those two applications, it has been successfully used in a wide range of applications, from hydrology, to crop modelling and oil extraction. Could you rephrase the sentence to show the broad range of applications for EnKF including some that may be closer to glaciology? Maybe add other references too?

this has been changed to:
However, the many applications in geoscience with large and non-linear models have shown that the method remains robust in general and EnKFs are used in several operational centres with atmosphere, ocean and hydrology models (e.g. Sakov et al., 2012; Houtekamer et al., 2009; Hendricks Franssen et al., 2011). While firstly developed for numerical weather and ocean prediction where the forecasts are very sensitive to the model initial state, the method is also widely used, e.g. in hydrology, for join state and parameters estimations (Sun et al., 2014).

– p. 6, Eq. (13): Could you define $\bar{X}^f$?

Done

– p. 9, l. 13: “the transient ASSIMILATION ON model projections”
Done

Assimilation of surface observations in a transient marine ice sheet model using an ensemble Kalman filter

Fabien Gillet-Chaulet[1]

[1]Univ. Grenoble Alpes, CNRS, IRD, IGE, F-38000 Grenoble, France

Correspondence: Gillet-Chaulet Fabien (fabien.gillet-Chaulet@univ-grenoble-alpes.fr)

Abstract.

Marine based sectors of the Antarctic Ice Sheet are increasingly contributing to sea level rise. The basal conditions exert an important control on the ice dynamics and can be propitious to instabilities in the grounding line position. Because the force balance is non-inertial, most ice flow models are now equipped with time-independent inverse methods to constrain the basal conditions from observed surface velocities. However, transient simulations starting ~~form~~ from this initial state usually suffer from inconsistencies and are not able to reproduce observed trends. Here, using a synthetic flow line experiment, we assess the performance of an ensemble Kalman filter for the assimilation of transient observations of surface elevation and velocities in a marine ice sheet model. The model solves the shallow shelf equation for the force balance and the continuity equation for ice thickness evolution. The position of the grounding line is determined by the floatation criterion. The filter analysis estimates both the state of the model, represented by the surface elevation, and the basal conditions, with the simultaneous inversion of the basal friction and topography. The idealized experiment reproduces a marine ice sheet that is in the early stage of an unstable retreat. Using observation frequencies and uncertainties consistent with current observing systems, we find that the filter allows to accurately recover both the basal friction and topography after few assimilation cycles with relatively small ensemble sizes. In addition it is found that assimilating the surface observations has a positive impact to constrain the evolution of the grounding line during the assimilation window. Using the initialised state to perform century-scale forecast simulations, we show that grounding line retreat rates are in agreement with the reference, however remaining uncertainties in the basal conditions may lead to significant delays in the initiation of the unstable retreat. These results are encouraging for the application to real glacial systems.

1 Introduction

Despite recent significant improvements in ice-sheet models, the projected magnitude and rate of the Antarctic and Greenland ice sheets contribution to 21st century sea-level rise (SLR) remains poorly constrained (Church et al., 2013). Improving our

ability to model the century-scale magnitude and rates of mass loss from marine ice sheets remains a key scientific objective (Scambos et al., 2017).

Improving SLR estimates requires, amongst others, to correctly model the dynamics of the grounding line (GL), i.e. the ~~line~~ location where the ice detaches from its underlying bed and goes afloat on the ocean (Durand and Pattyn, 2015). In the GL vicinity, the stress regime changes from a regime dominated by vertical shearing in the grounded part to a buoyancy driven flow dominated by longitudinal stretching and lateral shearing (Pattyn et al., 2006; Schoof, 2007). Because this transition occurs on horizontal dimensions that are smaller than the typical grid size of large scale ice sheet models, many studies have focussed on the ability of the numerical model to properly simulate grounding line migration using synthetic experiments ~~(e.g. Vieli and Payne, 2005; Durand et al., 2009; Gladstone et al., 2012; Seroussi et al., 2014)~~(e.g., Vieli and Payne, 2005; Durand et al., 2 Two Marine Ice-Sheet Model Intercomparison Projects (MISMIP) have allowed to identify the minimum requirements to properly resolve GL motion: *(i)* inclusion of membrane stresses and *(ii)* a sufficiently small grid size or a subgrid interpolation of the GL (Pattyn et al., 2012, 2013). These results suggest that, in realistic applications, the numerical error could be reduced below the errors associated with uncertainties in the model initial state, in the model parameters and in the forcings from the atmosphere and ocean.

For obvious reasons of inaccessibility, the basal conditions (topography and friction) are an important source of uncertainties. Because of the intrinsic instability of marine ice sheets resting over a seaward up-sloping bed, the resolution of the bed topography in the coastal regions can significantly affect short term ice-sheet forecasts (Durand et al., 2011; Enderlin et al., 2013). Analytical developments have shown that the flux at the grounding line depends on the friction law and its coefficients (Schoof, 2007; Tsai et al., 2015). The sensitivity of model projections to the basal friction has been confirmed by several numerical studies both on synthetic and real applications (Joughin et al., 2010; Ritz et al., 2015; Brondex et al., 2017, 2018). In particular, ~~Brondex et al. (2017)~~ Brondex et al. (2017) have shown that, for unbuttressed ice sheets, spatially varying friction coefficients can lead to stable GL positions also in up-sloping bed regions.

Uncertainties in the model state and parameters can be reduced by data assimilation (DA). The objective of formal DA methods is to update the model using observations in a framework consistent with the model, the data and their associated uncertainties (Bannister, 2017). Most ice flow models are now equipped with variational methods to constrain the basal conditions from surface observations (e.g. MacAyeal, 1993; Vieli and Payne, 2003; Larour et al., 2012; Gillet-Chaulet et al., 2012). However most ~~of the~~ studies perform "snapshot" calibrations, where the inversion is performed at a unique initial time step. The state of the model produced from this calibration is therefore sensitive to inconsistencies between the different datasets. The resulting transient artefacts are usually dissipated during a relaxation period where the model drift from the observations.

Because historic remote sensing data collections are spatially incomplete as well as temporarily sparse, most distributed maps are mosaicked, stacked or averaged to maximize the spatial coverage at the expense of the temporal information (Mouginot et al., 2012). However, in the last few years, the development of spaceborne ice-sheet observations has entered a new era with the launch of new satellite missions, considerably increasing the spatial and temporal resolution of surface observations. Because they require linearized versions of the forecast model and of the observation operator, extending the existing variational methods implies important numerical developments (*e.g.* Goldberg et al., 2016; Larour et al., 2016; Hascoët and Morlighem,

2018). In ~~Goldberg and Heimbach (2013)~~Goldberg and Heimbach (2013), a time-dependent adjoint ice flow model is derived using a source-to-source algorithmic differentiation software combined with analytical methods. The DA capabilities are illustrated with a suite of synthetic experiments, including the simultaneous inversion of the basal topography and friction from surface observations and the assimilation of transient surface elevations to retrieve initial ice thicknesses. In a real-world application to a region of West Antarctica, they show that assimilating annually resolved observations of surface height and velocities between 2002 and 2011 allows to improve the model initial state, giving better confidences in projected committed mass losses (Goldberg et al., 2015). Because of the complexity of the code, ~~Larour et al. (2014)~~ Larour et al. (2014) use an operator-overloading approach to generate the adjoint and assimilate surface altimetry observations from 2003 to 2009 to constrain the temporal evolution of the basal friction and surface mass balance of the Northeast Greenland Ice Stream.

Ensemble DA methods, based on the ensemble Kalman filter (EnKF), have been ~~successfull~~ successful in solving DA problems with large and non-linear geophysical models. Comparative discussions of the performances and advantages of ~~variationnal~~ variational and ensemble DA methods can be found in, *e.g.* ~~Kalnay et al. (2007) and Bannister (2017).~~ Kalnay et al. (2007), Bannister (2017) and Carrassi et al. (2018). As they aim at solving similar problems, a recent tendency is to combine both methods to benefit from their respective advantages.

EnKF approximates the state and the error covariance matrix of a system using an ensemble that is propagated forward in time with the model, avoiding the computation of the covariance matrices and the use of linearised or adjoint models. ~~The observations are then assimilated~~ Contrary to time-dependent variational methods where the objective is to find the model trajectory that minimizes the difference with all the observations within an assimilation window, EnKF assimilates the observations sequentially in time ~~to update the ensemble,~~ as they become available using the analysis step of the Kalman Filter~~.~~, as illustrated in Fig. 1. The model trajectory is then discontinuous and, at a given analysis, the model is only informed by past and present observations. For the retrospective analysis of a time period in the past, *i.e.* a reanalysis, ensemble filters can easily be extended to smoothers to provide analyses that are informed by all past, present and future observations (Evensen and van Leeuwen , 2000; Li and Navon, 2001; Cosme et al., 2012; Nerger et al., 2014). Since the first version introduced by ~~Evensen (1994)~~ Evensen (1994) many variants have been ~~developped~~ developed mainly differing in the way the Kalman Filter analysis is ~~rewrited~~ rewritten and the analysed error covariance matrix resampled ~~(e.g. Pham et al., 1998; Bishop et al., 2001;~~ A review of the most popular EnKFs using common notations can be found in ~~Vetra-Carvalho et al. (2018)~~Vetra-Carvalho et al. (2018). Efficient and parallel algorithms have been developed, and because they are independent of the forward model, several open-source toolboxes that implements various EnKFs are now available, e.g. OpenDA (https://www.openda.org), PDAF (http://pdaf.awi.de).

As Monte-Carlo methods, EnKFs suffer from under sampling issues as often the size of the ensemble is much smaller than the size of the system to estimate. Localisation and inflation are popular methods to counteract these issues and to increase the stability of the filtering. Because they are based on the original Kalman Filter equations, EnKFs are optimal only for Gaussian distributions and linear models. However, the many applications in ~~meteorology and oceanography show that these methods are efficient in practice.~~ geoscience with large and non-linear models have shown that the method remains robust in general and EnKFs are used in several operational centres with atmosphere, ocean and hydrology models (e.g. Sakov et al., 2012; Houtekamer et al., 20

While firstly developed for numerical weather and ocean prediction where the forecasts are very sensitive to the model initial state, the method is also widely used, *e.g.* in hydrology, for join state and parameters estimations (Sun et al., 2014).

In the context of ice-sheet modelling, encouraging results have been obtained by ~~Bonan et al. (2014)~~ Bonan et al. (2014) for the estimation of the state and basal conditions of an ice-sheet model using the Ensemble Transform Kalman Filter (ETKF, Bishop et al., 2001; Hunt et al., 2007). ~~Using a~~ They study the performance of the method using idealised twin experiments where perturbed observations generated from a model run are used in the DA framework to retrieve the true model states and parameters. Using a flowline shallow ice model~~and synthetic 1D twin experiments~~, they show that both the basal topography and basal friction can be retrieved with good accuracy from surface observations with realistic noise levels, even for relatively small ensembles. The method has been further developed to assimilate the margin position in a shallow ice model that explicitly tracks the boundaries with a moving mesh method (Bonan et al., 2017).

The purpose of this paper is to explore the performance of ensemble Kalman filtering for the initialisation of a marine ice sheet model that includes GL migration. In particular, we want to address *(i)* the quality of the analysis for the simultaneous estimation of the basal topography and friction in the context of a marine ice sheet that is undergoing an unstable GL retreat, and *(ii)* the effects of the remaining uncertainties for the predictability of GL retreat. The ice flow model and the EnKF used in this study are described in Section 2. To test the DA framework, we define a twin experiment in Section 3. Section 4 presents the results for both the transient assimilation and the forecasts. Finally, perspectives and challenges for real applications are discussed in Section 5, before concluding remarks.

2 Methods

2.1 Ice flow model

The gravity-driven free surface flow of ice is solved using the finite-element ice flow model Elmer/Ice (Gagliardini et al., 2013).

For the force balance, we solve the shelfy-stream approximation (SSA) equation (MacAyeal, 1989) in one horizontal dimension. This is a vertically integrated model that derives from the Stokes equations for small aspect ratio and basal friction. In 1D, this leads to the following non-linear partial differential equation for the horizontal velocity field u

$$\frac{\partial}{\partial x}\left(4\bar{\eta}H\frac{\partial u}{\partial x}\right) - \tau_b = \rho_i g H \frac{\partial z_s}{\partial x} \tag{1}$$

whith ρ_i the ice density, g the gravity norm, $H = z_s - z_b$ the ice thickness with z_s and z_b the top and bottom surface elevations, respectively. Using the Glen's constitutive flow law, the vertically averaged effective viscosity $\bar{\eta}$ is given by

$$\bar{\eta} = \frac{1}{H}\int_{z_b}^{z_s} \frac{1}{2} A^{-1/n} D_e^{(1-n)/n} \mathrm{d}z \tag{2}$$

where D_e is the second invariant of the strain-rate tensor, equal here to $D_e^2 = (\partial u/\partial x)^2$, A is the rate factor and n is the creep exponent, taken equal to the usual value $n = 3$ in the following. The basal friction τ_b is null under floating ice and is represented

with the non-linear Weertman friction law for grounded ice

$$\tau_b = Cu^m \tag{3}$$

with C and m the friction coefficient and exponent, respectively. In the following, we use the classical power law with $m = 1/n = 1/3$. When in contact with the ocean, the ice is assumed to be in hydrostatic equilibrium. The floating condition is evaluated directly at the integration points and τ_b in Eq. (1) is set to 0 wherever ice is floating (Seroussi et al., 2014).

The time dependency is introduced by the evolution of the top and bottom free surfaces. Because of the hydrostatic equilibrium, the ice sheet topography is fully defined by the bed elevation b and only one prognostic variable. Equation (1) is then coupled with the vertically integrated mass conservation equation for the evolution of the ice thickness H

$$\frac{\partial H}{\partial t} + \frac{\partial (uH)}{\partial x} = a_s - a_b \tag{4}$$

with a_s the surface accumulation rate and a_b the basal melt rate. The free surfaces z_s and z_b are obtained from the floating condition which, for z_s, using a constant sea level $z_{sl} = 0$, gives

$$\begin{cases} z_s = b + H & \text{for } H \geq -b\dfrac{\rho_w}{\rho_i} \\ z_s = H\left(1 - \dfrac{\rho_i}{\rho_w}\right) & \text{otherwise} \end{cases} \tag{5}$$

with ρ_w the sea water density.

2.2 Data Assimilation

2.2.1 Filter Algorithm

For the assimilation, we use the Error Subspace Ensemble Transform Kalman Filter (ESTKF, Nerger et al., 2012). Originally derived from the singular evolutive interpolated Kalman filter (SEIK, Pham et al., 1998), ESTKF leads to the same ensemble transformations as the ETKF but at a slightly lower computational cost. In practice we use the local version of the filter implemented in PDAF (http://pdaf.awi.de Nerger et al., 2005b) and coupled to Elmer/Ice in an offline mode. This section outlines the ESTKF algorithm.

As an EnKF, ESTKF approximates the state $\boldsymbol{x}^k$ and the error covariance matrix $\mathbf{P}_k$ of a system at time t_k using an ensemble of N_e realisations $\boldsymbol{x}_i^k$, $i = 1, ..., N_e$. The state vector, of size N_x, contains the prognostic variables and model parameters to be estimated and is approximated by the ensemble mean

$$\bar{\boldsymbol{x}}^k = \frac{1}{N_e} \sum_{i=1}^{N_e} \boldsymbol{x}_i^k \tag{6}$$

while the error covariance matrix is ~~given by~~ approximated by

$$\mathbf{P}_k = \frac{1}{N_e - 1} \mathbf{X'}_k \mathbf{X'}_k^T \tag{7}$$

where $\mathbf{X'}_k = (\boldsymbol{x}_1^k - \bar{\boldsymbol{x}}^k, ..., \boldsymbol{x}_{N_e}^k - \bar{\boldsymbol{x}}^k) \in \mathbb{R}^{N_x \times N_e}$ is the ensemble perturbation matrix.

The algorithm can be decomposed in two steps, the *forecast* and the *analysis*. Superscripts ^{f} (resp. ^{a}) denote quantities related to each step respectively. ~~During the forecast, we use the numerical model M_k~~The *forecast* propagates the state and the error covariance matrix of the system forward in time, from a previous analysis at $t = t_{k-1}$ to the next observation time $t = t_k$. For this, the numerical model $\mathcal{M}_k$, assumed perfect in the sequel, ~~to propagate the ensemble from a previous analysis at $t = t_{k-1}$ to~~ is used to propagate each ensemble member individually during n_{dt} model time steps

$$\boldsymbol{x}_i^{f,k} = \mathcal{M}_k(\boldsymbol{x}_i^{a,k-1}) \tag{8}$$

At $t = t_k$ ~~where new observations are available~~

$$\boldsymbol{x}_i^{f,k} = M_k(\boldsymbol{x}_i^{a,k-1})$$

~~This step allows to propagate the state and the error covariance matrix of the system forward in time. During the analysis the ensemble is updated using the available observations and in the~~ , a vector of observations $\boldsymbol{y}_o^k$ of size N_y (with usually $N_y << N_x$) is available. $\boldsymbol{y}_o^k$ is related to the true system state $\boldsymbol{x}^f$ by $\boldsymbol{y}_o^k = \mathcal{H}(\boldsymbol{x}^f) + \boldsymbol{\epsilon}^k$ where the observation error $\boldsymbol{\epsilon}^k$ is assumed to be a white Gaussian distributed process with known covariance matrix $\mathbf{R}^k$, and $\mathcal{H}$ is the observation operator that relates the state variables to the observations. When $\boldsymbol{y}_o^k$ is the observed surface velocities, the relation between the observations and the system state, *i.e.*, the ice-sheet geometry, and ,parameters, *i.e.* the boundary conditions, is given by the force balance equation (1), thus $\mathcal{H}$ is a non-linear elliptic partial differential equation.

The *analysis* provides a new estimation of the system state by combining the informations from the forecast and the observations. In the following we will omit the time index ^{k} in the notations as all the analysis is performed at $t = t_k$.

~~The analysis step is based on the KF equations where the state vector and the covariance matrix are updated using the following equations~~ As others EnKFs, ESTKF uses the Kalman Filter update equations to compute the analysed system state $\bar{\boldsymbol{x}}^a$ and covariance matrix $\mathbf{P}^a$ from the forecast, the observations and their uncertainties:

$$\begin{cases} \bar{\boldsymbol{x}}^a = \bar{\boldsymbol{x}}^f + \mathbf{K}\boldsymbol{d} \\ \mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f \end{cases} \tag{9}$$

where ~~$\boldsymbol{d} = \boldsymbol{y} - H(\bar{\boldsymbol{x}}^f)$~~ $\boldsymbol{d} = \boldsymbol{y}_o - \mathcal{H}(\bar{\boldsymbol{x}}^f)$ is the *innovation* ~~, $\boldsymbol{y}$ is the vector of observations of size N_y, H is the observation operator, with the matrix $\mathbf{H}$ the linearised observation operator $\mathbf{K}$ is the~~ and $\mathbf{K}$ is the *Kalman gain* given by

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{10}$$

Here, $\mathbf{H}$ is the linearised observation operator at the forecast mean. However, in practice $\mathbf{H}$ does not need do be computed as it always acts as an operator to project the ensemble members in the observation space. Defining the forecast ensemble projected in the observation space by $\boldsymbol{y}_i^f = \mathcal{H}(\boldsymbol{x}_1^f)$, $i = 1, ..., N_e$ with $\bar{\boldsymbol{y}}^f$ the ensemble mean, we make the linear approximation

$$\mathbf{K} = \mathbf{P}\mathbf{Y}^f = \mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{X}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{11}$$

and $\mathbf{R}$ is the observation error covariance matrix. With with $\mathbf{X}^f = (\boldsymbol{x}^f_1, ..., \boldsymbol{x}^f_{N_e}) \in \mathbb{R}^{N_x \times N_e}$ the forecast ensemble matrix and $\mathbf{Y}^f = (\boldsymbol{y}^f_1, ..., \boldsymbol{y}^f_{N_e}) \in \mathbb{R}^{N_y \times N_e}$ its equivalent in the observation space.

In practice, with large models ($N_x >> 1$), it is not possible to form the matrix the covariance matrices $\mathbf{P}^f$. ESTKF as other EnKFs, performs the analysis in a low-dimensional subspace, denoted the error subspace, which approximates the full error space. This makes the implementation efficient and tractable for large models. The use of an error subspace is possible because and $\mathbf{P}^a$ of size $N_x \times N_x$ can not be formed, so that, to be implemented, the analysis (Eq. 9) needs to be reformulated. Moreover, the sample covariance matrix $\mathbf{P}_k \in \mathbb{R}^{N_x \times N_x}$ in approximated with an ensemble of size N_e (Eq. (7) has a rank 7) is only a low-rank approximation of the true covariance matrix and its rank is at most $N_e - 1$. However, the drawback of approximating the covariance matrix by a low-rank matrix is that the analysis will adjust the model state only in this subpace, ignoring errors directions not accounted for ESTKF uses this property to write the analysis in a $(Ne-1)$-dimensional subspace spanned by the ensemble (Hunt et al., 2007).

To perform the projection and referred to as the error subspace (Nerger et al., 2005a). The forecast covariance matrix $\mathbf{P}^f$ is then rewritten as

$$\mathbf{P}^f = \frac{1}{N_e - 1}\mathbf{L}\mathbf{L}^T \tag{12}$$

where $\mathbf{L} \in \mathbb{R}^{N_x \times N_e - 1}$ is given by

$$\mathbf{L} = \mathbf{X}^f \mathbf{\Omega} \tag{13}$$

The matrix $\mathbf{\Omega} \in \mathbb{R}^{N_e \times N_e - 1}$ defined as

$$\Omega_{ij} = \begin{cases} 1 - \frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i = j,\ i < N_e \\ -\frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i \neq j,\ i < N_e \\ -\frac{1}{\sqrt{N_e}} & \text{for } i = N_e \end{cases} \tag{14}$$

projects the ensemble matrix $\mathbf{X}^f$ onto the error subspace, ESTKF use the deterministic matrix $\mathbf{\Omega} \in \mathbb{R}^{N_e \times N_e - 1}$

$$\Omega_{ij} = \begin{cases} 1 - \frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i = j,\ i < N_e \\ -\frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & \text{for } i \neq j,\ i < N_e \\ -\frac{1}{\sqrt{N_e}} & \text{for } i = N_e \end{cases}$$

The basis vectors of the error subspace are then given by . The multiplication with $\mathbf{X}^f$ subtracts the ensemble mean and a fraction of the last column of the ensemble perturbation matrix $\mathbf{X}'^f$ from all other columns.

After some algebra using Eq. (12) and Eq. (9), $\mathbf{P}^a$ can be written as a transformation of $\mathbf{L}$

$$\mathbf{P}^a = \mathbf{L}\mathbf{A}\mathbf{L}^T \tag{15}$$

with the transform matrix $\mathbf{A} \in \mathbb{R}^{N_e-1 \times N_e-1}$ given by

$$\mathbf{L}\mathbf{A}^{-1} = \mathbf{X}\rho(N_e-1)\mathbf{I} + (\mathbf{Y}^f\mathbf{\Omega})^T\mathbf{R}^{-1}\mathbf{Y}\mathbf{\Omega} \tag{16}$$

where ~~$\mathbf{X}^f = (\boldsymbol{x}_1, ..., \boldsymbol{x}_{N_e}) \in \mathbb{R}^{N_x \times N_e}$~~ $\rho \in [0,1]$ is the *forgetting factor* discussed in section 2.2.2.

Finally, the update step ~~can be written~~ is obtained as a single equation for the transformation of ~~$\mathbf{L}$~~ the forecast ensemble $\mathbf{X}^f$ to the analysed ensemble $\mathbf{X}^a$ as

$$\mathbf{X}^a = \bar{\mathbf{X}}^f + \mathbf{L}\mathbf{X}^f\mathbf{\Omega}(\bar{\mathbf{W}} + \mathbf{W}) \tag{17}$$

where $\bar{\mathbf{X}}^f$ is the matrix where the columns are given by the forecast ensemble mean, $\bar{\mathbf{W}}$ is a matrix where the columns are given by the vector

$$\bar{\boldsymbol{w}} = \mathbf{A}(\mathbf{H}\mathbf{L}\mathbf{Y}\mathbf{\Omega})^T\mathbf{R}^{-1}(\boldsymbol{y}_o - \bar{\boldsymbol{y}}^f) \tag{18}$$

and

$$\mathbf{W} = \sqrt{N_e-1}\mathbf{C}\mathbf{\Omega}^T$$

~~The transform matrix $\mathbf{A} \in \mathbb{R}^{N_e-1 \times N_e-1}$ is computed as~~ $\mathbf{W}$ is given by

$$\mathbf{A}^{-1}\mathbf{W} = \rho(N_e-1)\mathbf{I} + (\mathbf{H}\mathbf{L})\sqrt{N_e-1}\mathbf{C}\mathbf{\Omega}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L} \tag{19}$$

~~and~~ where $\mathbf{C}$ ~~in Eq. 19 is taken as its~~ is the symmetric square root ~~. Multiplicative covariance inflation is tuned using the *forgetting factor* ρ, with $0 < \rho \leq 1$. By increasing~~ of $\mathbf{A}$ obtained by singular value decomposition.

Finally, the analysed ensemble ~~spread, inflation is a popular method to counteract the effects of undersampling and enhance the filter robustness~~ $\mathbf{X}^a$ is used as the initial ensemble for the next forecast, and so on up to the end of the data assimilation window.

~~The product $\mathbf{H}\mathbf{L}$ in the equations above is computed, using~~ We draw attention to several remarks on the algorithm:

- To compute the innovation $\boldsymbol{d}$, we have made the same linear approximation $\mathcal{H}(\bar{\boldsymbol{x}}^f) = \bar{\boldsymbol{y}}^f$ as Hunt et al. (2007). This choice is consistent with the computation of the covariance matrices $\mathbf{P}^f\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ in Eq. (10) using the linear approximation Eq. (11) (Houtekamer and Mitchell, 2001).

- Several ensembles can have the same mean and covariance matrix, this is why there is several EnKFs that exactly satisfy Eq. (9) but lead to different ensemble transformations and thus different analysed ensembles (Vetra-Carvalho et al., 2018).

With the same arguments several variants of ESTKF can be introduced, *e.g.* by replacing $\mathbf{\Omega}$ in Eq. (~~14) , as $(\mathbf{H}\mathbf{X}^f)\mathbf{\Omega}$. The product $\mathbf{H}\mathbf{X}^f$ project each ensemble member in the observation space andcan then be computed using the non-linear observation operator applied to each member $H(\boldsymbol{x}_i)$. Further, the multiplication with $\mathbf{\Omega}$ implicitly subtracts the observed ensemble mean leading to consistent formulations for the covariance matrices $\mathbf{P}^f\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$~~ 19) by a random matrix with the same properties or using a Cholesky decomposition to compute $\mathbf{C}$.

- As written here, the ESTKF leads to the same ensemble transformation as the ETKF. However, as the computations are not performed in the same sub-space tiny differences due to the finite precision of the computations may grow leading to slight differences at the end of the assimilation window (Nerger et al., 2012).

- The leading computational cost of the ensemble transformation in ESTKF is $\mathcal{O}\left(N_y(N_e-1)^2+N_e(N_e-1)^2+N_xN_e(N_e-1)\right)$, so it scales linearly with N_x and N_y (Nerger et al., 2012). Naturally, increasing N_e also requires to increase the number of model runs and, in general, the objective is to get the ensemble size as small as possible. The performance of the algorithm also depends on the evaluation of the product of $\mathbf{R}^{-1}$ with some vectors, which can become more expensive when the observation errors are spatially correlated.

2.2.2 Filter stabilisation: inflation and localisation

In practice for large scale problems, EnKFs as Monte-Carlo methods can suffer from under-sampling issues. First, because of the rank deficiency of the covariance matrix $\mathbf{P}^f$, the analysis adjust the model state only in the error subspace, ignoring error directions not accounted for by the ensemble (Hunt et al., 2007). This can result in an analysis that is overconfident and underestimates the true variances. On the long run, the ensemble spread will become too small and the analysis will give to much weight on the forecast finally disregarding the observations and diverging from the true trajectory. A common simple *ad-hoc* remedy is to inflate the forecast covariance matrix with a multiplicative factor (Pham et al., 1998; Anderson and Anderson, 199). Here, inflation has been introduced in Eq. (~~10) (Houtekamer and Mitchell, 2001). For consistency, the vector $H(\bar{\boldsymbol{x}}^f)$ used for the computation of the innovation $\boldsymbol{d}$ in Eq. (18) is then also computed as the observed ensemble mean. The same linear approximation is made by Hunt et al. (2007) for the derivation of the local ETKF.~~

~~Here,~~ 16) using the forgetting factor $\rho \in [0,1]$ with $\rho = 1$ corresponding to no inflation (Pham et al., 1998). It is the inverse of the inflation factor used by Bonan et al. (2014).

Second, the rank deficiency of $\mathbf{P}^f$ leads to the appearance of spurious correlations between parts of the system that are far away. As these correlations are usually small, a common remedy is to damp these correlations with a procedure called localisation. In covariance localisation, localisation is applied by using an ensemble covariance matrix that results from the Schur product of $\mathbf{P}^f$ with an *ad-hoc* correlation matrix that drops long range correlations (Hamill et al., 2001; Houtekamer and Mitchell, 20 However, this localisation technique is not practical for square-root filters where $\mathbf{P}^f$ is never explicitly computed. Here, as in Bonan et al. (2014), we use a ~~local formulation of the filter. The localisation algorithm is~~ localisation algorithm based on domain localisation and observation localisation ~~(Hunt et al., 2007). It~~ (Ott et al., 2004; Hunt et al., 2007). Both methods are illustrated in Sakov and Bertino (2011) who conclude that they should yield to similar results. Domain localisation assumes

that observations far from a given location have negligible influence. ~~This effectively drops long-range correlations that may truly exist but, on the other hand, that are unlikely properly estimated with a small ensemble size.~~ In practice, the state vector in each single mesh node is updated independently during a loop through the nodes that can easily be parallelized for numerical efficiency. For each local analysis, only the observations within a given radius r from the current node are used. In addition to avoid an abrupt cut-off, the observation error covariance matrix $\mathbf{R}$ is modified so that the inverse observation variance decreases to zero with the distance from the node using a fifth-order polynomial function which mimics a Gaussian function but has compact support (Gaspari and Cohn, 1999). Because it drops spurious long-range correlations and allows the local analyses to choose different linear combinations of the ensemble members in different regions, localisation implicitly increases the rank of the covariance matrix, leading to a larger dimension of the error subspace, implicitly increasing the effective ensemble size and the filter stability (Nerger et al., 2006; Hunt et al., 2007). However, it has been reported that localisation could produce imbalanced solutions (Mitchell et al., 2002). Here, because the force balance are non-inertial and the SSA assumes that the ice-shelves are in hydrostatic equilibrium, this shouldn't be an issue. Another disadvantage is that, when long-range correlations truly exist, the analysis will ignore useful informations that could have been used from distant observations.

Here, the forgetting factor ρ and the localisation radius r will be used as tuning parameters of the filter. Improving the theoretical understanding of these *ad hoc* procedures and developing adaptive scheme is an active research area and interested readers can refer to review articles (e.g. Bannister, 2017; Carrassi et al., 2018; Vetra-Carvalho et al., 2018).

3 Experimental design

To evaluate the performance of the DA framework we perform a twin experiment. In this section we first describe the synthetic reference simulation that will be used to assess the performance of the DA framework. From this reference, we generate a set of synthetic noisy observations that will be used by the assimilation scheme. Finally, we describe the initial ensemble constructed using *a priori* or background informations.

3.1 Reference simulation

We start by building an initial steady marine ice-sheet. The domain extends from $x = 0\,\mathrm{km}$ where we apply a symmetry condition, $u = 0$ in Eq. (1), to $x = 800\,\mathrm{km}$ where we have a fixed calving front. We use 1D linear elements with a uniform mesh resolution of $200\,\mathrm{m}$, leading to 4001 mesh nodes.

Following ~~Durand et al. (2011)~~Durand et al. (2011), we generate a synthetic bed geometry that reproduces a typical large-scale overdeepening with some small scale roughness. The bed $b = b_{trend} + b_r$ is the sum of a general trend b_{trend} defined as

$$b_{trend} = \begin{cases} -1100 + x & \text{for } x \leq 450\,\mathrm{km} \\ -650 - 5(x - 450) & \text{for } x > 450\,\mathrm{km} \end{cases} \tag{20}$$

and a roughness signal b_r that is computed at $200\,\mathrm{m}$ resolution using ~~an iterative random mid-point algorithm with a prescribed fractal dimension of 1.3. The~~ a random midpoint displacement method (Fournier et al., 1982). This is a classical algorithm for artificial landscape generation. In 1D, the algorithm recursively subdivide a segment and a random value drawn from a normal distribution $\mathcal{N}(0, \sigma^2)$ is added to the elevation of the midpoint. The standard deviation σ is decreased by a factor 2^h between two recursions. Here we have used 12 recursions using an initial standard deviation $\sigma = 500\,\mathrm{m}$ and a roughness $h = 0.7$. The resulting bed is shown in Fig. 2.

For the basal friction, we use a synthetic sinusoidal function with two wavelengths for C ($\mathrm{MPa\,m^{-\frac{1}{3}}\,a^{-\frac{1}{3}}}$)

$$C = 0.020 + 0.015\sin(5\frac{2\pi x}{L})\sin(100\frac{2\pi x}{L}) \tag{21}$$

with $L = 800\,\mathrm{km}$ (Fig. 3).

While not tuned to match any specific glacier, this synthetic design compares relatively well to the conditions found in Thwaites Glacier (Antarctica). Thwaites has been the focus of many recent studies as it is undergoing rapid ice loss and, connected to deep marine-based basins, its retreat could trigger a large scale collapse of the West Antarctic Ice Sheet over the next centuries (Scambos et al., 2017). In Fig. 4, C and b are compared with model results from Brondex et al. (2018) along three streamlines. In Brondex et al. (2018) , C has been inferred from the observed surface velocities using a time-independent control inverse method and a SSA model. We can see that our synthetic design is realistic both in terms of amplitude and spatial variations. As the other characteristics (geometry, small flow divergence/convergence) are also similar, the model velocities have the good order of magnitude.

Using an uniform ice rigidity $B = (2A)^{-1/n} = 0.4\,\mathrm{MPa\,a^{\frac{1}{3}}}$, we grow an ice sheet to steady state using a uniform surface accumulation $a_s = 0.5\,\mathrm{m\,a^{-1}}$ and no basal melting $a_b = 0$. The steady state GL is located at $x = 440\,\mathrm{km}$, just downstream of the region of overdeepening (Fig. 2).

~~At $t = 0$,~~ In Jenkins et al. (2018), observed ice-flow accelerations in the Amundsen sea sector have been attributed to the decadal oceanic variability, where warm phases associated with increased basal melt induce a thinning of the ice shelves reducing their buttressing effect initiating short lived periods of unstable retreat of the most vulnerable GLs. In a flow line experiment the ice shelf do not exert any buttressing effect. Using a suite of melting and calving perturbation experiments for Pine Island Glacier, Favier et al. (2014) have shown that, when initiated, the dynamics of the unstable retreat is fairly independent of the type and magnitude of the perturbation. Here, to trigger the initial acceleration, we instantaneously decrease the ice rigidity to $B = 0.3\,\mathrm{MPa\,a^{\frac{1}{3}}}$ at $t = 0$, keeping all the other parameters constant.

This initial perturbation induce an acceleration, a thinning and a retreat of the GL. The model is then run for 200 years with a time step $dt = 5\,10^{-3}\,\mathrm{a^{-1}}$. After a short stabilisation at $x = 437.2\,\mathrm{km}$ between $t = 13\,\mathrm{a}$ and $t = 32\,\mathrm{a}$, the GL retreats at a rate of approximately $1\,\mathrm{km\,a^{-1}}$ during the following 100 years, then the rate decreases as the GL enters an area of down-slopping bed (Fig. 2). The retreat rate shows small variations associated with spatial variations of the topography and basal friction.

3.2 Synthetic Observations

From the reference run, we generate synthetic noisy observations that are typical of the resolution and performance of actual observing systems.

For the bed, we mimic an airborne radar survey conducted perpendicular to the ice flow with an along flow resolution of approximately $15\,\mathrm{km}$. For this, we randomly select 54 locations between $x = 0$ and $x = 800\,\mathrm{km}$, and then linearly interpolate the true bed and add a random uncorrelated Gaussian noise with a standard deviation ~~of $20\,\mathrm{m}$~~ $\sigma_b^{obs} = 20\,\mathrm{m}$ (Fig. 3).

We assume that the surface elevation and velocities are observed at an annual resolution at each mesh node. We then add an uncorrelated Gaussian noise with a standard deviation ~~of $10\,\mathrm{m}$~~ $\sigma_{z_s^{obs}} = 10\,\mathrm{m}$ for the surface elevation and ~~$20\,\mathrm{m\,a^{-1}}$~~ $\sigma_u^{obs} = 20\,\mathrm{m\,a^{-1}}$ for the velocity. The most recent velocity products are now posted with a monthly to annual resolution (Mouginot et al., 2017; Joughin et al., 2018). The reported uncertainty for individual velocity estimates using the 6- and 12-day image pairs from the Sentinel 1A/B satellites is 6.2 and $17.5\,\mathrm{m\,a^{-1}}$ for the two horizontal velocity components in stable conditions; however this could be underestimated in the coastal areas. For the surface elevation, the spatial and temporal resolution as well as the coverage and uncertainty will depend on the sensors. The ArcticDEM (http://arcticdem.org) is a collection of openly available digital surface models derived from satellite imagery and posted at ~~2 m~~ 2-m spatial resolution. After co-registration, a standard deviation ranging from 2 to $4\,\mathrm{m}$ has been reported for the uncertainty of elevation difference between two individual models of static surfaces (Dai and Howat, 2017). Using the same satellites, Greenland digital elevation models are now posted with a 3-month temporal resolution (https://nsidc.org/data/nsidc-0715).

3.3 Assimilation setup

We recall that our aim is to initialise the model using the DA framework to estimate the state together with the basal conditions. As a simplification to realistic experiments, we assume in the following that the ice rheological properties (represented by the Glen flow law and its ~~parameter~~parameters) and the forcing (represented by the surface and basal mass balances in Eq. (4)) are perfectly known. In addition, we assume that the form of the basal friction follows Eq. (3) with $m = 1/3$, so that only the spatially-varying friction coefficient C is uncertain.

In our model, as the force balance equation (1) contains no time derivative, the velocity is a diagnostic variable. Because of the flotation condition, the topography can be represented by only one prognostic variable. The state vector $\boldsymbol{x}$ is then given by the free surface elevation z_s at every mesh node, and we use the floatation Eq. (5) for the mapping between the ice thickness H and z_s. The state vector is augmented by the two parameters to be estimated, the bedrock topography b and the basal friction coefficient C. For the parameters we assume a persistence model, *i.e.* no time evolution, during the forecast step (Eq. 8). Because the velocities are insensitive to the basal conditions where ice is floating, these two parameters are included in the state vector only for the nodes where at least one member is grounded. In addition, to insure that C remains positive, we use the following change of variable for the assimilation $C = \alpha^2$. Although it does not insure uniqueness of the estimation as α and $-\alpha$ would lead to the same C, this change of variable is classical (MacAyeal, 1993) and was chosen as the reference friction

coefficient spans only one order of magnitude. Similar performances where found using the other classical change of variable $C = 10^{\alpha}$ as in Gillet-Chaulet et al. (2012).

~~As the~~ Because both z_s and b are included in the state vector, the analysis does not conserve the ice sheet volume, neither for the ensemble mean and the individual members. However, the estimation of the ice-sheet volume is improved at each analysis as more data are assimilated, and the final volume is the best estimation provided by the filter knowing the model, all the observations during the assimilation window and their uncertainties. As mentioned in the introduction, if the main interest is an analysis of past volume changes, as smoother might be more appropriate and the smoother extension of the ESTKF can be found in Nerger et al. (2014).

Kalman-based filters are based on the hypothesis of the independence between the background, *i.e. the initial ensemble*, and the observations that are used during the assimilation. As the synthetic bed observations ~~are~~ will be used to construct the initial ensemble (cf next section), we assimilate only the surface elevation and velocity observations, every year from $t = 1\,\mathrm{a}$ up to $t = 35\,\mathrm{a}$. The observation operator ~~H~~ $\mathcal{H}$ is a simple mapping for the surface elevation, and is given by the non-linear SSA equation (Eq. 1) for the surface velocities.

Finally, to illustrate the effect of the transient ~~assimilationon~~ assimilation on model projections on time scales relevant for sea level projections, the analysed states at $t = 20\,\mathrm{a}$ and $t = 35\,\mathrm{a}$ are used to run deterministic and ensemble forecasts up to $t = 200\,\mathrm{a}$. The deterministic forecast uses the ensemble mean produced by the analysis while the ensemble forecast propagates the full ensemble.

3.4 Initial ensemble

~~The initial~~ For atmosphere and ocean models, the initial state is usually sampled from a climatology, either observed or from a model run. This method can not be used for the parameters and the initial ensemble must reflect the background and the estimation of its uncertainty, available *a priori* before the assimilation. ~~Kalman-based filters are based on the hypothesis of the independence between the background and the observations that are used during the assimilation~~Following previous studies (Gudmundsson and Raymond, 2008; Pralong and Gudmundsson, 2011; Bonan et al., 2014; Brinkerhoff et al., 2016), we assume that the initial distributions for b and C are Gaussian with a given mean and a prescribed covariance model. Furthermore we assume no cross-correlation between the initial b, C and zs and we draw the initial ensembles independently.

For ~~the bed, we use the synthetic observations to draw conditional Gaussian simulations~~ b and C, the initial samples are drawn using the R package *gstat* (Pebesma and Wesseling, 1998). ~~To model the spatial dependence,~~ As classical in geostatistics, the covariance model is prescribed using a variogram $\gamma(d)$ that is half the variance of the difference between field values as a function of their separation d. It is usually defined by two parameters, the sill s that defines the semi-variance at large distances and the range r_a which, for asymptotic functions, is defined as the distance where the $\gamma(r_a) = 0.95s$. The package *gstat* allows directly to draw simulations, i.e. random realisations of the field, from the prescribed spatial moments (Pebesma and Wesseling, 1998).

For the bed we use an exponential function ~~for the variogram with a range of $50\,\mathrm{km}$, a sill of $4000\,\mathrm{m}^2$ and a nugget of $200\,\mathrm{m}^2$. For a large~~

$$\gamma(d) = s(1 - e^{-\frac{3d}{r}}) \tag{22}$$

with $r_a = 50\,\mathrm{km}$ and $s = 4000\,\mathrm{m}^2$. We also add a nugget model defined by

$$\gamma(d) = \begin{cases} 0 & d = 0 \\ nug & d > 0 \end{cases} \tag{23}$$

with $nug = 200\,\mathrm{m}^2$. This model is meant to represent the bed measurement error. To draw the initial ensemble, the ~~ensemble mean should converge to the prediction obtained by kriging~~simulations are conditioned with the bed observations.This procedure gives an initial ensemble that is drawn from the posterior probability distribution that would be obtained using ordinary Kriging with the same observations and variograms. The ensemble mean and spread for a 50-members ensemble are shown in Fig. 3 and the first three members are shown in Fig. 5. As expected, the ensemble spread increases with the distance from the observations. At the observation locations, the spread is controlled by the nugget~~which is meant to represent the bed measurement error. However, for~~ . For the individual members, the nugget controls the small scale variability, resulting in a roughness larger than the reference. When averaged this roughness disappears, and the ensemble mean has a much smoother topography.

For the friction coefficient, we assume that we know the mean value ~~$C_{mean} = 0.020\,\mathrm{MPa\,m^{-\frac{1}{3}}\,a^{-\frac{1}{3}}}$~~ $C_{mean} = 0.020\,\mathrm{MPa\,m^{-\frac{1}{3}}\,a^{-\frac{1}{3}}}$ and draw unconditional simulations. For the spatial dependence, we use a Gaussian function $\gamma(d) = s(1 - e^{-3(\frac{d}{r})^2})$ for the variogram using a range ~~of $2.5\,\mathrm{km}$,~~ $r_a = 2.5\,\mathrm{km}$ and a sill ~~of $8.10^{-5}\,\mathrm{MPa^2\,m^{-\frac{2}{3}}\,a^{-\frac{2}{3}}}$~~ $s = 8.10^{-5}\,\mathrm{MPa^2\,m^{-\frac{2}{3}}\,a^{-\frac{2}{3}}}$. This results in initial ensemble members that have approximately the same maximal amplitude as the reference, as shown in Fig. 5.

For the free surface, we initialise all the members using the observed (noisy) free surface at $t = 0$. Doing so, we implicitly assume that the spread of the ensemble induced by the uncertain initial conditions at the first analysis is small compared to the spread induced by the uncertain parameters. This is motivated by the fact that divergence anomalies induced by uncertainties in model parameters can typically reach tens to hundreds of meters per years in fast flowing areas (Seroussi et al., 2011).

4 Results

4.1 Assimilation

To assess the performance of the DA in retrieving the basal conditions we compute the root-mean-square error (RMSE) between the analysed ensemble mean and the reference for both the bed and the friction coefficient, RMSE_b and RMSE_C respectively. After each analysis, the RMSE is computed using all the nodes where the basal conditions have been updated by the assimilation, i.e. at least one member is grounded, and where $x \geq 300\,\mathrm{km}$. The later value is close to the position reached by the grounding line after 200 years in the reference simulation, moreover, during the assimilation window, the reference velocity at

this location is close to $80\,\mathrm{m\,a^{-1}}$ (Fig. 6), so that the relative noise is $\sim 25\%$ and we don't expect too much improvement from the DA upstream as the velocity tends to 0.

Here the size of the state vector $\boldsymbol{x}$, N_x, is approximately 8400, *i.e.* z_s at every node and the basal conditions, b and C, in the grounded part. To test the performances of DA in conditions that would be numerically affordable for real applications, we run the assimilation with relatively small ensemble sizes $N_e = 30$, $N_e = 50$ and $N_e = 100$. In this case, inflation and localisation are required to counteract the effects of undersampling and we test a range of forgetting factors ρ and localisation radius r. The errors obtained at $t = 20\,\mathrm{a}$ relative to the errors from the initial ensemble mean are shown in Fig. 7. The performances of the assimilation for $N_e = 50$ and $N_e = 100$ are very similar. The filter diverges and produces errors larger than the initial errors for a localisation radius $r \leq 4\,\mathrm{km}$. However, for larger localisation radii, the assimilation is relatively robust for a wide range of r and ρ, with errors reduced by $\sim 30\%$ for b and $\sim 40\%$ for C. Decreasing the ensemble sizes reduces the filter performance but there is still a reduction of the errors by $\sim 20\%$ and $\sim 30\%$, respectively, with $N_e = 30$. For the two smallest ensembles, there is an optimal value for r and increasing r above this value decreases the filter performance. In general, this optimal value for r increases as ρ decreases, because the ensemble spread reduction induced by assimilating more observations is counterbalanced by the inflation.

In the sequel we discuss the results obtained with an ensemble size $N_e = 50$. As a compromise between the performances in retrieving b and C, we choose a forgetting factor $\rho = 0.92$ and a localisation radius $r = 8\,\mathrm{km}$. The evolution of the RMSEs as a function of assimilation time together with the initial and final ensembles are shown in Fig. 3. RMSE_b decreases steadily from $\sim 25\,\mathrm{m}$ for the initial ensemble at $t = 0$ to $\sim 12\,\mathrm{m}$ at $t = 35\,\mathrm{a}$. For the basal friction, RMSE_C is decreased by a factor 1.75 during the first ten years, then there is still a slight but much smaller improvement as new observations are assimilated.

At the end of the assimilation, for both fields, the spatial variations are well reproduced by the ensemble mean and, compared to the initial ensemble, the difference from the reference is decreased everywhere except between 300 and $325\,\mathrm{km}$ for C. The reduction in the error is also accompanied by a diminution of the ensemble spread, represented by the minimum and maximum values in Fig. 3. This reduction is the most important just upstream of the grounding line where the relative noise for the velocity is the smallest. For the first $100\,\mathrm{km}$ upstream of the grounding line, the ensemble standard deviation increases by a factor 4, from approximately 4 to $17\,\mathrm{m}$ for b and from 1.10^{-3} to $4.10^{-3}\,\mathrm{MPa\,m^{-\frac{1}{3}}\,a^{-\frac{1}{3}}}$ for C. Downstream of the GL where all members are floating, the model is insensitive to the basal conditions and the initial ensemble is unchanged.

We expect that uncertainties in the ice-sheet interior should not affect short-term forecast of the coastal regions (Durand et al., 2011), however for completeness we also show the results for the first $300\,\mathrm{km}$ in Fig. 8. For the bed there is only a small improvement of the ensemble mean with an RMSE decreasing from $50\,\mathrm{m}$ to $45\,\mathrm{m}$ after 35 years. Because the relative observation error on the velocity is very high in the first kilometres, the reduction of the ensemble spread due to the assimilation of new observations is very small and eventually outperformed by the inflation leading to an ensemble spread that becomes larger than before the assimilation. The model seems more sensitive to the basal friction and this effect is less pronounced for C with a continuous decrease of the RMSE and a small reduction of the ensemble spread everywhere.

Figure 2 shows that some members undergo a fast GL retreat of few kilometres before assimilation at the end of the first year. Interestingly, as the assimilation updates both the thickness and the bed, it also corrects the GL position which never departs by more than few nodes from the reference for the rest of the assimilation period.

As in realistic simulations, the true bed and friction are not available to assess the performance of the DA, we also look at the variables assimilated by the model. Figure 9 shows the RMSEs between the ensemble mean and the reference for the velocity u (RMSE_u) and the free surface zs (RMSE_{z_s}), computed for the entire domain ($0 \leq x \leq 800\,\text{km}$). We also report the evolution of the ensemble spread, computed as the square root of the averaged ensemble variance. The velocities before and after the analysis at $t = 1\,\text{a}$ and $t = 35\,\text{a}$ are shown in Fig. 6. The RMSEs are largely decreased during the first few years, especially for the velocity with an error of more than $300\,\text{m}\,\text{a}^{-1}$ before the first assimilation to approximately the noise level, $20\,\text{m}\,\text{a}^{-1}$, at $t = 20\,\text{a}$. For ~~zs~~z_s, RMSE_{z_s} is already below the noise level before the first analysis and decreases relatively steadily to reach $\sim 2\,\text{m}$ after 35 years. RMSE_u increases at the end of the period when the reference GL leaves the stable region. As can be shown in Fig. 6, the error is dominated by the larger difference over the ice-shelf due to the few members that still have their GL at the stable location, largely affecting the ensemble mean.

In general, during the first and last years of the assimilation period, the error and the ensemble spread increase during the forecast step. The analysis step reduces both the error and the ensemble spread (Fig. 9). With the stabilisation of the grounding line, both the error and the spread remain relatively stable during the forecast, and as RMSE_u and RMSE_{z_s} have already reached levels comparable to the observation noise, there is no much improvement during the analysis. After few assimilation steps, as expected for a reliable ensemble, the error and the spread have similar values.

4.2 ~~forecast simulations~~

Similar conclusions are drawn is the assimilation is pursued up to $t = 50\,\text{a}$. Because of the sensitivity of the ice-shelf velocities to the grounding line position, $RMSE_u$ shows a higher variability, but expect for few exceptions, stay close to the noise level. RME_b and $RMSE_c$ stagnate as continue to improve the reconstruction mostly in the first few tens of kilometres upstream of the GL where the relative noise on u is the smallest.

To asses the influence of the observation uncertainties in the performance of the DA, we repeat the experiment with the same localisation and inflation but different levels for the uncertainties on the observed surface velocity (σ_u^{obs}) and surface elevation ($\sigma_{z_s}^{obs}$) (*cf* section 3.2). We recall that these uncertainties are not correlated spatially and temporally. As shown in Fig. 10, the performance of the DA to retrieve both b and C increases when the uncertainty on the velocity observation σ_u^{obs} decreases. However, when looking at the model velocities and surface elevation, this improvement is not significant as the RMSEs were already below the noise level. As shown in Fig. 11, as expected, decreasing $\sigma_{z_s}^{obs}$ improves the analysis for the surface elevation. However, it doesn't necessarily reflect on the basal conditions and, on the contrary, reducing $\sigma_{z_s}^{obs}$ below $10\,\text{m}$ leads to an increase of $RMSE_C$ from 0.004 to $0.005\,\text{Mpa}\,\text{m}^{-1/3}\,\text{a}^{1/3}$. However, again, this effect does not reflect on the model velocities that are retrieved with the same accuracy.

4.2 Forecast simulations

We now discuss model projections from the initial state to $t = 200\,\mathrm{a}$.

Without assimilation, the deterministic forecast, i.e. using the ensemble mean basal conditions, rapidly leads to the fastest GL retreat and, after few years the GL position is no more included within the previsions from the ensemble (Fig.2B). This is due to the fact that the ensemble mean is smoother than the reference and any of the ensemble members. The reference GL position is included in the ensemble and at the end of the simulations most of the members are within $\pm 25\,\mathrm{km}$ from the reference. However, for few members the GL remains very stable near its initial position for tens to hundreds of years, eventually never switching to an unstable regime during the duration of the simulation. Retreat rates are relatively variable from one member to the other, depending on the basal conditions.

With assimilation, the ensemble mean is improved and the difference from the reference reduced. The deterministic forecast cannot be distinguished from the ensemble members any more (Fig.2C-D). Retreat rates are closer to the reference, with the previsions from all the ensemble members being more or less parallel to the reference. We note however that, when the forecast starts after an assimilation window of 20 years, i.e. during a period of stable GL position for the reference, the deterministic forecast leaves the stable position with a delay of approximately 25 years, and a few members remain stable for the entire simulation. On average between ~~t=13 to t=32 a~~$t = 13$ to $t = 32\,\mathrm{a}$, the thinning rate at the GL in the reference simulation, is approximately 0.6 m/a, reducing to 0.25 m/a during the last two years. The total thinning between two analyses is then much lower than the noise in the observed surface elevation and cannot be captured accurately by the DA. In addition, at the GL, the difference between the minimum and maximum bed elevation given by the ensemble is approximately 20 meters. This remaining uncertainty induces a difference of more than two meters for the floatation surface, and combined with the small thinning rates explains the delays in the initiation of the instability.

Extending the assimilation window up to $t = 35\,\mathrm{a}$ when the reference has switched in a fast retreat, allows to force all the members in the unstable retreat. There is a very good agreement between the reference and the deterministic forecast up to $t = 110\,\mathrm{a}$. This is also true for the ensemble and after that the spread is larger and the predicted GLs are less retreated than for the reference.

These results can be summarized by looking at the distribution of the ensemble forecasts for the grounding line position and volume above floatation (VAF) at $t = 100\,\mathrm{a}$ in Fig. 12 where the relative VAF change is computed as $(VAF_{t=100} - VAF^{ref}_{t=0})/VAF^{ref}_{t=0}$ with $VAF^{ref}_{t=0}$ the reference VAF at $t = 0$. As expected there is a clear correlation between grounding line retreat and mass loss, higher retreat leading to higher mass loss. The distributions are clearly non Gaussian, however, even without assimilation there is already a mode close to the reference. The mode is more pronounced, with more members close to the reference as observations are assimilated. As discussed before, with no assimilation or a short assimilation up to $t = 20\,\mathrm{a}$ before the unstable retreat, the deterministic forecast can be very different from the mode of the ensemble forecast. However they are very similar if the assimilation is pursued up to $t = 35\,\mathrm{a}$, within 1% for the relative volume loss or 5km for the GL position.

5 Discussion

Here, we have tested an ensemble Kalman Filter to assimilate annually observed surface velocities and surface elevation in a marine ice-sheet model. Similarly to previous studies, we have shown that, in fast flowing regions, it is possible to accurately separate and recover both the basal topography and basal friction from surface observations (Gudmundsson and Raymond, 2008; Goldberg and Heimbach, 2013; Bonan et al., 2014; Mosbeux et al., 2016). In view of our results, because the synthetic bed observations were already used once to generate the initial ensemble, it seems unnecessary to assimilate these same observations again during each analysis as in ~~Bonan et al. (2014)~~Bonan et al. (2014).

Using a scheme that assimilates time-dependent observations provides a model state consistent with transient changes and that can directly serve as an optimal initial condition to run forecast simulations without the need of an additional relaxation (Goldberg and Heimbach, 2013; Goldberg et al., 2015). Interestingly the position of the grounding line is also corrected during the analysis step, and the ensemble quickly converges within few grid nodes from the reference. In addition, the ensemble framework naturally allows to estimate and propagate the uncertainty of the estimated parameters. We have shown that the remaining uncertainties in the basal conditions do not significantly affect GL retreat rates once the unstable retreat is engaged but can lead to considerable delays in the initiation of the instability.

~~Contrary to the variational methods that require the linearisation of the forward model and the computation of the adjoint, the DA framework presented here is fairly independent of the ice flow model. This in an advantage with models that are still relatively young, continuously under development and with a variety of applications and configurations that can use different force balance approximations and parametrisations~~Good results have been obtained with relatively small ensembles (50 to 100 members) for a state vector of size $N_x \approx 8400$ and $N_y = 4002$ observations. Similarly to Bonan et al. (2014), we still see an improvement with a 30-members ensemble but the performances to retrieve the basal conditions are not as good. Running 2D plane view simulations with such ensemble sizes is largely possible as demonstrated by Ritz et al. (2015) who, using an hybrid shallow ice-shallow shelf model, have run a 200 years ensemble forecast of the whole Antarctic Ice Sheet using 3000 members.

We have used inflation and localisation to stabilise the filter. The inflation giving the best results in Bonan et al. (2014) ($\rho = 0.87 - 1.02$) is similar to the values tested in this study. For the localisation radius r we have used values between 4 and 16 km, while it ranges from 80 to 120 km in Bonan et al. (2014). While this seems counter-intuitive as the velocities depends only on the local conditions with the shallow ice approximation used by Bonan et al. (2014), in fact, because we use a different grid size ($dx = 0.2$km compared to $dx = 5$km in Bonan et al. (2014)), for each node we assimilate twice as much observations. Our results are in agreement with the adaptive localisation radius proposed by Kirchgessner et al. (2014). Using three different models, Kirchgessner et al. (2014) have shown that good performances are obtained when r is such that the effective local observation dimension, defined as the sum of the weights attributed to each observation during the local assimilation, is equal to the ensemble size. Here, the value $r = 8$km used for the 50 members-ensemble corresponds to an effective observation dimension of 56. Future studies should investigate if this result can be transposed to realistic 2D simulations with unstructured meshes.

In the experiments presented above, we have used a depth integrated model for the force balance equations where GL migration is implemented through a hydrostatic floatation condition. This allows to fully describe the ice topography with only one prognostic variable. Adaptation of the framework to a full-Stokes model requires minimum adaptations, however these models do not rely on the floatation condition and solve a proper contact problem for the grounding line migration (Durand et al., 2009), this implies to incorporate the two prognostic free surfaces z_b and z_s in the state vector. These models might be more sensitive to unbalanced geometries that could result from the analyses, especially when localisation is used (Cohn et al., 1998; Houtekamer and Mitchell, 2001). However, the ESTKF, as the ETKF, induces a minimal transformation of the ensemble members and thus has better chances to preserve balance (Nerger et al., 2012).

~~If the~~ Before generalizing such methods to real glacial systems, several points must be taken in consideration. They are independent of the DA method but they will eventually be treated differently in a variational or in an ensemble framework.

First, if the implementation is not an issue, the computational cost implied by running a full-Stokes model might remain a limiting factor. Compared to the Stokes solution, the SSA is known to overestimates the effects of bed topography perturbations on the surface profile for wavelengths less than few ice thicknesses (Gudmundsson, 2008). How this issue can affect the reconstruction of the basal properties has never been quantified, however snapshot basal friction inversions have shown that the solution is sensitive to the force balance approximation (Morlighem et al., 2010). In addition, the MISMIP experiments have shown that the GL position and its response to a perturbation depend on the force balance solved by the models (Pattyn et al., 2012, 2013). In real applications, the performance of ~~EnKFs~~ DA can be improved by explicitly taking into account the model error. Several strategies have been developed to account for this error, one approach with EnKFs being to use different versions of the model for different ensemble members (Houtekamer et al., 2009). Further studies could investigate the potential benefits of using ensembles that combine several force balance approximations.

~~Before generalizing such methods to real glacial systems, three points must be taken in consideration. They are independent of the DA method but they will eventually be treated differently in a variational or in an ensemble framework.~~

~~First, the~~ Second, the quality of the analysis and the accuracy of the error estimates depends on the observation error covariance matrix $\mathbf{R}$. It is then important to provide meaningful error estimates. Recent velocity maps provide an error estimates reported as the 1σ value for each individual location (Mouginot et al., 2012; Joughin et al., 2018). In general, this value agrees well with independent estimates, however care must be taken when the maps results from a composite of different sensors or different periods and in general it might be difficult to properly estimate $\mathbf{R}$.

~~Second, the~~ In a review paper Tandeo et al. (2018) illustrate the impacts of badly calibrated observation and model error covariance matrices in a sequential DA framework and discuss available methods and challenges for their joint estimation. For the question of the impact of systematic errors, *i.e.* bias, either in the model and in the observations, and their correction by augmenting the system state in variational and ensemble DA, interested readers are referred to Dee (2005).

Third, the results depends on prior assumptions on the control variables and their variability, represented here by the initial ensemble. For the basal topography, current reference maps provide local error estimates (Fretwell et al., 2013; Morlighem et al., 2017), however they do not provide informations about spatial correlations so that generating initial ensembles with the correct statistics might be problematic. In addition, the gridding can result in a loss of information for some regions of dense

measurements, or can lead to too smooth terrains in sparsely sampled areas. With the aim of generating terrains that have the correct high-resolution roughness, ~~Graham et al. (2017)~~ Graham et al. (2017) propose a synthetic 100-m resolution Antarctic bed elevation that combines the reference topography bedmap2 (Fretwell et al., 2013) with an unconditional simulation where the spatial correlation is fitted from dense radar measurements. This method could be used to generate initial ensembles but requires to have access to the initial high resolution measurements. Generating initial ensembles for the basal friction might be more problematic as there is in general no independent *a-priori* informations about the magnitude and spatial variability of the basal friction. If there is a correlation between the basal drag and the seismic observations of the bed conditions at large scale, a proper physical theory is still missing to quantitatively incorporate such informations in the models (Kyrke-Smith et al., 2017). It could be interesting to investigate how the existing multi-model basal friction reconstructions, based on snapshot inversions, could be used to derive initial uncertainty statistics and reduce the initial ensemble spread.

Finally, in our synthetic applications, we have not accounted for all potential sources of uncertainty which are, for example:

- the ice flow law: the ice viscosity depends on the englacial temperature which itself is function of the ice sheet history and the boundary forcing, including the geothermal heat flux (e.g. Van Liefferinge and Pattyn, 2013). Several other processes also affect the ice viscosity, including damage and strain-induced mechanical anisotropy (e.g. Pimienta et al., 1987; Schulson and Duval, 2009; Borstad et al., 2013). For the stress exponent, if the values $n = 3$ is used by most models, published values ranges between 1 and 5 (e.g. Gillet-Chaulet et al., 2011).

- The friction law: more and more direct or indirect evidences show that the friction under fast ice streams is at least partially controlled by the presence of sediments leading to a Coulomb type friction law (e.g. Tulaczyk et al., 2000; Murray, 1997; Joughin et al., 2010; Gillet-Chaulet et al., 2016). For hard beds, the development of subglacial cavities also implies deviations for the classical Weertman friction law (Schoof, 2005; Gagliardini et al., 2007).

- The density: the firn layer is not accounted for in most models, however its depth and density affect the floatation condition and thus the GL position (e.g. Griggs and Bamber, 2011). Direclty assimilating the GL position, using *e.g.* the moving mesh approach develop by ~~Bonan et al. (2017)~~Bonan et al. (2017), would certainly be beneficial in realistic applications to reduce the discrepancy between the modelled and observed GL (Goldberg et al., 2015).

- The external forcings from the atmosphere and the ocean: increasing mass loss rates from the ice sheets, in a large portion, can be attributed to a response to oceanic forcing, but multiple challenges remain for a proper assessment of their magnitude (Joughin et al., 2012).

Realistic simulations with ice flow models cover a wide range of spatial and temporal scales, and the relative importance of these uncertainties as well as their representation in the models will certainly have to be evaluated partly in a case by case basis, requiring to develop robust framework for a variety of applications.

6 Conclusions

Developing model initialisation strategies that properly reproduce the ice-sheets dynamical mass losses observed over the last decades, requires to develop transient assimilation frameworks that are able to account for the growing availability of dense time series, especially from space observations. Here, we presented a synthetic twin experiment demonstrating the possibility to calibrate a marine ice model using an ensemble Kalman Filter which requires less numerical developments than variational methods.

Using resolutions and noise levels consistent with current observing systems, good performances are obtained to recover both the basal friction and basal topography with an ensemble of at least 50 members. Localisation and inflation have been tuned manually, however the results are consistent over relatively wide ranges. Future studies should investigate how these values can be transposed to realistic applications. Nevertheless, there is an abundant and growing literature in other geophysical fields to overcome ~~problem~~ problems that we might be facing in future studies.

Once the GL enters an unstable region, retreat rates largely depends on the basal conditions, thus using DA to reduce the associated uncertainties largely increases the skill of the model to predict rates and magnitude of GL retreat for time scales relevant for sea level rise projections. In our simplified application, the assimilation of the surface observations was sufficient to capture the GL migration during the assimilation window, without explicitely assimilating the observed position. However, for the GL to enter a irreversible retreat, the thickness must reach a tipping point, *i.e.* the thickness at the GL must reach floatation. This can seriously impact the predictability of the system as, for small perturbations, remaining uncertainties on the basal conditions can lead to an uncertainty on the residence time of the GL on stabilisation points, that can be similar to the simulation timescale.

Finally, we have discussed the main challenges to tackle before generalizing transient DA in ice-sheet modelling. This includes a better assessment of the uncertainties in the model and in the observations used for the background and for the assimilation.

Code availability. Elmer/Ice code is publicly available through GitHub (https://github.com/ElmerCSC/elmerfem, Gagliardini et al. (2013)). PDAF is distributed under the GNU General Public License, version 3, and is available at http://pdaf.awi.de.

Appendix A: Notations

Competing interests. The authors declare that they have no competing interests.

Table A1. Notations and values used in this study associated with the ice flow model

Prognostic variables:		
$H = z_s - z_b$	m	Thickness
z_s	m	top surface elevation
z_s	m	bottom surface elevation
Diagnostic variable:		
u	$\mathrm{m\,a^{-1}}$	horizontal velocity
Parameters:		
$a_b = 0.0$	$\mathrm{m\,a^{-1}}$	basal melting
$a_s = 0.5$	$\mathrm{m\,a^{-1}}$	surface accumulation
b	m	bed elevation
$B = 0.4$	$\mathrm{Mpa\,a^{1/3}}$	ice rigidity
C	m	basal friction coefficient
$m = 1/3$		friction law exponent
$n = 3$		Glen's creep exponent
$\rho_i = 900$	$\mathrm{kg\,m^3}$	ice density
$\rho_w = 1000$	$\mathrm{kg\,m^3}$	sea water density
Numerical parameters:		
$dt = 5\,10^{-3}$	a	model time step
$dx = 200$	m	mesh resolution

Table A2. Notations and values used in this study associated with the ensemble filter

Variables:		
$\boldsymbol{x} = (z_s, b, C)$		state vector
$\mathbf{P}$		covariance matrix
Stabilisation parameters:		
r	m	localisation radius
ρ		forgetting factor
Sizes:		
N_e		ensemble size
N_x		state vector size
N_y		observation vector size
Others:		
$\Delta t = 1$	a	time interval between two analyses

Acknowledgements. This work was funded by the French National Research Agency (ANR) through the TROIS-AS (ANR-15-CE01-0005-01). The author thanks G. Durand, O. Gagliardini and J. Mouginot for valuable comments on the first drafts of the manuscript.

References

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, https://doi.org/10.1175/1520- 0493(1999)127<2741:A AMCIOT>2.0.CO;2, 1999.

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation: Ensemble-variational Data Assimilation, Quarterly Journal of the Royal Meteorological Society, 143, 607–633, https://doi.org/10.1002/qj.2982, 2017.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Monthly Weather Review, 129, 420–436, https://doi.org/10.1175/1520-0493, 2001.

Bonan, B., Nodet, M., Ritz, C., and Peyaud, V.: An ETKF approach for initial state and parameter estimation in ice sheet modelling, Nonlin. Processes Geophys., 21, 569–582, 2014.

Bonan, B., Nichols, N. K., Baines, M. J., and Partridge, D.: Data assimilation for moving mesh methods with an application to ice sheet modelling, Nonlin. Processes Geophys., 24, 515–534, https://doi.org/10.5194/npg-24-515-2017, 2017.

Borstad, C. P., Rignot, E., Mouginot, J., and Schodlok, M. P.: Creep deformation and buttressing capacity of damaged ice shelves: theory and application to Larsen C ice shelf, The Cryosphere, 7, 1931–1947, https://doi.org/10.5194/tc-7-1931-2013, 2013.

Brinkerhoff, D.J., Aschwanden, A., Truffer, M.: Bayesian Inference of Subglacial Topography Using Mass Conservation. Front. Earth Sci. 4:8, https://doi.org/10.3389/feart.2016.00008, 2016

Brondex, J., Gagliardini, O., Gillet-Chaulet, F., and Durand, G.: Sensitivity of grounding line dynamics to the choice of the friction law, J. Glaciol., 63, 854–866, https://doi.org/10.1017/jog.2017.51, 2017.

Brondex, J., Gillet-Chaulet, F., and Gagliardini, O.: Sensitivity of centennial mass loss projections of the Amundsen basin to the friction law, The Cryosphere, 13, 177–195, https://doi.org/10.5194/tc-13-177-2019, 2019

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, https://doi.org/10.1175/1520- 0493(1998)126<1719:ASITEK>2.0.CO;2, 1998.

Carrassi, A, Bocquet, M, Bertino, L, Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, WIREs Clim Change, https://doi.org/0.1002/wcc.535, 2018.

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., Merrifield, M. A., Milne, G. A., Nerem, R. S., Nunn, P. D., Payne, A. J., Pfeffer, W. T., Stammer, D., and Unnikrishnan: Sea level change, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Cohn, S. E., da Silva, A., Guo, J., Sienkiewicz, M., and Lamich, D.: Assessing the Effects of Data Selection with the DAO Physical-Space Statistical Analysis System*, Monthly Weather Review, 126, 2913–2926, https://doi.org/10.1175/1520-0493, 1998.

Cosme E., Verron J., Brasseur P., Blum J. and Auroux D. : Smoothing problems in a Bayesian framework and their linear Gaussian solutions. Mon. Wea. Rev. 140: 683—695, 2012.

Dai, C. and Howat, I. M.: Measuring Lava Flows With ArcticDEM: Application to the 2012-2013 Eruption of Tolbachik, Kamchatka, Geophys. Res. Lett., 44, 12,133–12,140, https://doi.org/10.1002/2017GL075920, 2017.

Dee, D. P.: Bias and data assimilation, Q. J. Roy. Meteor. Soc., 131, 3323–3343, https://doi.org/10.1256/qj.05.137, 2005.

Durand, G. and Pattyn, F.: Reducing uncertainties in projections of Antarctic ice mass loss, The Cryosphere, 9, 2043–2055, https://doi.org/10.5194/tc-9-2043-2015, 2015.

Durand, G., Gagliardini, O., Zwinger, T., Le Meur, E., and Hindmarsh, R. C.: Full Stokes modeling of marine ice sheets: influence of the grid size, Annals Glaciol., 50, 109–114, https://doi.org/10.3189/172756409789624283, 2009.

Durand, G., Gagliardini, O., Favier, L., Zwinger, T., and Meur, E. l.: Impact of bedrock description on modeling ice sheet dynamics, Geophys. Res. Lett., 38, https://doi.org/201110.1029/2011GL048892, 2011.

Enderlin, E. M., Howat, I. M., and Vieli, A.: High sensitivity of tidewater outlet glacier dynamics to shape, The Cryosphere, 7, 1007–1015, https://doi.org/10.5194/tc-7-1007-2013, 2013.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10 143, https://doi.org/10.1029/94JC00572, 1994.

Evensen G., van Leeuwen P.J.:An ensemble Kalman smoother for nonlinear dynamics. Mon. Wea. Rev., 128, 1852–1867, 2000.

Favier, L., Durand, G., Cornford, S. L., Gudmundsson, G. H., Gagliardini, O., Gillet-Chaulet, F., Zwinger, T., Payne, A. J., and Le Brocq, A. M.: Retreat of Pine Island Glacier controlled by marine ice-sheet instability, Nat. Clim. Change, 4, 117–121, https://doi.org/10.1038/nclimate2094, 2014.

Fournier, A., D. Fussell, and L. Carpenter: Computer rendering of stochastic models, Commun. Assoc. Cornput. Mach., 25, 371-384, 1982.

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., Bianchi, C., Bingham, R. G., Blankenship, D. D., Casassa, G., Catania, G., Callens, D., Conway, H., Cook, A. J., Corr, H. F. J., Damaske, D., Damm, V., Ferraccioli, F., Forsberg, R., Fujita, S., Gim, Y., Gogineni, P., Griggs, J. A., Hindmarsh, R. C. A., Holmlund, P., Holt, J. W., Jacobel, R. W., Jenkins, A., Jokat, W., Jordan, T., King, E. C., Kohler, J., Krabill, W., Riger-Kusk, M., Langley, K. A., Leitchenkov, G., Leuschen, C., Luyendyk, B. P., Matsuoka, K., Mouginot, J., Nitsche, F. O., Nogi, Y., Nost, O. A., Popov, S. V., Rignot, E., Rippin, D. M., Rivera, A., Roberts, J., Ross, N., Siegert, M. J., Smith, A. M., Steinhage, D., Studinger, M., Sun, B., Tinto, B. K., Welch, B. C., Wilson, D., Young, D. A., Xiangbin, C., and Zirizzotti, A.: Bedmap2: improved ice bed, surface and thickness datasets for Antarctica, The Cryosphere, 7, 375–393, https://doi.org/10.5194/tc-7-375-2013, 2013.

Gagliardini, O., Cohen, D., Råback, P., and Zwinger, T.: Finite-element modeling of subglacial cavities and related friction law, J. Geophys. Res., 112, https://doi.org/10.1029/2006JF000576, 2007.

Gagliardini, O., Zwinger, T., Gillet-Chaulet, F., Durand, G., Favier, L., de Fleurian, B., Greve, R., Malinen, M., Martín, C., Råback, P., Ruokolainen, J., Sacchettini, M., Schäfer, M., Seddik, H., and Thies, J.: Capabilities and performance of Elmer/Ice, a new-generation ice sheet model, Geosci. Model Dev. 6, 1299–1318, https://doi.org/10.5194/gmd-6-1299-2013, 2013.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, https://doi.org/10.1002/qj.49712555417, 1999.

Gillet-Chaulet, F., Hindmarsh, R. C. A., Corr, H. F. J., King, E. C., and Jenkins, A.: In-situ quantification of ice rheology and direct measurement of the Raymond Effect at Summit, Greenland using a phase-sensitive radar, Geophys. Res. Lett., 38, https://doi.org/201110.1029/2011GL049843, 2011.

Gillet-Chaulet, F., Gagliardini, O., Seddik, H., Nodet, M., Durand, G., Ritz, C., Zwinger, T., Greve, R., and Vaughan, D. G.: Greenland ice sheet contribution to sea-level rise from a new-generation ice-sheet model, The Cryosphere, 6, 1561–1576, https://doi.org/10.5194/tc-6-1561-2012, 2012.

Gillet-Chaulet, F., Durand, G., Gagliardini, O., Mosbeux, C., Mouginot, J., Rémy, F., and Ritz, C.: Assimilation of surface velocities acquired between 1996 and 2010 to constrain the form of the basal friction law under Pine Island Glacier, Geophys. Res. Lett., 43, 2016GL069 937, https://doi.org/10.1002/2016GL069937, 2016.

Gladstone, R. M., Payne, A. J., and Cornford, S. L.: Resolution requirements for grounding-line modelling: sensitivity to basal drag and ice-shelf buttressing, Annals Glaciol., 53, 97–105, https://doi.org/10.3189/2012AoG60A148, 2012.

Goldberg, D. N. and Heimbach, P.: Parameter and state estimation with a time-dependent adjoint marine ice sheet model, The Cryosphere, 7, 1659–1678, https://doi.org/10.5194/tc-7-1659-2013, 2013.

Goldberg, D. N., Heimbach, P., Joughin, I., and Smith, B.: Committed retreat of Smith, Pope, and Kohler Glaciers over the next 30 years inferred by transient model calibration, The Cryosphere, 9, 2429–2446, https://doi.org/10.5194/tc-9-2429-2015, 2015.

Goldberg, D. N., Narayanan, S. H. K., Hascoet, L., and Utke, J.: An optimized treatment for algorithmic differentiation of an important glaciological fixed-point problem, Geosci. Model Dev., 9, 1891–1904, https://doi.org/10.5194/gmd-9-1891-2016, 2016.

Graham, F. S., Roberts, J. L., Galton-Fenzi, B. K., Young, D., Blankenship, D., and Siegert, M. J.: A high-resolution synthetic bed elevation grid of the Antarctic continent, Earth Syst. Sci. Data, 9, 267-279, https://doi.org/10.5194/essd-9-267-2017, 2017.

Griggs, J. and Bamber, J.: Antarctic ice-shelf thickness from satellite radar altimetry, J. Glaciol., 57, 485–498, https://doi.org/10.3189/002214311796905659, 2011.

Gudmundsson, G.H. and Raymond, M.: On the limit to resolution and information on basal properties obtainable from surface data on ice streams. The Cryosphere 2, 167—178, 2008.

Gudmundsson, G. H.: Analytical solutions for the surface response to small amplitude perturbations in boundary data in the shallow-ice-stream approximation, The Cryosphere, 2, 77–93, 2008.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Mon. Weather Rev., 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2, 2001.

Hascoët, L. and Morlighem, M.: Source-to-source adjoint Algorithmic Differentiation of an ice sheet model written in C, Optimization Methods and Software, 33, 829–843, https://doi.org/10.1080/10556788.2017.1396600, 2018.

Hendricks Franssen, H. J., Kaiser, H. P., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R., and Kinzelbach, W.: Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating, Water Resour. Res., 47, W02532, https://doi.org/10.1029/2010WR009480, 2011.

Houtekamer, P. L., and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., 126, 796-811, https://doi.org/10.1175/1520- 0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 129, 123–137, https://doi.org/10.1175/1520-0493, 2001.

Houtekamer, P. L., Mitchell, H. L., and Deng, X.: Model Error Representation in an Operational Ensemble Kalman Filter, Monthly Weather Review, 137, 2126–2143, https://doi.org/10.1175/2008MWR2737.1, 2009.

Houtekamer, P.L., X. Deng, H.L. Mitchell, S. Baek, and Gagnon, N.: Higher Resolution in an Operational Ensemble Kalman Filter. Mon. Wea. Rev., 142, 1143–1162, https://doi.org/10.1175/MWR-D-13-00138.1, 2014

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D: Nonlinear Phenomena, 230, 112–126, https://doi.org/10.1016/j.physd.2006.11.008, 2007

Jenkins, A., Shoosmith, D., Dutrieux, P., Jacobs, S., Kim, T. W., Lee, S. H., Ha, H. K., and Stammerjohn, S.: West Antarctic Ice Sheet retreat in the Amundsen Sea driven by decadal oceanic variability, Nat. Geosci., 11, 733–738, https://doi.org/10.1038/s41561-018-0207-4, 2018.

Joughin, I., Smith, B. E., and Holland, D. M.: Sensitivity of 21st century sea level to ocean-induced thinning of Pine Island Glacier, Antarctica, Geophys. Res. Lett., 37, https://doi.org/10.1029/2010GL044819, 2010.

Joughin, I., Alley, R. B., and Holland, D. M.: Ice-Sheet Response to Oceanic Forcing, Science, 338, 1172–1176, https://doi.org/10.1126/science.1226481, 2012.

Joughin, I., Smith, B. E., and Howat, I.: Greenland Ice Mapping Project: ice flow velocity variation at sub-monthly to decadal timescales, The Cryosphere, 12, 2211–2227, https://doi.org/10.5194/tc-12-2211-2018, 2018.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?, Tellus A: Dynamic Meteorology and Oceanography, 59, 758–773, https://doi.org/10.1111/j.1600-0870.2007.00261.x, 2007.

Kirchgessner, P., Nerger, L. and Bunse-Gerstner, A.,: On the Choice of an Optimal Localization Radius in Ensemble Kalman Filter Methods, Mon. Wea. Rev., 142, 2165—2175, https://doi.org/0.1175/MWR-D-13-00246.1, 2014

Kyrke-Smith, T.M., Gudmundsson, G.H., Farrell, P.E.: Can Seismic Observations of Bed Conditions on Ice Streams Help Constrain Parameters in Ice Flow Models? J. Geophys. Res.: Earth Surface 122, 2269—2282. https://doi.org/doi.org/10.1002/2017JF004373, 2017

Larour, E., Morlighem, H. S. a. M., and Rignot, E.: Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM), J. Geophys. Res., 117, https://doi.org/201210.1029/2011JF002140, 2012.

Larour, E., Utke, J., Csatho, B., Schenk, A., Seroussi, H., Morlighem, M., Rignot, E., Schlegel, N., and Khazendar, A.: Inferred basal friction and surface mass balance of the Northeast Greenland Ice Stream using data assimilation of ICESat (Ice Cloud and land Elevation Satellite) surface altimetry and ISSM (Ice Sheet System Model), The Cryosphere, 8, 2335–2351, https://doi.org/10.5194/tc-8-2335-2014, 2014.

Larour, E., Utke, J., Bovin, A., Morlighem, M., and Perez, G.: An approach to computing discrete adjoints for MPI-parallelized models applied to Ice Sheet System Model 4.11, Geosci. Model Dev., 9, 3907–3918, https://doi.org/10.5194/gmd-9-3907-2016, 2016.

Li, Z. and Navon, I. M.: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. Q.J.R. Meteorol. Soc., 127: 661–683. https://doi.org/10.1002/qj.49712757220, 2001

MacAyeal, D. R.: Large-scale ice flow over a viscous basal sediment: Theory and application to ice stream B, Antarctica, J. Geophys. Res.: Solid Earth, 94, 4071–4087, https://doi.org/10.1029/JB094iB04p04071, 1989.

MacAyeal, D. R.: A tutorial on the use of control methods in ice-sheet modeling, J. Glaciol., 39, 91–98, 1993.

Mitchell HL, Houtekamer PL, Pellerin G.: Ensemble size,balance, and model-error representation in an Ensemble Kalman Filter. Mon. Weather Rev., 130, 2791–2808, https://doi.org/10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2, 2002..

Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Dhia, H. B., and Aubry, D.: Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models for Pine Island Glacier, West Antarctica, Geophys. Res. Lett., 37, https://doi.org/10.1029/2010GL043853, 2010.

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., Catania, G., Chauché, N., Dowdeswell, J. A., Dorschel, B., Fenty, I., Hogan, K., Howat, I., Hubbard, A., Jakobsson, M., Jordan, T. M., Kjeldsen, K. K., Millan, R., Mayer, L., Mouginot, J., Noël, B. P. Y., O'Cofaigh, C., Palmer, S., Rysgaard, S., Seroussi, H., Siegert, M. J., Slabon, P., Straneo, F., van den Broeke, M. R., Weinrebe, W., Wood, M., and Zinglersen, K. B.: BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation, Geophys. Res. Lett., 44, 11,051–11,061, https://doi.org/10.1002/2017GL074954, 2017.

Mosbeux, C., Gillet-Chaulet, F., Gagliardini, O.: Comparison of adjoint and nudging methods to initialise ice sheet model basal conditions. Geosci. Model Dev. 9, 2549–2562, https://doi.org/10.5194/gmd-9-2549-2016, 2016.

Mouginot, J., Scheuchl, B., and Rignot, E.: Mapping of Ice Motion in Antarctica Using Synthetic-Aperture Radar Data, Remote Sensing, 4, 2753–2767, https://doi.org/10.3390/rs4092753, 2012.

Mouginot, J., Rignot, E., Scheuchl, B., and Millan, R.: Comprehensive Annual Ice Sheet Velocity Mapping Using Landsat-8, Sentinel-1, and RADARSAT-2 Data, Remote Sensing 2017, 9(4), , https://doi.org/10.3390/rs9040364, 2017

Murray, T.: Assessing the paradigm shift: deformable glacier beds, Quaternary Science Reviews, 16, 996–1016, 1997.

Nerger, L., Hiller, W., and Schröter, J.: A comparison of error subspace Kalman filters, Tellus A: Dynamic Meteorology and Oceanography, 57:5, 715-735, https://doi.org/10.3402/tellusa.v57i5.14732

Nerger, L., Hiller, W., and Schröter, J.: PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman Filtering, in: Use of High Performance Computing in Meteorology, pp. 63–83, WORLD SCIENTIFIC, Reading, UK, https://doi.org/10.1142/9789812701831_0006, 2005.

Nerger, L., Danilov, S., Hiller, W., and Schröter, J.: Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter, Ocean Dynamics, 56, 634–649, https://doi.org/10.1007/s10236-006-0083-0, 2006.

Nerger, L., Schröter, J., and Hiller, W.: A Unification of Ensemble Square Root Kalman Filters, Monthly Weather Review, 140, 2335–2345, https://doi.org/10.1175/MWR-D-11-00102.1, 2012.

Nerger, L. , Schulte, S. and Bunse-Gerstner, A. : On the influence of model nonlinearity and localization on ensemble Kalman smoothing , Q. J. R. Meteorol. Soc., 140, 2249–2259, https://doi.org/10.1002/qj.2293, 2014.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, A.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus A, 56, 415–428, https://doi.org/10.1111/j.1600- 0870.2004.00076.x, 2004.

Pattyn, F., Huyghe, A., De Brabander, S., and De Smedt, B.: Role of transition zones in marine ice sheet dynamics, J. Geophysi. Res., 111, https://doi.org/10.1029/2005JF000394,2006.

Pattyn, F., Schoof, C., Perichon, L., Hindmarsh, R. C. A., Bueler, E., de Fleurian, B., Durand, G., Gagliardini, O., Gladstone, R., Goldberg, D., Gudmundsson, G. H., Huybrechts, P., Lee, V., Nick, F. M., Payne, A. J., Pollard, D., Rybak, O., Saito, F., and Vieli, A.: Results of the Marine Ice Sheet Model Intercomparison Project, MISMIP, The Cryosphere, 6, 573–588, https://doi.org/10.5194/tc-6-573-2012, 2012.

Pattyn, F., Perichon, L., Durand, G., Favier, L., Gagliardini, O., Hindmarsh, R. C., Zwinger, T., Albrecht, T., Cornford, S., Docquier, D., FüRst, J. J., Goldberg, D., Gudmundsson, G. H., Humbert, A., HüTten, M., Huybrechts, P., Jouvet, G., Kleiner, T., Larour, E., Martin, D., Morlighem, M., Payne, A. J., Pollard, D., RüCkamp, M., Rybak, O., Seroussi, H., Thoma, M., and Wilkens, N.: Grounding-line migration in plan-view marine ice-sheet models: results of the ice2sea MISMIP3d intercomparison, J. Glaciol., 59, 410–422, https://doi.org/10.3189/2013JoG12J129, 2013.

Pebesma, E. J. and Wesseling, C. G.: Gstat: a program for geostatistical modelling, prediction and simulation, Computers & Geosciences, 24, 17–31, https://doi.org/10.1016/S0098-3004(97)00082-4, 1998.

Pham, D. T., Verron, J., and Christine Roubaud, M.: A singular evolutive extended Kalman filter for data assimilation in oceanography, Journal of Marine Systems, 16, 323–340, https://doi.org/10.1016/S0924-7963(97)00109-7, 1998.

Pimienta, P., Duval, P., and Lipenkov, V. Y.: Mechanical behavior of anisotropic polar ice, Internationnal Association of Hydrological Sciences Publication 170 (Symposium at Vancouver 1987 - Physical Basis of Ice Sheet Modelling), 57–66, 1987

Pralong, M.R., Gudmundsson, G.H.,: Bayesian estimation of basal conditions on Rutford Ice Stream, West Antarctica, from surface data. J. Glaciol. 57, https://doi.org/10.3189/002214311796406004, 2011

Ritz, C., Edwards, T. L., Durand, G., Payne, A. J., Peyaud, V., and Hindmarsh, R. C. A.: Potential sea-level rise from Antarctic ice-sheet instability constrained by observations, Nature, https://doi.org/10.1038/nature16147, 2015.

Sakov, P. and Bertino, L.: Relation between two common localisation methods for the EnKF, Comput. Geosci. 15, 225–237, https://doi.org/0.1007/s10596-010-9202-6, 2011.

Sakov, P., Counillon, F., Bertino, L., Lisaeter, K. A., Oke, P. R., and Korablev, A.: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Sci., 8, 633-656, https://doi.org/10.5194/os-8-633-2012, 2012.

Scambos, T., Bell, R., Alley, R., Anandakrishnan, S., Bromwich, D., Brunt, K., Christianson, K., Creyts, T., Das, S., DeConto, R., Dutrieux, P., Fricker, H., Holland, D., MacGregor, J., Medley, B., Nicolas, J., Pollard, D., Siegfried, M., Smith, A., Steig, E., Trusel, L., Vaughan, D., and Yager, P.: How much, how fast?: A science review and outlook for research on the instability of Antarctica's Thwaites Glacier in the 21st century, Global and Planetary Change, 153, 16–34, https://doi.org/10.1016/j.gloplacha.2017.04.008, 2017.

Schoof, C.: The effect of cavitation on glacier sliding, Proc. R. Soc. A, 461, 609–627, https://doi.org/10.1098/rspa.2004.1350, 2005.

Schoof, C.: Marine ice-sheet dynamics. Part 1. The case of rapid sliding, J. Fluid Mech., 573, 27, https://doi.org/10.1017/S0022112006003570, 2007.

Schulson, E. M. and Duval, P.: Creep and Fracture of Ice, Cambridge University Press, 2009.

Seroussi, H., Morlighem, M., Rignot, E., Larour, E., Aubry, D., Dhia, H. B., and Kristensen, S. S.: Ice flux divergence anomalies on 79north Glacier, Greenland, Geophys. Res. Lett., 38, https://doi.org/10.1029/2011GL047338, 2011.

Seroussi, H., Morlighem, M., Larour, E., Rignot, E., and Khazendar, A.: Hydrostatic grounding line parameterization in ice sheet models, The Cryosphere, 8, 2075–2087, https://doi.org/10.5194/tc-8-2075-2014, 2014.

Sun, L., Seidou, O., Nistor, I. and Liu, K.: Review of the Kalman-type hydrological data assimilation, Hydrol. Sci. J., 61(13), 2348–2366, https://doi.org/10.1080/02626667.2015.1127376, 2016

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M. and Zhen, Y.: Joint Estimation of Model and Observation Error Covariance Matri- ces in Data Assimilation: a Review. Mon. Weather Rev., submitted, available at: https://arxiv.org/abs/1807.11221v2, 2018.

Tsai, V. C., Stewart, A. L., and Thompson, A. F.: Marine ice-sheet profiles and stability under Coulomb basal conditions, J. Glaciol., 61, 205–215, https://doi.org/10.3189/2015JoG14J221, 2015.

Tulaczyk, S., Kamb, W. B., and Engelhardt, H. F.: Basal mechanics of Ice Stream B, West Antarctica 1. Till mechanics, J. Geophys. Res., 105, 463–481, 2000.

Van Liefferinge, B. and Pattyn, F.: Using ice-flow models to evaluate potential sites of million year-old ice in Antarctica, Clim. Past, 9, 2335–2345, https://doi.org/10.5194/cp-9-2335-2013, 2013.

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., Kirchgessner, P., and Beckers, J.-M.: State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems, Tellus A: Dynamic Meteorology and Oceanography, 70, 1445 364, https://doi.org/10.1080/16000870.2018.1445364, 2018.

Vieli, A. and Payne, A. J.: Application of control methods for modelling the flow of Pine Island Glacier, West Antarctica, Annals Glaciol., 36, 197–204, 2003.

Vieli, A. and Payne, A. J.: Assessing the ability of numerical ice sheet models to simulate grounding line migration, J. Geophys. Res.: Earth Surface, 110, F01 003, https://doi.org/10.1029/2004JF000202, 2005.

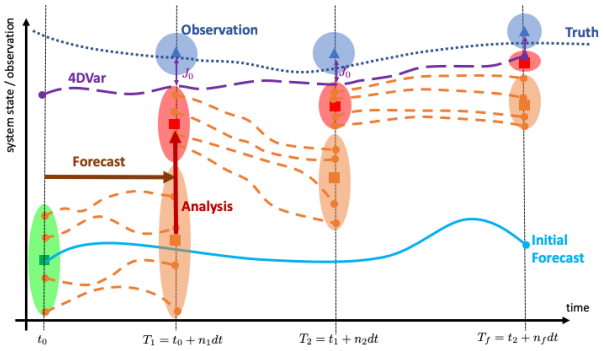

Figure 1. Principle of data assimilation (Adapted from Carrassi et al. (2018)). Having a physical model able to forecast the evolution of a system from time $t = t_0$ to time $t = T_f$ (cyan curve), the aim of DA is to use available observations (blue triangles) to correct the model projections and get closer to the (unknown) truth (dotted line). In EnKFs, the initial system state and its uncertainty (green square and ellipsoid) is represented by N_e members. The members are propagated forward in time during n_1 model time steps dt to $t = T_1$ where observations are available (Forecast phase, orange dashed lines). At $T = t_1$ the analysis uses the observations and their uncertainty (blue triangle and ellipsoid) to produce a new system state that is closer to the observations and with a lower uncertainty (red square and ellipsoid). A new forecast is issued from the analysed state and this procedure is repeated until the end of the assimilation window at $t = T_f$. The model state should get closer to the truth and with lower uncertainty as more observations are assimilated. Time dependent variational methods (4D-Var) iterate over the assimilation window to find the trajectory that minimizes the misfit (J_0) between the model and all observations available from t_0 to T_f (violet curve). For linear dynamics, Gaussian errors and infinite ensemble sizes, the states produced at the end of the assimilation window by the two methods should be equivalent (Li and Navon, 2001).

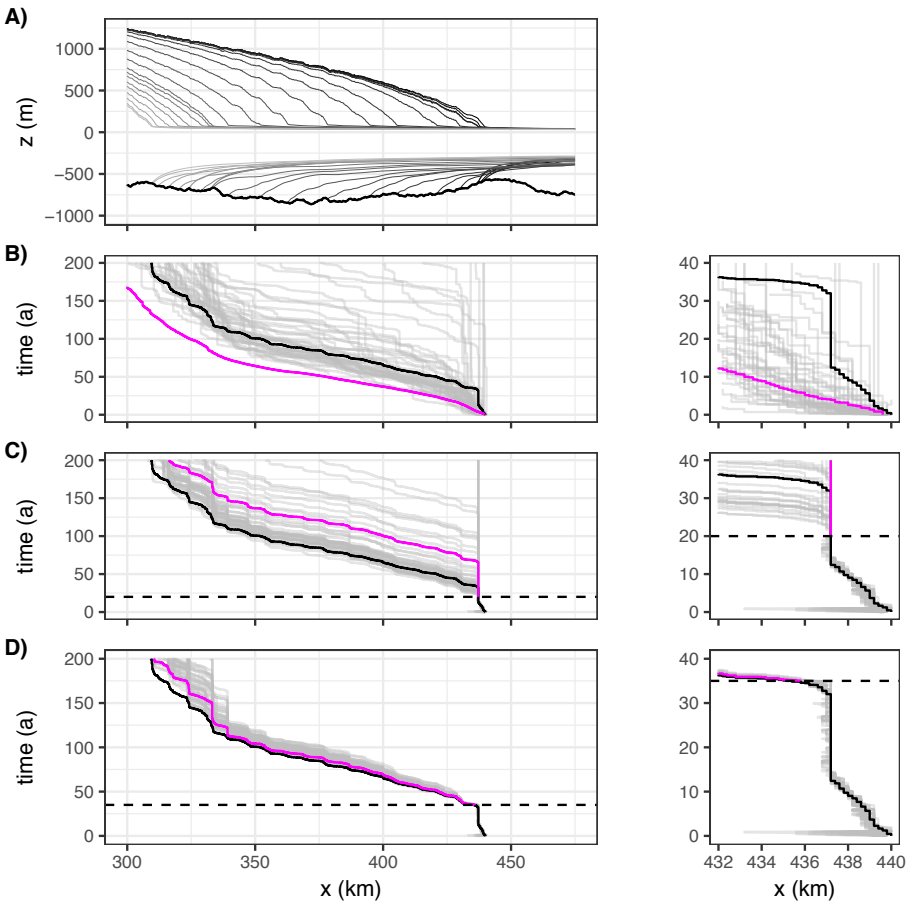

Figure 2. (A) Reference ice sheet topography every 10 years from $t = 0$ to $t = 200\,\mathrm{a}$ (black to grey). (B-D) GL position as a function of the simulation time for the reference (black line), for the ensemble (grey lines), and for the deterministic forecast (magenta line) (B) without assimilation, (C) with assimilation up to $t = 20\,\mathrm{a}$ and (D) with assimilation up to $t = 35\,\mathrm{a}$. The right column show a zoom on the first 40 years. In C-D), the horizontal dashed line show the end of the assimilation window.

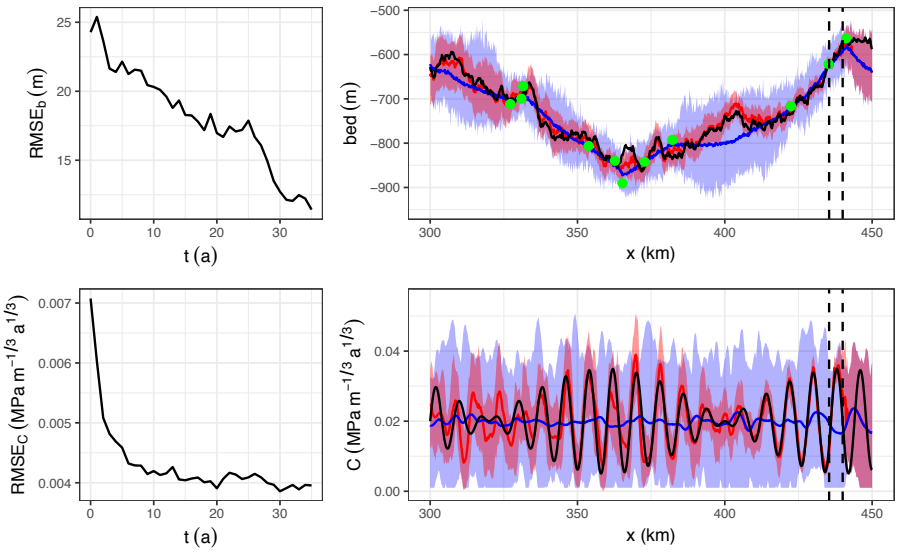

Figure 3. (left) RMSE between the reference and the analysed ensemble mean for the bed and friction coefficient. (Right) Bed and friction coefficient, the reference is shown in black, the synthetic bed measurements in the top panel are shown as green dots, the ensemble mean before assimilation is in blue and at $t = 35\,\mathrm{a}$ in red. The shading shows the ensemble spread between the minimum and maximum values, before assimilation (blue) and at $t = 35\,\mathrm{a}$ (red). The dashed vertical lines show the GL position at $t = 0$ and $t = 35\,\mathrm{a}$.

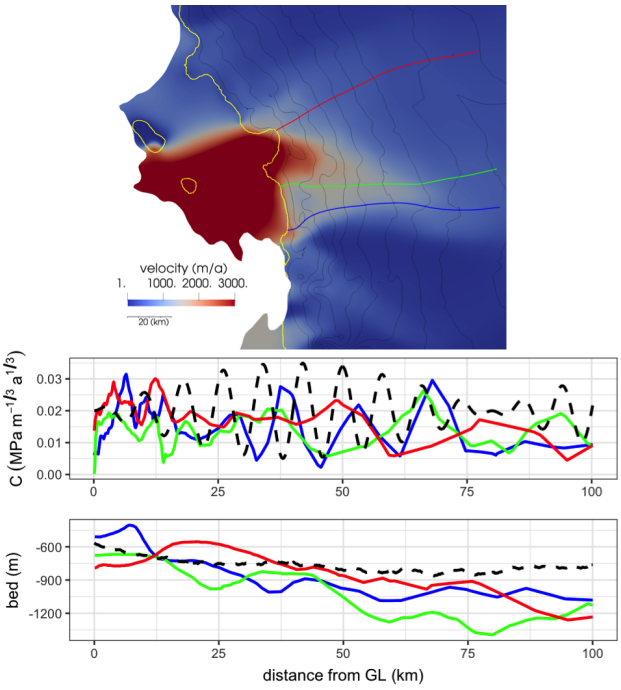

Figure 4. Thwaites Glacier (Antarctica). Model results from Brondex et al. (2018): model velocities (top) and friction coefficient C and bed elevation b extracted along three streamlines (same color code). Synthetic values used in this study are shown with black dashed lines. Note that the mesh resolution varies from $\backsim$ 200m close to the GL, shown in yellow in the top panel, to $\backsim$ 10km at the upstream end of the streamlines.

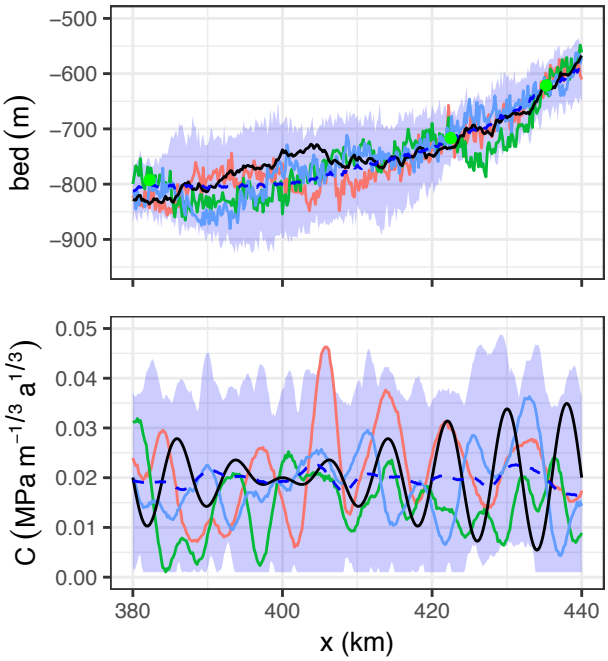

Figure 5. Initial ensemble for the bed and friction coefficient, the reference is shown in black, the synthetic bed measurements are shown as green dots in the top panel, the ensemble mean is the dashed blue curve and the shading shows the ensemble spread. Coloured solid lines show the first 3 members.

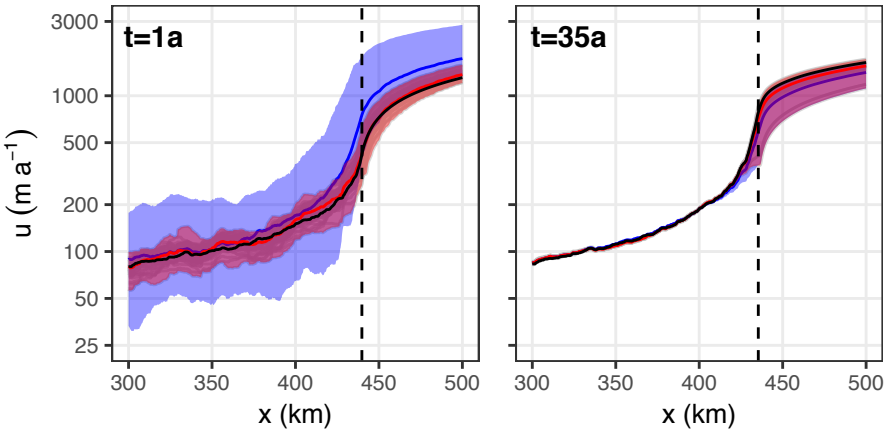

Figure 6. Velocity, u, at $t = 1$a and $t = 35$a. The reference is in black, the ensemble mean before and after the analysis is in blue and red, respectively. The shading shows the ensemble spread between the minimum and maximum. The dashed vertical black indicates grounding line position.

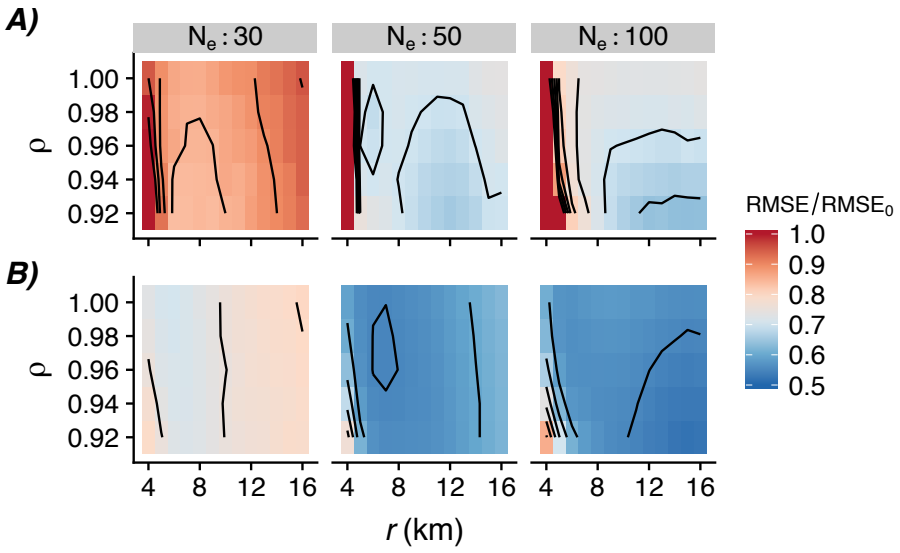

Figure 7. RMSE at $t = 20$a, relative to the initial (before assimilation) RMSE for the 50-member ensemble as a function of the forgetting factor ρ and the localisation radius r for different ensemble sizes N_e. A) for the bed and B) for the friction coefficient. Black lines as isovalues spaced by 5%.

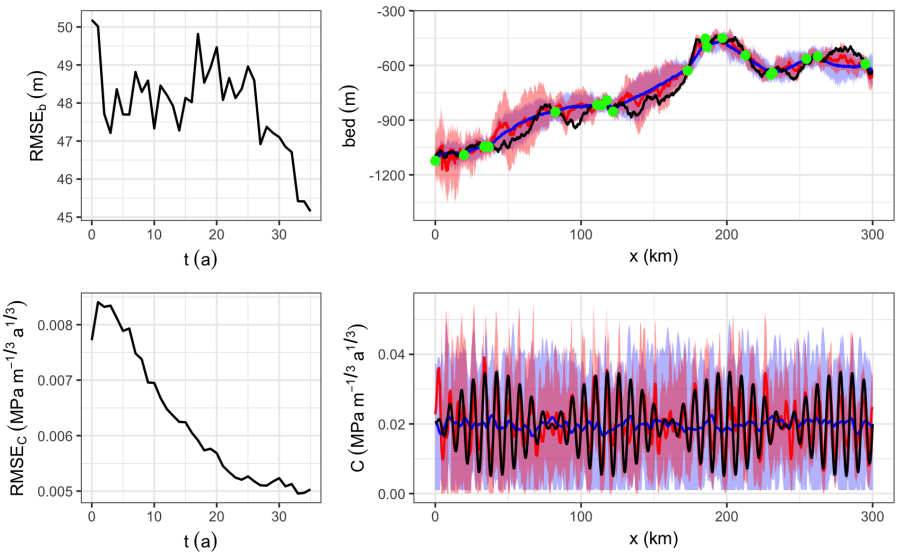

Figure 8. Same as Fig. 3 but with the RMSE computed for $x \in [0, 300]$ km.

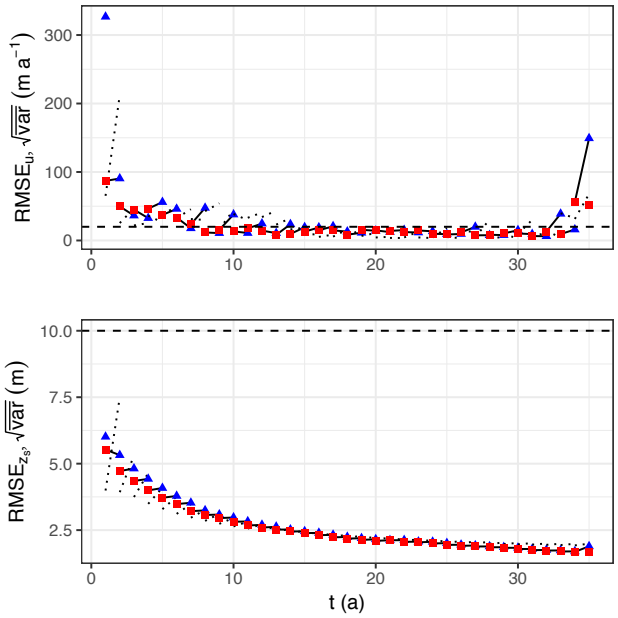

Figure 9. RMSE (solid lines) and square root of the averaged ensemble variance (dashed lines) during the assimilation window for (top) the velocity, u, and (bottom) the free surface, z_s. Each year, the blue triangle and the red square are the RMSEs before and after the analysis, respectively. Each segment represent a 1-year forecast step.

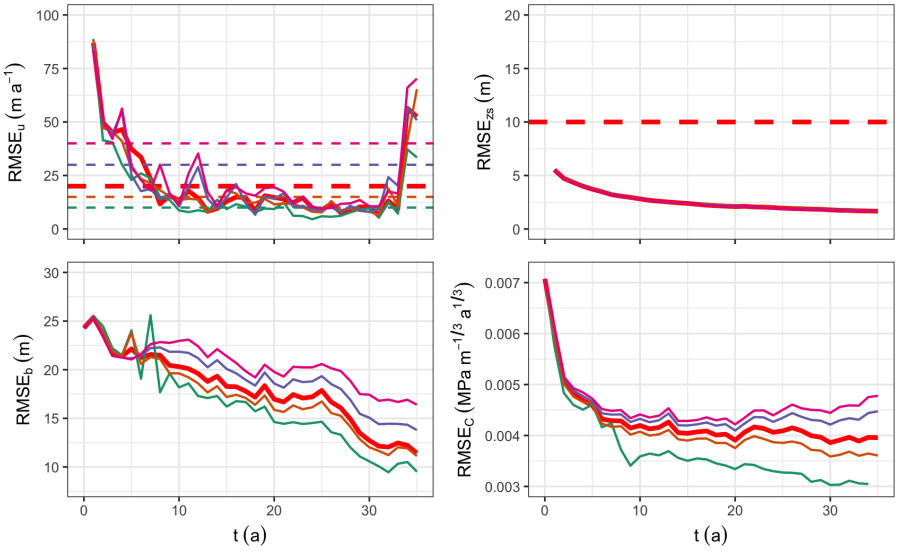

Figure 10. Sensitivity to the surface velocities observation error σ_v^{obs} : RMSEs after each analysis, computed only for $x \geq 300km$ for b and C.. The thick red lines correspond to the results with $\sigma_u^{obs} = 20$ ma^{-1} and $\sigma_{z_s^{obs}} = 10$ m shown in Figs. 3 and 9. The horizontal dashed lines correspond to the observation errors σ_u^{obs} and the results are presented with solid lines using the same color code. $\sigma_{z_s^{obs}} = 10$ m for all the experiments.

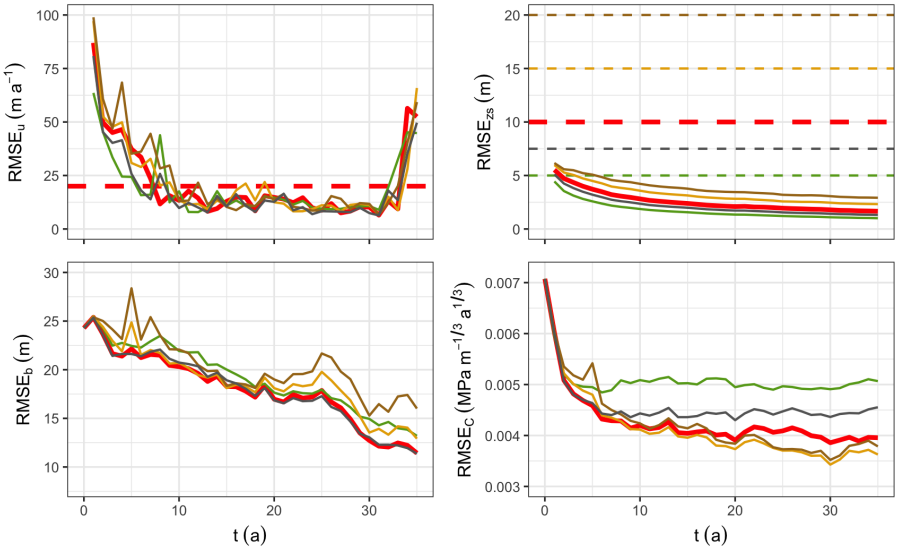

Figure 11. Sensitivity to the surface elevation observation error $\sigma_{z_s^{obs}}$: RMSEs after each analysis, computed only for $x \geq 300km$ for b and C. The thick red lines correspond to the results with $\sigma_u^{obs} = 20$ ma^{-1} and $\sigma_{z_s^{obs}} = 10$ m shown in Figs. 3 and 9. The horizontal dashed lines correspond to the observation errors $\sigma_{z_s^{obs}}$ and the results are presented with solid lines using the same color code. $\sigma_u^{obs} = 20$ ma^{-1} for all the experiments.

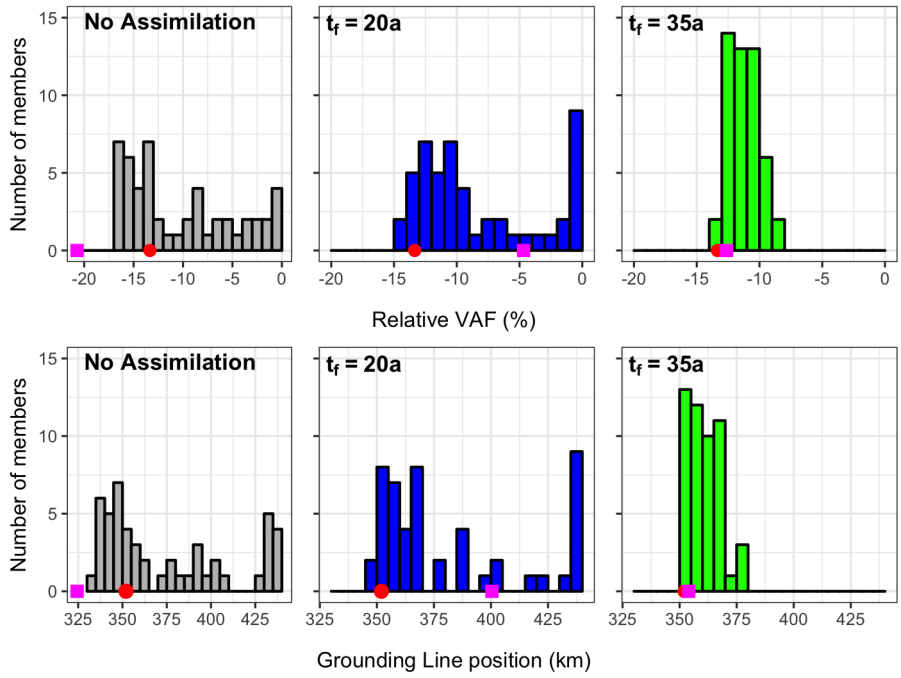

Figure 12. Ensemble forecast at $t = 100$ a: (top) relative change of volume above floatation (VAF) and (bottom) GL position with (left) no assimilation, (center) assimilation up to $t = 20$ a and (right) assimilation up to $t = 35$ a. The red circle correspond to the reference run and the magenta square to the deterministic forecast.