Linearized physics: progress and issues

Philippe Lopez

ECMWF
Reading, United Kingdom
philippe.lopez@ecmwf.int

Abstract

Linearized physics has become an essential component of four-dimensional variational (4D-Var) data assimilation, in particular to extract information from direct or indirect cloud and precipitation observations. Tangent-linear and adjoint versions of an increasing number of physical shemes are currently used during the minimization of the 4D-Var cost function to evolve the model state to the actual observation time and to convert it into an observed equivalent. However, necessary discrepancies between linearized physics packages and their fully-fledged non-linear counterparts, resolution differences between minimizations and trajectories as well as strong non-linearities in the forecast model can affect the optimality of the 4D-Var analysis, in particular when assimilating observations that are affected by clouds or precipitation. The application of some smoothing in time to either observations or to the non-linear trajectory can help to alleviate these problems. In addition, the "0-value" issue currently limits the coverage of cloud and precipitation observations that can be successfully assimilated as well as our ability to move weather systems that are misplaced in the model background. Current assimilation in saturated regions also exhibits an asymmetry in analysis departures, which indicates that it is always easier to reduce than to increase cloud/precipitation. Recommendations for future work are given.

1. Introduction

In the context of four-dimensional variational (4D-Var) data assimilation (DA) (e.g. Rabier *et al.* 2000), the optimal model 3D representation of the atmospheric state (the *analysis*) is obtained by combining information coming from a set of observations available over a certain time window (typically up to 12 hours) with a priori information coming from the forecast model (the *background*, usually a previous short-range forecast). To achieve this goal, the following cost function

$$J = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=1}^{nsteps}\left(\mathbf{y}_{0_i} - HM_i(\mathbf{x}_0)\right)^T \mathbf{R}_i^{-1}\left(\mathbf{y}_{0_i} - HM_i(\mathbf{x}_0)\right) \qquad (1)$$

is minimized in a least square sense, where $\mathbf{x}_0$ denotes the model state vector at the beginning of the assimilation window and $\mathbf{y}_o$ the vector of available observations. Superscript b indicates the background model state. The model state typically consists of temperature, moisture, wind and surface pressure. **B** and **R** are the error covariance matrices for model background and observations, respectively. H and M correspond to the observation operator and to the (non-linear) forecast model, respectively. Operator M is used to evolve the model state from the beginning of the assimilation window to the actual time of each observation (time subscript i). Operator H permits the conversion of the model state to observation space, for instance through a radiative transfer model. The summation in Eq.(1) is performed over all time steps inside the assimilation window.

The minimization of the cost function, J, requires the computation of its gradient with respect to the initial model state, $\mathbf{x}_0$,

$$\nabla_{x_0} J = \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \sum_{i=1}^{nsteps} \mathbf{M}_i^T\, \mathbf{H}^T\, \mathbf{R}_i^{-1}\left(\mathbf{y}_{0_i} - HM_i(\mathbf{x}_0)\right) \qquad (2)$$

where **M** and **H** are the tangent-linear versions (i.e. the matrices of derivatives w.r.t. the model state) of the forecast model and observation operator, respectively. The transpose of these matrices which appears in Eq.(2) corresponds to their adjoint versions. Since the forecast model, M, contains physical parameterizations (e.g. radiation, vertical diffusion, gravity wave drag, convection and large-scale condensation), past studies have demonstrated that it can be beneficial to include a representation of these processes in the linearized physics (LP) used in the minimization. However, the use of LP in 4D-Var requires to achieve the best compromise between

- linearity (one of the most constraining assumptions of 4D-Var),
- simplicity (to reduce computational cost and avoid coding nightmares), and
- realism (w.r.t. the true atmospheric behaviour but also w.r.t. the non-linear forecast model).

Reaching this goal can be extremely challenging and time consuming. In particular, it implies that LP paramaterizations are usually simplified versions of their fully-fledged non-linear counterparts. Special care has also to be taken in order to avoid the inclusion of processes that will lead to the spurious growth of initial perturbations in the tangent-linear model.

2. Status of LP

Table 1 provides a non-exhaustive list of LP parameterizations currently used in operational global DA systems around the world. The type of DA control variable(s) used to describe moist processes is also mentioned. Many centres (except JMA and ECMWF) only have a partial description of physical processes in their LP packages. In particular, the inclusion of convection is clearly the biggest challenge. In other respects, the large variety of moist control variable formulations reflects the lack of consensus among the DA community on which variable analysis increments should be applied to.

3. Minimization/trajectory agreement

Sub-optimality can appear during a 4D-Var assimilation cycle as a result of

- Simplifications in the LP compared to the fully-fledged non-linear model,
- Strong non-linearities in the forward model, in particular in the presence of moist processes which are often associated with thresholds (saturation, autoconversion), switches or discontinuities.
- Horizontal resolution differences between minimizations and trajectories. Minimizations are run at a lower resolution (e.g. T159≈130km and T255≈80km at ECMWF) than trajectories (e.g. T1279≈15km at ECMWF), mainly to reduce computational cost.

Status June 2010	Environ. Canada	ECMWF	JMA	Météo-France	Met Office	NCEP	NRL
DA method	4D-Var	4D-Var	4D-Var	4D-Var	4D-Var	3D-Var	4D-Var
RAD		X	X			NR	
VDIF	X	X	X	X	X	NR	X
GWD	X	X	X	X		NR	X
CONV		X	X		X	NR	
LS COND	X	X	X	X	X	NR	
DA moist Cont. Var.	$\ln(q_v)$	$\frac{\delta RH}{\sigma(RH_b)}$	$\ln(q_v)$	$\frac{\delta q_v}{\sigma\left(q_v^b\right)}$	$q_t=q_v+q_c$	q_v/q_{sat}	$q_v/q_{sat}(T_b)$

Table 1: Summary of linearized physics parameterizations and data assimilation moist control variable used in various global assimilation systems around the world. RAD=radiation, VDIF=vertical diffusion, GWD=gravity wave drag, CONV=convection, LS COND=large-scale condensation, q_v=specific humidity, q_{sat}=saturation specific humidity, q_c=cloud condensate amount, RH=relative humidity. Subscript or superscript b denote background values. NR (not relevant) indicates that LP is not required in 3D-Var.

An illustration of the mismatch between minimization departures, $D_{min}=\mathbf{y}_o-(HM(\mathbf{x}_0^b)+\mathbf{HM}\delta\mathbf{x}_0)$, and subsequent trajectory departures, $D_{traj}=\mathbf{y}_o-HM(\mathbf{x}_0^b+\delta\mathbf{x}_0)$ is given in Fig.1, for the first T95 minimization and updated T799 trajectory of a single 4D-Var assimilation cycle with the ECMWF system. D_{min} is computed using low resolution increments, $\delta\mathbf{x}_0$, and simplified LP (operator $\mathbf{M}$) while D_{traj} uses the increments evolved at high resolution with the full non-linear forecast model (M). This Taylor diagram displays for each observation type assimilated in ECMWF's 4D-Var (symbols; see legend at the top) the standard deviation ratio (SDR; radial direction) and correlation (CORR; azimuthal direction) of D_{traj} and D_{min}. A perfect match between the two quantities is obtained when SDR=1 and CORR=1 (black square). The further away from this black square, the poorer the agreement between minimization and trajectory and the stronger the sub-optimality of the 4D-Var analysis. Fig.1 shows that for observations that are insensitive to clouds and precipitation the level of agreement between D_{min} and D_{traj} is rather good, with SDR between 1 and 1.5 and CORR ranging from 0.7 to 0.92 for radiosonde temperature measurements and AMSU-B brightness temperatures (TBs). For observations that are directly affected by clouds or precipitation, such as SSM/I and AMSR-E all-sky microwave TBs, non-linearities and interpolation errors lead to a degradation of the match since SDR increases up to 1.7 while CORR drops down to 0.46. When NCEP Stage IV combined radar and gauge hourly precipitation data (NCEP-RR; Lin and Mitchell 2005) are assimilated, the results become dramatically poor, with SDR exceeding 4.5 and CORR around 0.2, which could lead to very bad 4D-Var performance. However, if NCEP Stage IV observations are accumulated over 3, 6 or 12 hours (NCEP-RR[3,6,12]h) prior to their assimilation, the level of matching between D_{min} and D_{traj} makes a spectacular recovery, with SDR and CORR reaching similar values to those for SSM/I and AMSR-E TBs. The optimal accumulation period seems to be 6 hours, which is the value chosen in the latest direct 4D-Var assimilation experiments with NCEP Stage IV rain data run at ECMWF (Lopez 2010). These results suggest that smoothing cloud or precipitation observations in time can reduce the risk of sub-optimality in 4D-Var.

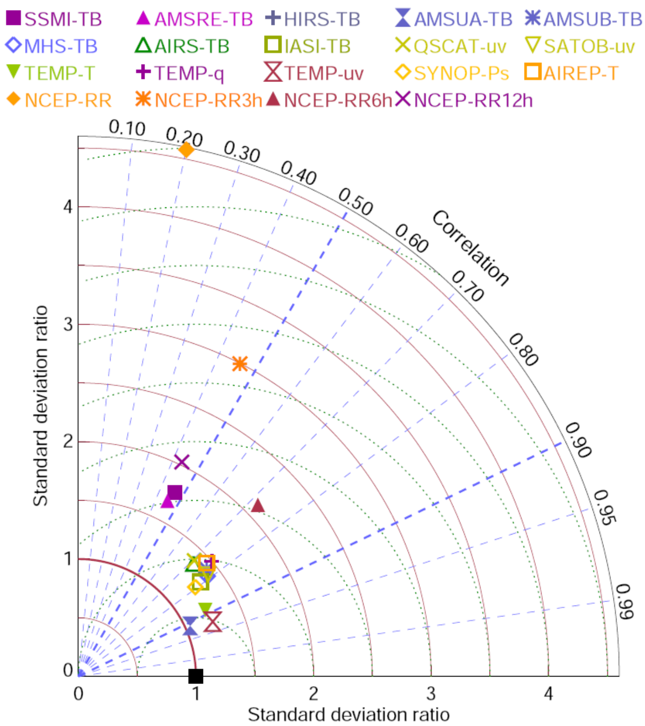

Figure 1: Taylor diagram showing the level of matching between minimization and trajectory departures for various observation types (see legend at the top) from a single T799 L91 4D-Var assimilation cycle. A perfect match is indicated by the black square (see text for details).

4. Excessive non-linearities

It is common practice in the LP community to check that the linearity hypothesis is valid for initial perturbations, $\delta\mathbf{x}_0$, of all sizes, at the range of horizontal resolutions and over the time window length used in 4D-Var minimizations. The typical test verifies the relationship

$$M(\mathbf{x}_0 + \delta\mathbf{x}_0) - M(\mathbf{x}_0) \approx \mathbf{M}\delta\mathbf{x}_0 \quad (3)$$

where the notations of Eq.(2) are used. The left hand side of Eq.(3) corresponds to the difference between two integrations of the non-linear model, M, started from slightly different intial conditions (perturbation $\delta\mathbf{x}_0$). The right hand side is the evolution of the initial perturbation with the tangent-linear model, $\mathbf{M}$.

Experimentation to assess the validity of the linearity assumption when initializing the model with small-amplitude global perturbations recently evidenced some extreme growth of non-linearities in the non-linear model, which the tangent-linear model will never be able to represent. As an example, Fig.2 shows the time evolution of an initial T95 L60 white noise temperature and wind perturbation with maximum amplitude of 10^{-5} K and 10^{-5} m s^{-1}, respectively (not shown). The evolved field displayed in Fig.2 is temperature at the lowest model level.

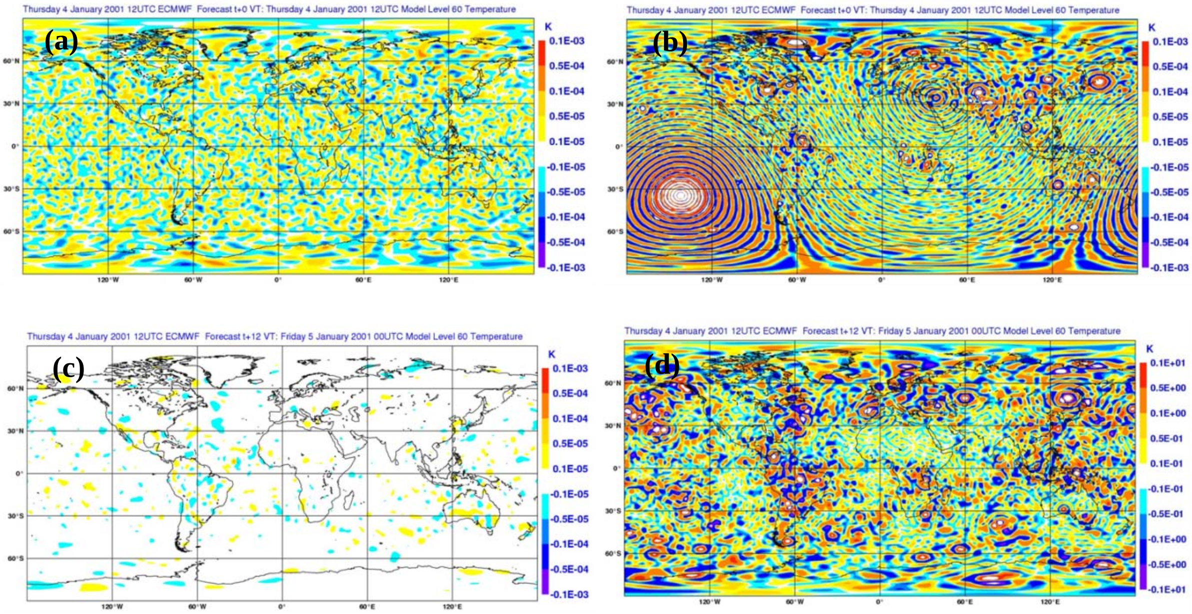

Figure 2: Comparison of $\mathbf{M}\delta\mathbf{x}_0$ (left panels) and $M(\mathbf{x}_0+\delta\mathbf{x}_0)–M(\mathbf{x}_0)$ (right panels) after 30 mn (top panels) and 12 hours (bottom panels) using an initial T95 L60 white noise perturbation with amplitude 10^{-5} K and 10^{-5} m s^{-1}. Displayed field is temperature at the lowest model level. Note the change of scale in panel (d).

Note that all physical parameterizations were activated in both the non-linear and the tangent-linear models. Panels (a) and (b) display $\mathbf{M}\delta\mathbf{x}_0$ and $M(\mathbf{x}_0+\delta\mathbf{x}_0)–M(\mathbf{x}_0)$ after a single time step (30 mn), respectively, while panels (c) and (d) are valid for the 24th time step (12h). While $\mathbf{M}\delta\mathbf{x}_0$ does not show any sign of growth (the perturbation amplitude actually decreases in time), $M(\mathbf{x}_0+\delta\mathbf{x}_0)–M(\mathbf{x}_0)$ exhibits a very strong amplification of the initial perturbation after 24 steps (panel (d)), leading to a temperature perturbation in the order of 50 K over Eastern Siberia (note the change of scale in panel(d)). Panel (b) shows that even after a single time step, the perturbation has already amplified by several orders of magnitude over the South Pacific. It should be emphasized that such excessive growth does not occur with larger amplitude global initial perturbations, as is currently the case in 4D-Var. It is believed that thresholds and switches present in the non-linear versions of the vertical diffusion, large-scale condensation and convection schemes are responsible for the extreme non-linear behaviour seen in Fig.2. A possible way to alleviate this problem could be to apply some smoothing to some of the physical outputs in the non-linear trajectory (e.g. Stiller 2009: smoothing of convective tendencies in Met Office 4D-Var).

5. The "0-value" issue

When cloud or precipitation retrievals, radar reflectivity or lidar backscatter are to be assimilated, a problem arises whenever the observed equivalent computed from the model background is equal to zero, while the observation is non-zero. Indeed, in this case, the sensitivity of the moist physics outputs (cloud water or precipitation) to the input model variables (typically temperature, moisture, wind and surface pressure) is zero, which means that no information can be extracted from the observation (the analysis is equal to the background). A possible way to overcome this problem could be to define a first-guess which is artificially moistened compared to the background (instead of using the background as first-guess). Such approach was recently tested by Caumont et al. (2010) in their 1D+3D-Var mesoscale assimilation of 3D radar reflectivities. However, it is not clear whether such modification would be recommended and easily implementable in 4D-Var.

Conversely, when the observation is zero and the observed equivalent from the model background is non-zero, the sensitivities of the moist physics to the input model variables are non-zero and the analyzed state can get closer to the observation than the background (in observation space). However, in this case, a strong ambiguity remains regarding the actual state of the atmosphere, since a non-cloudy or non-rainy situation can be associated with a large variety of atmospheric profiles. For instance, a non-rainy observation can equally be obtained in extremely dry or nearly saturated conditions. Only the availability of other observation types (e.g. radiosoundings) in the vicinity can alleviate this ambiguity. One should also keep in mind that a misleading 0-observation from a given instrument can be obtained if the detection threshold of this instrument is not exceeded.

This is the reason why in ECMWF's current 4D-Var assimilation experimentation with NCEP Stage IV precipitation data over the USA (Lopez 2010), only points that are rainy in both model background and observations are assimilated, which limits observation coverage and our ability to correct possible misplacements of weather systems in the analysis. However, it should be underlined that the assimilation of 6-hourly accumulated precipitation amounts (instead of hourly accumulations) lessens the effect of the "0-value" issue because of the propagation of weather systems. It should also be noted that the "0-value" issue does not affect the assimilation of microwave TBs since these are sensitive not only to clouds and precipitation, but also to water vapour.

6. Asymmetry of analysis departures

In the previously mentioned direct 4D-Var assimilation experiments with NCEP Stage IV precipitation data, an asymmetry could be identified when plotting the joint PDF of analysis versus background departures in precipitation space, as illustrated in Fig. 3. It appears to be always easier to reduce than to increase cloud or precipitation in the model. This asymmetry, which was already evidenced in previous studies of the assimilation in cloudy and rainy regions, is partly attributable to the capping effect of the saturation threshold on relative humidity increments in the analysis (Hólm *et al.* 2002). An additional explanation is the existence of asymmetries in the moist physics sensitivities to input model temperature and moisture. Introducing a prognostic variable for cloud condensate in the DA control vector (i.e. the ability to produce increments of cloud condensate directly) might help to reduce this problem. This is currently under development at ECMWF.

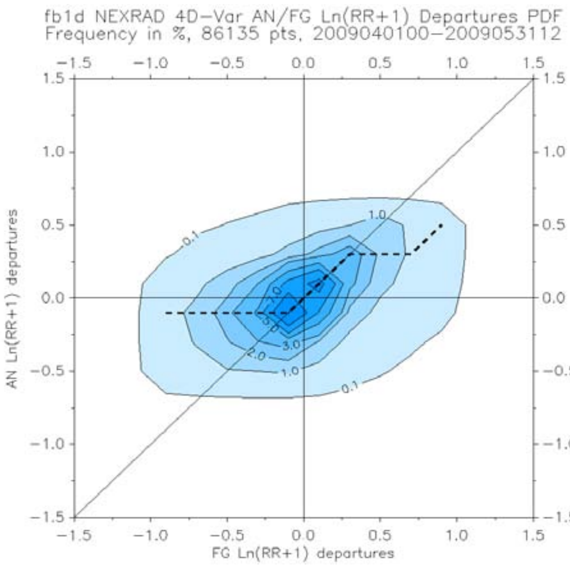

Figure 3: Joint PDF of analysis versus background departures in ln(RR_{6h}+1) space from a direct 4D-Var assimilation experiment with NCEP Stage IV 6-hourly precipitation accumulations (RR_{6h}). Statistics are computed over April-May 2009 (86,135 points). Blue shading shows the frequency in each class, with contour intervals 0.1, 1, 2, 3, 5, 7, 10 and 15%. The black dotted line indicates the location of the PDF mode in each class.

7. Summary and recommendations

The amount of work needed for the mere maintenance and small adjustments of LP packages (e.g. every time the non-linear forecast model is modified) should be neither underestimated nor overlooked. New developments of more complex LP parameterizations, which are necessary in order to avoid the divergence from the non-linear forecast model, will become more and more time consuming as a result of their growing complexity and range of applications involved (4D-Var, singular vectors used in ensemble prediction systems, forecast sensitivities). Furthermore, solving new issues in LP linked to increases in horizontal and vertical resolutions and to the general evolution of the DA approach (e.g. long-window weak-constraint 4D-Var, ensemble 4D-Var) will also require more human resources. In my opinion, the trend is unfortunately towards a shrinking of the LP work force rather than the opposite, which might put these future developments at risk or at least slow down their progress.

Some efforts should be devoted to investigate further the benefits of smoothing either the observations or the non-linear trajectory in time (or in space) since there is evidence that this can help the performance of 4D-Var, especially when assimilating cloud/precipitation observations.

The “0-value” issue should also be addressed, maybe through the usage of a first-guess which is modified from the background so as to produce clouds/precipitation. However, it is not clear whether this approach is applicable and desirable in the 4D-Var context. The inclusion of a cloud condensate variable in the DA control vector might also help.

A reduction of the asymmetry of analysis departures found when assimilating cloud/precipitation observations might be obtained through the relaxation of the saturation constraint imposed on increments or through the extension of the DA control vector to clouds and precipitation.

Experiments should be conducted in the mesoscale DA community to assess the validity of the tangent-linear approximation in time at high resolutions (5 km or better) and to determine whether there is an upper limit to the level of complexity achievable in LP (e.g. in terms of microphysical processes representation).

References

Caumont, O., V. Ducrocq, E. Wattrelot, G. Jaubert and S. Pradier-Vabre, 2010: 1D+3D-Var assimilation of radar reflectivity data: a proof of concept, *Tellus*, **62A**, 173-187.

Hólm, E., E. Andersson, A. Beljaars, P. Lopez, J.-F. Mahfouf, A. J. Simmons and J.-N. Thépaut, 2002: Assimilation and Modelling of the Hydrological Cycle: ECMWF's Status and Plans, *ECMWF Technical Memorandum No.* ***383***, 55 pages, available from ECMWF, Reading, UK.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV Hourly Precipitation Analyses: Development and Applications, *in Proceedings of the 19th AMS Conference on Hydrology, San Diego, CA (USA),* 5-14 January 2005.

Lopez, P., 2010: Direct 4D-Var assimilation of NCEP Stage IV radar and gauge precipitation data at ECMWF, *ECMWF Technical Memorandum No.* ***627***, 27 pages, available from ECMWF, Reading, UK.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of the four-dimensional variational assimilation. Part I: Experimental results with simplified physics, *Q. J. R. Meteorol. Soc.*, **126**, 1143-1170.

Stiller, O., 2009: Efficient moist physics schemes for data assimilation. II: Deep convection, *Q. J. R. Meteorol. Soc.*, **135**, 721-738.