I first apologize to the authors for the long delay in sending this review.

The additional results that the authors have obtained have significantly improved the paper, and lead to a much more precise evaluation of the compared performance of DBFN and 4DVar. I have first two comments on these additional results.

1. The new Figure 13 shows that changing the initial conditions of the experiments improves he results in three cases out of four (and does not modify them in the fourth case). It would be useful to explain briefly how the initial conditions have been perturbed (just by random perturbations ?).

2. The authors (ll. 601-603) explain the differences in the results produced by DBFN and 4DVar by the fact that the residual error at initial time must be concentrated by the DBFN in the stable components of the flow. That fact is certainly true, but it is true as well of 4DVar, and for the same reasons (this has been discussed in detail by Pires *et al.*, 1996; see also comment 9 below). So it cannot explain (at least in the absence of additional information) differences between the results produced by DBFN and 4DVar.

And I have a number of additional comments which, as said in my previous review, I had not mentioned then since they were in my mind of lesser importance.

3. The authors do not really define what they mean by 'nudging'. The only definition they give is the very general equation (1). Kalman Filter is actually of form (1), although it is clear that it is not 'nudging' for the authors. Obviously, nudging requires at least the gain matrix $\boldsymbol{K}$ in eq. (1) to be 'simple', but that is only implicit, and nowhere discussed clearly. Actually, there is a continuum of methods which, starting from a diagonal gain matrix, go to a full Kalman matrix which represents the whole spatial structure of the uncertainty on $\boldsymbol{x}$. It is certainly not obvious where the limit is to be put.

It does not make sense in my mind to try and distinguish, as the authors seem to do in the Introduction, between methods that use nudging and methods that do not. Unless they can give a precise definition of nudging, I suggest they remove that aspect of the discussion.

As for the statement (l. 22) *The first appearance of nudging in the geophysical literature was in 1974*, it may be true of the word, but not of the thing. The idea of, well, 'nudging' the fields produced by a numerical model towards the observations, with the expectation that this would lead those fields closer to reality, was already known and studied by 1974. A basic paper in that respect is Charney *et al.* (1969). I also mention Morel *et al.* (1971), who did use a form of forward-backward iterative adjustment of a model to observations.

4. Ll. 124-125 … *forces the state variables to fit the observations as well as possible*. This cannot be what you mean. In order to fit the observations as well as possible, you would have to define $\boldsymbol{x}$ in such a way that $\mathcal{H}(\boldsymbol{x}) = \boldsymbol{x}_{obs}$. This is obviously not what you want to achieve, even in the limit of infinitely many iterations of the forward-backward nudging (on the other hand, you certainly hope that the state variables will fit reality as well as possible).

5. Ll. 138-139. *The BFN novelty with respect to conventional nudging methods is the model integration backward in time*. Backward integration for assimilation existed before BFN, even in a context of simple 'nudging'. See the quoted paper by Morel *et al.* (1971) or Talagrand (1981) (and the authors mention a paper by Kalnay *et al.*, 2000, which apparently also deals with assimilation).

6. LL. 212-229. What is said here about the incremental method introduced by Courtier *et al*. is erroneous. The incremental method was not introduced for dealing with the problem of possible multiple minima of the cost function (9). It was introduced for reducing the overall cost (code development and computing time) of 4D-Var, by solving the global nonlinear minimization of (9) through a sequence of more economical linear minimizations of form (13). The need for inner loops does not result from increments becoming too large. Each new linearization of $\mathcal{M}$ and $\mathcal{H}$ is performed about the nonlinear solution of the model (2) emanating from the result of the latest linear minimization.

And, incidentally, the question of possible multiple minima of (9) has been studied in the quoted paper by Pires *et al*. (1996).

7. Sentence starting l. 236, *The model solves six prognostic equations, namely* […] *the hydrostatic equilibrium,* …. The 'hydrostatic equilibrium' is not an equation, and is not expressed by a prognostic equation anyway. I do not know the NEMO model, but I think it actually solves five prognostic equations, for the variables that are mentioned ll. 448-449. Please check the whole paper on this point.

8. Ll. 511-512, *Iterations compensate for the lack of a priori information on the model errors* …. Taken literally, iterations, which only crunch numbers, cannot compensate for information that is not there in the first place. Do you mean that a (possibly implicit) hypothesis is made somewhere about the model error ?

9. Ll. 538-541. *The poor performance of the 4Dvar* […] *is related to spurious increments due to the fact that in one assimilation window there is only one set of observation available. If this set is at the end of the window this can complicate the minimization process and the iterations may stop before convergence.* All that seems to be very speculative, and is not an agreement with the immense experience that has now been accumulated on 4Dvar. Actually, inverting a model (*i.e.*, reconstructing a model state at time t from exact observations at a later time $t+T$) is a standard way of checking that a code for 4Dvar does what it is meant to do. But if observations are concentrated at the final time of the assimilation window, one can expect that it will be difficult to accurately reconstruct at initial time the components of the fields along the stable modes of the flow. That is obvious from the fact that an initial perturbation on the stable modes will have very little impact on the fields at the end of the assimilation window (see comment 2 above). On the other hand, the fields will of course be perfectly reconstructed at the end of the window. If the authors know that it is what they observe, then they must say it clearly (but then, the error at the end of the assimilation window must be small). If they do not observe that, I think it is preferable not to propose an explanation for the performance of 4Dvar.

10. Figure 5 does not provide a fair comparison between 4Dvar and DBFN. A proper comparison would be between the computing times necessary to reach a given level of accuracy. Please remove completely the figure, or replace it with a more instructive one.

11. Figure 6 and associated text. I suggest to replace the words *mean trajectory* with *temporal evolution (over the assimilation window)* (at least, that is what I understand).

12. L. 466. The *Ordinary Nudging (ONDG)* mentioned here does not seem to have been defined.

13. L. 28, … *in the recent work of Auroux and Blum (2005)*. Well, not so recent …

REFERENCES

Charney, J., M. Halem, and R. Jastrow, 1969, Use of Incomplete Historical Data to Infer the Present State of the Atmosphere. *J. Atmos. Sci.*, **26**, 1160–1163.

Morel, P., G. Lefevre and G. Rabreau, 1971, On Initialisation and Non-Synoptic Data Assimilation, *Tellus*, **23**, 197-206.

Pires, C., R. Vautard and O. Talagrand, 1996, On extending the limits of variational assimilation in nonlinear chaotic systems, *Tellus*, **48A**, 96-121.

Talagrand, O., 1981, A study of the dynamics of four-dimensional data assimilation, *Tellus*, **33**, 43-60.