

The paper presents experiments on assimilation of observations of the flow of an ice sheet. A specific, and original, feature of the paper is that, following an earlier work of the authors (Bonan *et al.*, 2016), the flow model is built on a moving mesh that allows description of the motion of the front of the ice sheet while avoiding unpleasant spatial interpolations. The experiments are of the identical twin type, in which the ‘observations’ that are assimilated are extracted from an earlier integration of the assimilating model. The authors use two different assimilation algorithms, one that they call ‘3D-Var’ (see comment 12 below about the use of that expression), and a standard Ensemble Transform Kalman Filter (ETKF). Both algorithms are basically successful, with a general advantage for the latter.

The paper is well written and instructive, and I think its material is worth publishing, provided however a number of improvements are made. I give below my scientific comments and suggestions (in approximate order of decreasing importance), followed by editing remarks.

Scientific comments

1. There seems to be a basic contradiction in what the authors exactly do. It is said in subsection 2.2 (*Moving-point method*) that the evolution of the node position \hat{r} is determined by the condition that the corresponding mass fraction $\mu(\hat{r})$ is conserved in time (Eq. 9, later discretized in the form 11). It is then said in the following subsection that the discretized mass fractions μ_i are updated in the temporal integration (l. 20, p. 8, and Eq. 15).

The authors also write (eq. 26) that the state vector \mathbf{x} of their model consists of the ice thicknesses h_i at the points of the moving nodes, and of positions \hat{r}_i of those nodes. There should then be, in agreement with the general equation (16), prognostic equations for both the h_i 's and \hat{r}_i 's. There is one for \hat{r}_i 's, but none for the h_i 's. On the contrary, h is defined diagnostically by eq. (12) from the profile of the mass fraction.

I consider it is necessary for acceptance of the paper to first resolve that contradiction. I suspect that what is described in subsection 2.2 is actually not what is done in the numerical model. What is done must be precisely described. In particular, the prognostic equation for the thickness h (if there is any) must be explicitly mentioned. And, if the mass fractions are not conserved, by what is the evolution of the node positions \hat{r}_i determined?

2. The authors find that ETKF performs generally better than ‘3D-Var’ (one exception is actually shown on the left panel of Fig. 16; and, on Fig.13, the performance of both algorithms is the same at the end of the 10 years of assimilation). They give as an explanation of the better performance of ETKF the fact that the latter ‘remembers’ past observations (p. 17, l. 1, p. 24, l. 4). That really does not explain much. Actually ‘3D-Var’ also ‘remembers’ past observations as seen in many places in the paper (*e.g.*, right panel of Fig. 16) where it produces forecasts that are much more accurate than what is obtained when no observations are used.

It is possible to say more. The basic difference between ‘3D-Var’ and ETKF is that the former uses a background error covariance matrix \mathbf{B} that is defined from the start and remains static in time, while the latter computes a matrix \mathbf{B} from an ensemble of forecasts. Both algorithms evolve in time an estimate of the state of the observed system, but ETKF evolves in addition an estimate of the associated estimation error. That very likely leads to a better estimate of the matrix \mathbf{B} and to a better analysis. Actually, it is very generally observed that assimilation algorithms that carry in time an estimate of the estimation error (such as 4D-Var and the various forms of Kalman Filter) perform better than algorithms that use a static estimate of the estimation error (as does 3D-Var).

The importance of evolving in time an estimate of the estimation error must be stressed.

And I suggest the authors look in more detail at the consistency of the ‘predicted’ and ‘observed’ background. The authors state repeatedly that the reference lies in their ETKF experiments

within the range of the predicted ensemble. What about the ‘3D-Var’ ? It does not produce an explicit range, but it uses a known matrix \mathbf{B} to which the observed background error can be compared. And how does the matrix \mathbf{B} compare between the two algorithms ? Does it tend to be larger in either one of them? Such a comparison can fundamentally be made from Figs 3 and 7, but these figures are not in the same format (covariances in Fig. 3, correlations in Fig. 7).

3. An important question in assimilation is the degree of stability of the observed system. That question is always present, but is particularly obvious in sequential assimilation, such as 3D-Var and ETKF. In sequential assimilation, the evolution of analysed state \mathbf{x}^a results of the combined influence of, on the one hand, the unstable modes of the system and of the possible model errors, which together tend to increase the estimation error, and on the other hand, of the stable modes and the introduction of observations at analysis time, which together tend to decrease the estimation error (there are also neutral modes, which have at most marginal effect on the evolution of the uncertainty). Were there instabilities in the present case ? The fact that the forecast error tends to remain constant after the end of the assimilations (see, e. g., Fig. 16) suggests that there are no instabilities in the system. That is not surprising for the motion of highly viscous fluid. But assimilation, in the case of a system which has no instabilities used with an error-free model (which is the case here) is relatively easy. Any assimilation algorithm, unless it is devised or implemented in a particularly unfortunate way, will normally be ‘successful’ in that it will gradually move the analyzed state towards the reference from which observations are extracted.

This seems to be the case here. That does not decrease the value of the paper for the study of ice sheet modelling. But it should be mentioned that the problem considered in the paper is relatively easy from the point of view of the assimilation, in that all difficulties that might result from the presence of instabilities in the system and/or of model errors are absent.

4. A number of statements are made (for some as a way of explanation of a specific observed feature) which seem unjustified.

- 4.1. P. 12, ll. 5-6, ... *this approach ensures that this moving-point framework produces positive estimates of ice thickness variables and a smooth interior profile ...* What is the evidence, particularly as concerns the positiveness of thickness variables (I presume the question can arise only in the vicinity of the ice margin) ?

- 4.2. P. 13, ll. 15-16, *The formulation of \mathbf{B}_r aims to ensure that the order between mesh points defined by Eq. (13) is preserved by the 3D-Var algorithm. Since the distance between nodes evolves in time, it is even more important than in the previous case to use a flow-dependent background error covariance matrix \mathbf{B} .* In what does the particular formulation of \mathbf{B}_r helps preserving the order of mesh points ?

- 4.3. P. 16, ll. 17-18. ... *the experiment shows the sensitivity of 3D-Var to current observations because of the use of fixed variances in the prescribed covariance matrix \mathbf{B} .* I am not sure to understand what you mean. But if you imply that the increase of error at the second analysis time is due to the use of fixed variances, that is unfounded.

- 4.4. P. 20, l. 4, *This is because the ensemble spread is too small in that area.* The middle panel of Fig. 11 does not show a smaller ensemble spread beyond year 7.

- 4.5. P. 21, l. 10, and p. 22, l. 1, *The observation operator for surface velocities is nonlinear (...). For that reason, even if the ensemble is large, inflation is mandatory for the ETKF.* Is there any objective evidence for a link between the nonlinearity of the observation operator and the need for

inflation of the analyzed ensembles ? I suggest you only mention that you have observed that, even though the ensembles are large, inflation is necessary in the present case.

5. The authors write in the conclusion (p. 24, ll. 1-3) *3D-Var can have issues with assimilating observations if they are located outside the forecast domain; the ETKF can overcome these issues if at least one member of the ensemble has its numerical domain large enough to include the location of these observations*. The problem raised by observations that are located outside the background domain is hardly discussed in the main text (from p. 20, l. 19 to p. 21, l. 1). If they consider that aspect to be important enough to be mentioned in the conclusion, the authors must discuss it at more length in the main text, and not wait the presentation of the ‘advanced configuration’ experiments to discuss it.

I first mention that what I understand of the first line of p. 21 would be better expressed as ... *since it is observed that there is always at least one member of the ensemble whose domain is larger than the domain of the background* (I think that, contrary to what the authors write, it is the background that matters here, not the reference).

Second, what is done in 3D-Var with those outlying observations ? Are they simply ignored ?

And, in ETKF, even if one or more ensemble members can accommodate those observations, but not all, how is the ensemble increased back after the analysis to its full dimension N_e ?

6. Eq. (34-35). I understand that equation originates from Eq. (4). Say it. And give explanations (or at least a reference) for the complicated expression on the right-hand side.

7. Subsection 5.3. The errors on the observations of surface velocity and of position of the margin are apparently not mentioned.

8. P. 23, caption of Fig. 16, last two lines. You write *The ETKF performs better than the 3D-Var with respect to the position of the margin, but 3D-Var seems to give better results for the ice thickness at $r = 0$* . Do you imply the better performance of ETKF on the left panel of the figure is real, but the better performance of 3D-Var on the right panel might be only apparent ?

9. Figures 4 and 5. The left panel shows results at the first analysis time $t_1 = 500$ yr. It would be preferable to show also the analogous results at the second time $t_2 = 1500$ yr.

10. Since the surface altitude $s(t, r)$ is known, there is no need for making a distinction between ice thickness and surface elevation (Eqs 30 and 31). These observations are exactly equivalent.

Editing comments

11. Literally speaking, the expression *3D-Var* is inappropriate for the paper. The physical problem under consideration is one-, not three-dimensional, and the authors mention nowhere that they have used an explicit variational algorithm for determining the analysed state \mathbf{x}^a . Now, their algorithm possesses a property which, owing to a pure language convention in the trade of assimilation, is associated with the expression *3D-Var*. Namely, that it uses a time-invariant background error covariance matrix \mathbf{B} (that convention originates from ECMWF, which developed a *3D-Var* as a preliminary step towards its full-fledged *4D-Var*). I suggest that the authors, if they want to keep the expression *3D-Var*, say clearly that their algorithm is neither 3D nor variational, and use a different notation (*‘3D-Var’*, *3D-Var-like*, ...)

12. The right-hand side of Eq. (21) should read $\mathbf{B}_k \mathbf{H}_k^T (\dots)$ (and not $\mathbf{B}_k \mathbf{H}_k$)

13. P. 18, 1.10. You mention T_{clim} . Reference should be made here to Eq. (A4).

14. Eq. (A3). I think the unit for T is kelvin. Say it.