Response to Reviewer#4 of “Nonlinear effects in 4D-Var” by Massimo Bonavita, Peter Lean and Elias Holm.

Anonymous Referee #4

Received and published: 16 July 2018

We thank the Reviewer for her/his careful reading of our manuscript and the constructive suggestions, which we will mostly incorporate into the final version of our manuscript. We also provide below detailed answers to her/his comments.

1) *p1L15: Rabier and Courtier (QJRMS 1992) presented a good study to measure the accuracy of the tangent-linear with a "realistic" global baroclinic model, the IFS of the time but without the physical parameterizations. Lacarra and Talagrand (1988) did also conclude that the TL model was a reasonable approximation of the evolution of a perturbation for ~48-h. This is the basis of ECMWF's EPS.P4, line 24: “Holm et al., 2003 should be Holm et al. 2002. “.*

   Yes, this is an important early work on the validity of the TL approximation for a baroclinic wave evolution, it will be added in the references. Also, thanks for spotting the error in the Holm reference, it will be corrected in the final version.

2) *the reference to Pires et al. (1996) is missing in the list of references.*

   It is present, PG 13 L 27.

3) *p1L24: "the use of a linear model".*

   OK

4) *p2L22: although this is mentioned in the legend, the notation TL511, for example, should be explained: eg., 40 km (TL511, spectral truncation 511 with a linear grid). The same for TCo1279, which was a new one for me.*

   It is a good suggestion, this will be made clearer in the final version. Also, a relevant reference will be added for the reader interested in more details.

5) *p4 eq.(5): the Hessian should be written as* $\mathbf{B}^{-1} + \sum_k \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k$ *. The perturbation* $\delta \mathbf{x}_0$ *is not part of it.*

   Thanks for spotting this error. What we meant to write here is that the Hessian is a function of the reference trajectory in the nonlinear case. This will be rectified in the final version.

6) *p3L23: the notation* x^t *is usually used for the true state in the literature while here it refers to the trajectory. I suggest that the notation* x^{tr} *should be used instead.*

   We agree, x^t could be confusing. We will denote the guess trajectory as x^g in the final version of the manuscript.

7) *p5L4: I can understand that interpolation of a high resolution field with high variability would create a "noisy" low resolution field. However, such cases require special treatment of the interpolation to be representative (e.g, aggregation instead of geometrical interpolation). Can the authors comment on this?.*

   This effect has become more visible in recent years due to the increased difference in resolution between inner and outer loops. Work is under way to find a satisfactory solution in term of appropriate time and space interpolation/averaging from the high resolution to the low resolution trajectory.

8) *Rabier and Courtier (QJ 1992) have looked into this with a similar approach. This should be referred to and discussed in relation with the results of Fig.1. The difference*

$$\Delta\mathbf{x} = M(\mathbf{x}^{n-1} + \delta\mathbf{x}^n) - M(\mathbf{x}^{n-1} - \delta\mathbf{x}^n)$$

shows how long the linearity assumption holds regardless of how the TLM has been formulated. If $\Delta\mathbf{x}$ is not small, it implies that it is hopeless to be able to find a linear model that would provide a reasonable evolution of the perturbation in the initial conditions.

Yes, this paper will be referred to in relation with the results of Fig. 1.
The additional diagnostic quantity mentioned by the Referee is useful to evaluate the assimilation window length over which the linearity assumption holds regardless of the accuracy of the TL model. From our perspective in this work, the stronger constraint implied by Eq. (6) in the paper is the relevant quantity to look at to evaluate the goodness of the TL approximation for a fixed assimilation window length and the available TL model.

9) p5L15: reusing the observations implies that the accuracy of the background state has changed and the B-matrix should reflect that. The first-guess trajectory is therefore not the background state or is it? My interpretation is that the first-guess is just a starting point of the minimization as are each state used in an outer loop.

The interpretation of the Referee is correct. The guess trajectory is just a reference state around which the generalised observation operator **G** is linearised. Thus, no change in the B-matrix is required.

10) p5L23: this argument holds if the errors are Gaussian but non-Gaussian errors could lead to a non-zero mean ensemble.

True. This will be noted in the final version of the paper.

11) *p7L5: convergence needs to consider that the objective function represents the fit of a given realisation of the observations. Convergence to numerical accuracy is meaningless and it is justified to reduce the requested accuracy of the minimization.*

A similar point has been raised in the paper's discussion by an earlier Referee and has been addressed there (and will also be in the final version of the paper).

12) *p8, Fig10: it is difficult to say that the departures with respect to wind observations is smaller when more outer loops are used since the scales (color bars) are not the same. In fact, that of the first-guess has a value of 9.97, while for one outer iteration it is 11.91, with three we get 13.52 and finally with 5, 11.91. It is not so "visually" apparent that 5 outer loops is better.*

The scales used are the same except for the last interval, which is different in the four panels in order to accommodate the different maximum values. Apart from the visual impression, the increased analysis fit with increased number of outer loops is demonstrated by the smaller area-averaged values of the standard deviation of the analysis departures (as reported in the text and in the top captions of each panel).

13) *p8L15: this particular situation involves the physical parameterizations (convection). Looking at the initial physical tendencies (see Rodwell and Palmer, QJ2007) may reveal interesting information for this particular case.*

This is a good suggestion, thanks. We will pursue this idea in future work.

14) *p9L10 (Fig.11): a reduction of O-A is not a good measure of the quality of the analysis. This can be obtained by reducing the observation error but result in even unphysical forecasts. The O-B is a better measure in that sense and indicates that increasing the outer loop from 3 to 4 or 5 leads to a more modest gain.*

We beg to differ on this point. In the ECMWF 4D-Var, the O-A departures are computed through a nonlinear model integration started by the analysis fields at the beginning of the assimilation window. Thus, they represent the best fit of a full model trajectory to the observations over the whole assimilation window. Improving this fit without changing the input error statistics implies a better analysis. A different, but closely related question, is how much of this signal can the full model propagate in time to the next assimilation window. The O-B statistics give a measure of this second aspect. The fact that O-A and O-B statistics show changes which are qualitatively similar but quantitatively different are seen as confirmation that the analysis is behaving properly.

15) *With one outer loop, is the resolution of the analysis increment TL95 or does it correspond to that used in the third outer loop (TL399). If it is TL95, the degradation may be attributed to the degraded resolution and in that case, it would be better to redo it using the same higher resolution as used for the 3, 4 and 5 outer loops. Even in 3D-Var, TL95 would be considered too low.*

    This point has been raised in the paper's discussion by an earlier Referee and has been addressed there (and will also be in the final version of the paper). In short, the single resolution experiments have been run at the maximum resolution of the multi outer loop experiments (in this case TL255), so the relatively poor performance of the single resolution experiments cannot be attributed to this.

16) *p9L25: with significantly different observation errors, the minimization would focus first on those with small errors and it is only when convergence is reached for those that it would take care of others. This can happen when artificially large observation error are assigned to some satellite observations which were then incapable to have a significant influence on the analysis. In this particular case, it may be that satellite observations have now impacted the analysis significantly more than before.*

    We have conducted further experiments on this aspect and we will shortly report on their results in the final version of the paper. In summary, the problematic convergence of the stratospheric-peaking channels is due to the difference in timestep between inner and outer loops, which causes different wave propagation speed in the outer loop trajectories and in the inner loop minimisations. Running with same timestep in both inner and outer loop solves the problem.

17) *p10L10: an experiment assimilating only satellite measurements sensitive to humidity and precipitation could show more about the nonlinearity associated with those.*

    Yes, this is a useful suggestion.

18) *p10L14: "seen that the impact".*

    Thanks for spotting this typo.

19) *Conclusion: increasing the resolution implies a reduction of the nonlinear timescale. The assimilation window would have to be shortened. It is important to evaluate what this timescale is. If the TLM of the full model cannot be achieved, Tanguay et al. (1995) have shown with a simple model that even in the best of cases convergence cannot be reached. To what extent can we say that a weak constraint 4D-Var would be needed? Being at ECMWF hwere the weak constraint 4D-Var has been extensively studied, the authors are in a position to comment on this.*

    Generalised weak-constraint 4D-Var with a time-varying model error term would be a solution to the increased nonlinear effects in 4D-Var, as the Reviewer suggests. However, it comes with its own problems. A fundamental issue is the evaluation of accurate first and second moments of the model error: it is not clear how to do this, at this stage. Another, connected problem, is that this type of evolving model errors will be auto-correlated in time and cross-correlated with the estimated background errors. Thus, significantly more complex forms of the 4D-Var cost function would have to be implemented.

From our perspective, and as stated in our conclusions, the more promising path towards controlling nonlinearity in 4D-Var is through repeated re-linearisations in a DA framework where the length of the assimilation window is progressively increased. This is not a new idea (e.g., Pires et al., 1996; Jarvinen et al., 1996), but we think it is an idea whose time has come to be put into operational practice.