348

The Development of 12-hourly 4D-Var

F. Bouttier[1]

with Annexes supplied by M. Miller, M. Hortal
and L. Isaksen

Research Department

[1]MeteoFrance, Toulouse, France

September 2001

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading, Berks RG2 9AX
library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/pressroom/publications.html

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

The Development of 12-hourly 4D-Var

F. Bouttier

with Annexes supplied by M. Miller, M. Hortal and L. Isaksen

Abstract

The 12-hourly 4D-Var assimilation was implemented operationally in September 2001 after thorough validation. It is essentially a straightforward doubling of the time window used in the 4D-Var analysis. Preliminary testing has suggested that a time window of 12 hours has some advantages over a 6-hour window, but a 24-hour window would not be appropriate for the current data assimilation system. The doubling of the time window exacerbates some weaknesses of the existing incremental 4D-Var system and several refinements have been developed in order to alleviate them.

1. Introduction

The 4D-Var algorithm for data assimilation was first implemented at ECMWF in November 1997 (Rabier *et al.* 2000, Mahfouf *et al.* 2000, Klinker *et al.* 2000) as an upgrade of 3D-Var. The main justification for using 4D-Var was the provision of flow dependent structure functions, in other words, that the departures of the observations from the background are interpolated consistently with the linearized model dynamics and physics. It has been demonstrated (Fisher and Andersson, 2001, their Appendix D) that the flow dependence in 4D-Var is due to a combination of two separate effects: covariance propagation and propagation of the analysis increments.

For each observed datum, the flow dependence increases with the lag between the observation and the starting time of the 4D-Var window[1] (0300[2], 0900, 1500 and 2100 in 6-hourly 4D-Var; 0300 and 1500 in 12-hourly 4D-Var). The 4D-Var structure function for an observation near the starting time is the same as in 3D-Var; namely the one implied by the J_b background term (Derber and Bouttier 1999), which is static. Attempts to introduce flow dependence in the J_b-term using Kalman filtering (Fisher 1998) or flow-dependent variance statistics (Andersson *et al.* 2000) have not demonstrated significant benefits, so far.

Idealized studies with 4D-Var have demonstrated that longer windows provide structure functions with desirable properties: they conform to our physical knowledge of the synoptic dynamics, and they are more consistent with objective diagnoses of the forecast error growth such as singular vectors (Thépaut *et al.* 1993). Thus it is natural to try and maximize the return on investment in 4D-Var by increasing the length of the assimilation window. On the other hand, longer windows test the limits of validity of the incremental 4D-Var algorithms. The approximations of the incremental approach will need to be diagnosed and remedied as far as possible.

---

[1] The 4D-Var window is defined as the time interval during which observations are taken into account by the observation term of the 4D-Var cost-function. The cycling period of the ECMWF 4D-Var assimilation system is equal to the 4D-Var window length i.e. no observations are left out or used twice.

[2] All times are UTC.

2. Technical overview

The architectures of the 6- and 12-hourly data assimilation systems are compared in Fig. 1. The organization of the 4D-Var windows provides approximately the same dissemination times in both systems for the 1200 analysis and the ensuing forecasts. The observation cutoff times are approximately the same; giving the 12-hourly system access to slightly more late observations around 0600 and 1800 UTC, with no measurable impact. The static surface analysis modules (soil wetness and temperature, SST and ice, snow cover) are scheduled in order to maintain the same archiving as before. The schematic below indicates that the surface analyses at 0600 and 1800 in the 12-hourly system are largely diagnostic as their outputs are not directly used to initialize any forecast.

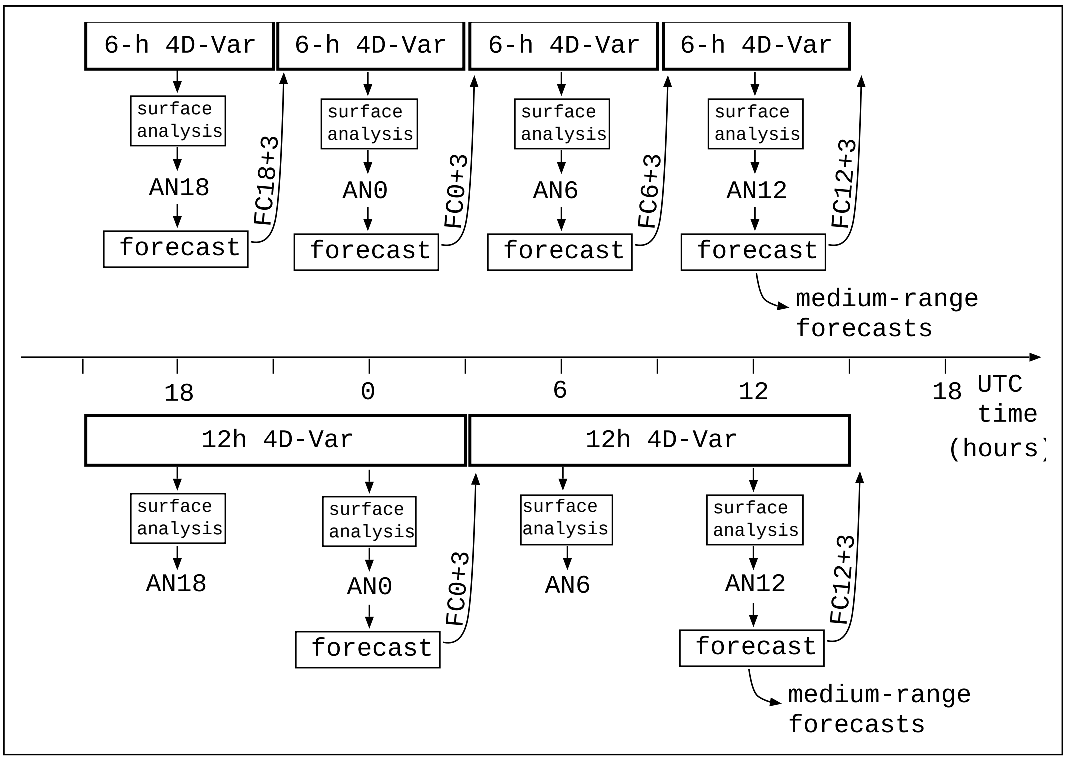

Fig 1: Schematic architecture of the 6- (old) and 12-hourly (new) data assimilation systems, as a function of meteorological time. They are depicted above and below the time axis, respectively. The fields are tagged according to type, time and step: ANhh for analysis at time hh, FChh+ss for forecast starting at time hh and range (or step) ss. All times are UTC.

The concept of 'analysis time' becomes blurred as the 4D-Var window is lengthened. The 4D-Var algorithm does not produce an analysis at a particular time; fundamentally it analyses the full sequence of atmospheric states inside the assimilation window, at the frequency of the model timestep, using all available observations simultaneously. For practical purposes this becomes significant with a 12-hourly cycling, as it also affects the concept of range in the ECMWF forecasts. This range depends on the forecast aspect under consideration:

- the time of the most recent observation used to prepare the forecast's initial conditions traditionally defines the nominal analysis time (here 0000 and 1200, plus three hours inclusive). The 12-hourly 4D-Var still follows this convention at 0000 and 1200, but not at 0600 or 1800.
- the starting time of the model spinup is, in an ideal implementation of 4D-Var, at the beginning of the window (0300 and 1500 for the 1200 and 0000 forecasts, respectively). In the implementation described

in Section 4.5 (“Late Start 4D-Var”), however, spinup starts at the nominal analysis time (0000 and 1200) for all imbalances introduced by the incremental formulation.

- the “best analysis”, i.e. the analysed state that is closest to the real atmosphere as measured by the distance to observations, is (as we shall see in Section 4.5) near the middle of the 12-hour window: at 0900 and 2100 for the 1200 and 0000 assimilations, respectively.

The concept of “first-guess” has become redundant. The corresponding type (FG) in the Mars archive has thus been made obsolescent. The forecast Mars type (FC) is used instead.

Apart from these changes, and unless mentioned below, the ingredients of the 6- and 12-hourly 4D-Var are currently identical: J_b statistics[1], tuning of observation errors, screening and thinning thresholds, inner and outer loop resolutions, linearized physics, initialization technique, number of iterations in the minimization, all have been kept identical in the comparisons between the two systems.

3. Summary of preliminary experimentation

A first set of experiments was run using the configuration available in 1999, with 3 window lengths: 6, 12 and 24 hours. The impact in terms of forecast scores is shown in Fig. 2. Some advantage was apparent for 12-

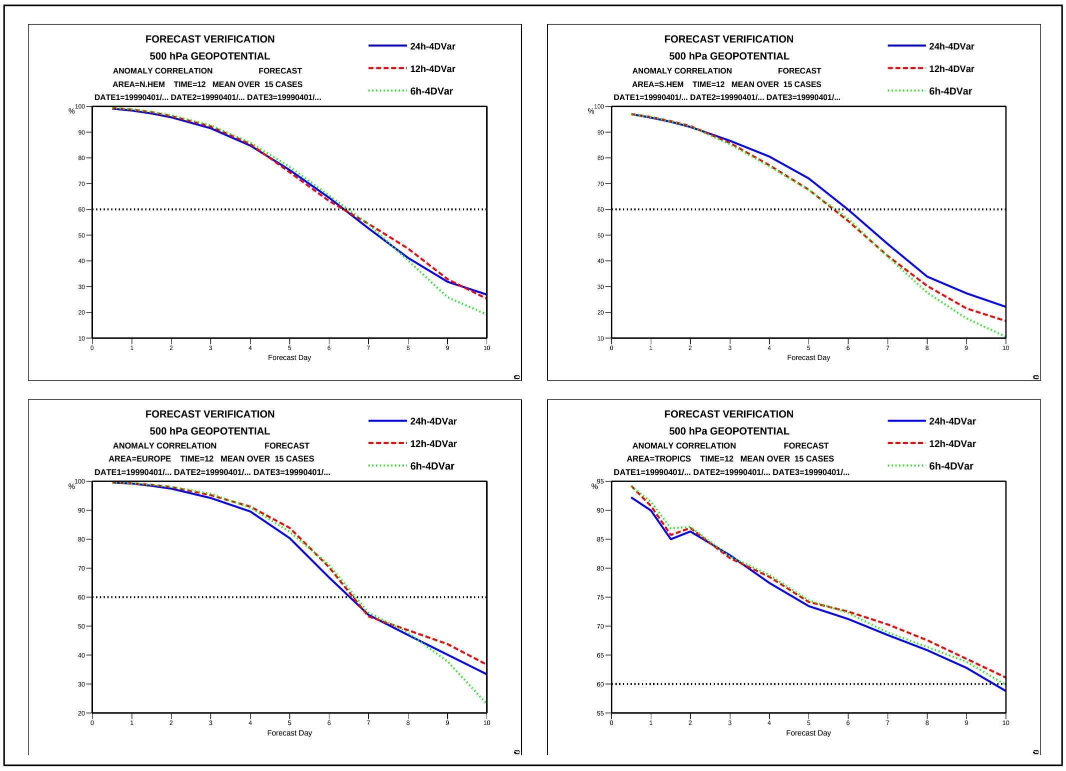

Fig 2: Forecast scores compared in data assimilation experiments using 4D-Var with windows of 6, 12 and 24 hours. These experiments used a low-resolution model of T159L50. The four panels show N. Hem, S. Hem, Europe and Tropics.

---

[1] A re-calibration of the J_b-statistics based on an ensemble of 12-hour 4D-Var assimilations is currently underway.

hourly 4D-Var, mainly in the European area and the Tropics (lower panels), whereas the 24-hour window performed well in the Southern Hemisphere (top right panel). More detailed investigations revealed that the 24-hour window led to a very poor fit of the analysis to observations, and to a strong, unphysical large-scale pattern of the analysis increments that suggests a problem with tidal waves and/or the diurnal cycle over land. The 12-hour window exhibited similar weaknesses, but in a much more acceptable manner. It was decided to experiment further with the 12-hourly system. In subsequent high-resolution experiments the 12-hourly system was found to consistently outperform the 6-hourly 4D-Var in two independent periods (a total of 35 forecasts).

An increase of inner-loop resolution from T63 to T106 was tested and results are shown in Fig. 3. The conclusion from this set of experiments is that higher-resolution inner loops improve both systems similarly. Diagnoses of incremental inconsistencies showed that observations are less affected by incremental inconsistencies with the T106 inner loops (Fig. 4).

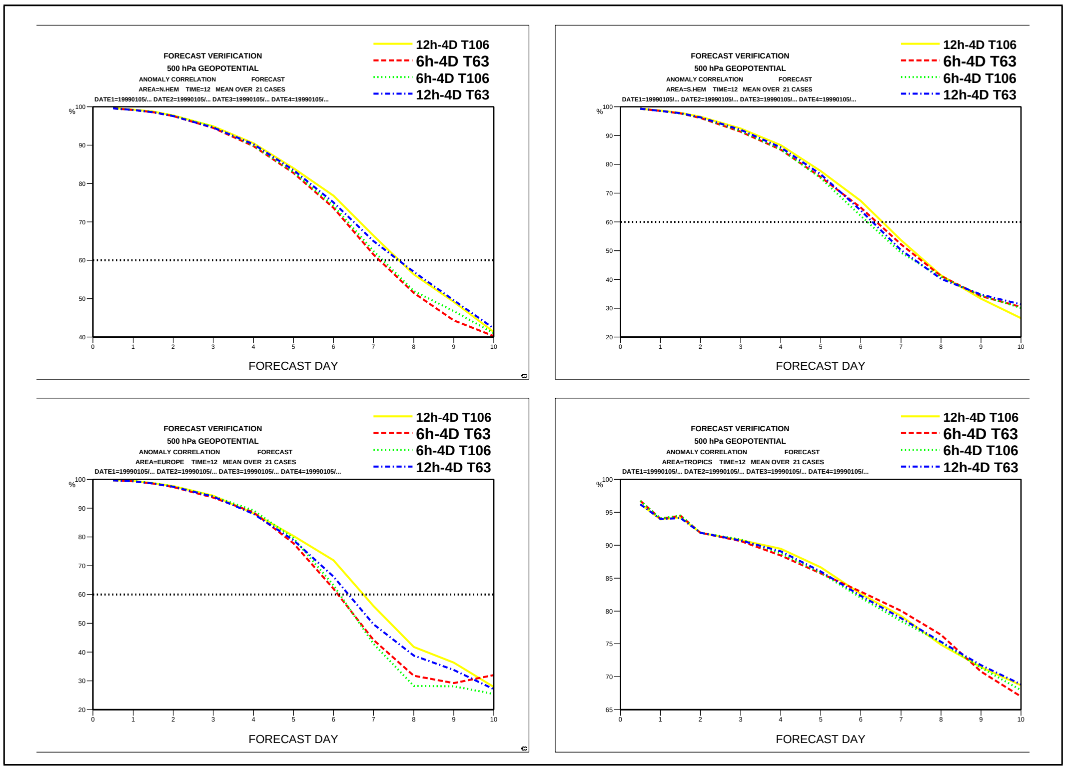

Fig 3: Impact on the forecast scores of increasing the truncation of the inner loop model from T63 to T106, in both 6- and 12-hourly 4D-Var.

In 1999 the vertical resolution of the operational system was upgraded from 50 to 60 levels. The experimentation of 12-hourly 4D-Var was thus redone with 60 levels. In this new framework most of the advantage of 12- over 6-hourly cycling seemed to have disappeared, and the impact was as described in Section 5 below. This was unexpected but not absurd since 4D-Var depends on a number of approximations such as the linearisation of the model and physics, and the properties of model errors. These approximations apparently have different consequences with 50 and 60 levels. Several experiments have been run to ensure that the loss of benefit of using 12-hourly 4D-Var was entirely due to the change in vertical resolution of the model, and not to other aspects of the 50-level system implementation such as the new J_b calibration technique (Fisher, 1999), or to a slowing down of the convergence of the minimization[1].

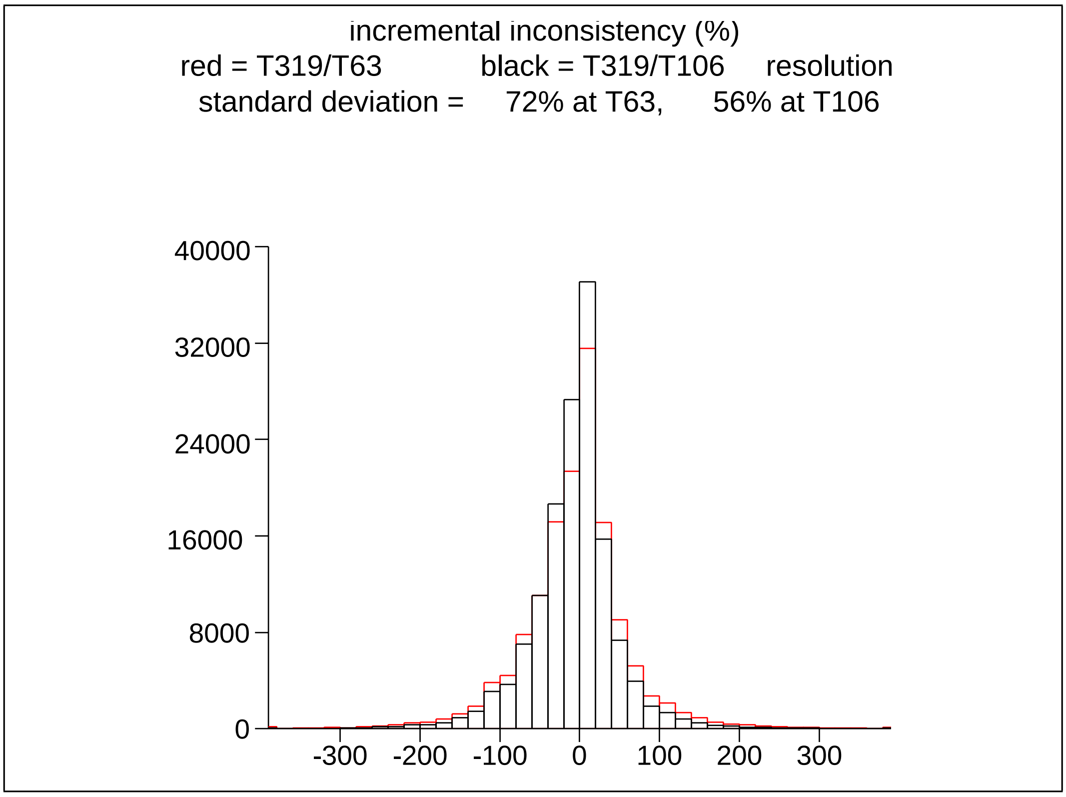

Fig 4: Impact of increasing the inner loop resolution on the use of observations, in terms of incremental inconsistencies, in the context of 12-hourly 4D-Var. The histogram represents the distribution of the quality of the incremental convergence. The abscissa is the percentage of inconsistency between high- and low-resolution increments in the first 4D-Var inner loop (high over low-resolution increment minus one). The ordinate is the number of data in each class. The statistics have been gathered for all used aircraft wind component data for three days of 12-hourly 4D-Var assimilation, with T63 increments (red) and with T106 increments (black). The T106 system exhibits better consistency and has used slightly more observations.

From then on the development of 12-hourly 4D-Var could no longer be justified by a positive impact on forecast scores. It was rather seen as a good basis for further improvements of the 4D-Var algorithm. Previous experimentation with the L31 and L50 system had demonstrated its potential for improving the forecasts. A longer time window was seen as a good framework to expose the strength and weaknesses of the 4D-Var approach. Keeping the 12-hourly system in parallel with the operational 6-hourly one for a longer time was not an option as it was consuming maintenance and experimentation resources. The strategy was thus to ensure that 12-hourly 4D-Var would not degrade the forecast performance and that it did not contain any scientific or technical flaws. An extensive evaluation of the 12-hourly 4D-Var performance was carried out using the latest operational system as reference. A few specific weaknesses were identified and corrected, as explained in the next section, and the resulting package was evaluated as explained in Section 5.

---

1. The 12-hourly 4D-Var minimization is currently being run with 20% more iterations than the 6-hourly one. This is a security measure to guarantee the robustness of the 12-hourly operational system. Tests with the same number of iterations as the 6-hourly system have shown that the meteorological quality of the system is not affected, and that less iterations could be used safely.

4. Related modifications of the assimilation system

4.1 Potential sources of errors in incremental 4D-Var

The theory of 4D-Var predicts (Pires *et al.* 1996) that this algorithm will misbehave if nonlinearities or model errors become significant. Here, model errors mean "errors in the evolution of analysis increments as predicted by the model inside each 4D-Var window" (which excludes external or slow-growing errors of the assimilation system). In the incremental formulation of 4D-Var as used at ECMWF (Courtier *et al.* 1994), this concern extends to "errors in the predictions of the linearised, low-resolution model":

- errors in the formulation of the model (physics and dynamics),
- linearization errors (in terms of the evolution of the analysis increments)
- inconsistencies between the high-resolution assimilating model and the simplified low-resolution model used in the incremental minimization.

The net effect of these errors is to degrade the fit of the analysis to the observations, which is a feature of 12- vs 6-hourly 4D-Var (see Section 5 for more comments). The distribution in time of the analysis fits (Fig. 5) demonstrates that the most realistic analysis is near the middle of the time window, and that the errors in the formulation of the model (responsible for the convexity of the curve of analysis fit, see Talagrand 1999) are small compared to the other sources of errors which are responsible for the overall level of the curve. This suggests that it is premature to worry about an explicit representation of model errors in the 4D-Var algorithm.

4.2 Diagnoses of incremental errors

By monitoring how observations are used inside inner and outer loops of the incremental 4D-Var, it is possible to diagnose the effect of errors related to the incremental formulation. The outline of the incremental technique is as follows:

i) compare the observations with the forecast starting from the high-resolution background, x_{bhr}

ii) prepare the background fields at low resolution, x_{blr}

iii) linearize the model at low resolution, in the vicinity of the model trajectory starting at x_{blr}

iv) compute the 4D-Var optimal increments δx_{lr} using the linearized, low-resolution model and observation operators to define the structure functions; this defines $\delta H_{lr} = H(\delta x_{lr})$, the vector of low-resolution increments on the simulated observed parameters (H is the so-called observation operator)

v) compare the observations with the forecast that starts from the corrected high-resolution background: $x_{bhr} + T(\delta x_{lr})$ where T is a transform operator that maps the low-resolution increment into the high-resolution model state. This defines $\delta H_{hr} = H[x_{bhr} + T(\delta x_{lr})] - H[x_{bhr}]$, the vector of high-resolution increments on the simulated observed parameters.

A necessary condition for the technique to work is that step (iv) and step (v) should yield similar increments in terms of the observed quantities, since (iv) is what 4D-Var uses in the optimization, and (v) determines the fit to observations. In a non-incremental 4D-Var these increments are identical. If they are different it means that the actual effect of the analysis increments in the forecast model is different from what was implied by the 4D-Var cost function; in other words, it means that the incremental technique is distorting the 4D-Var analysis in a spurious way.

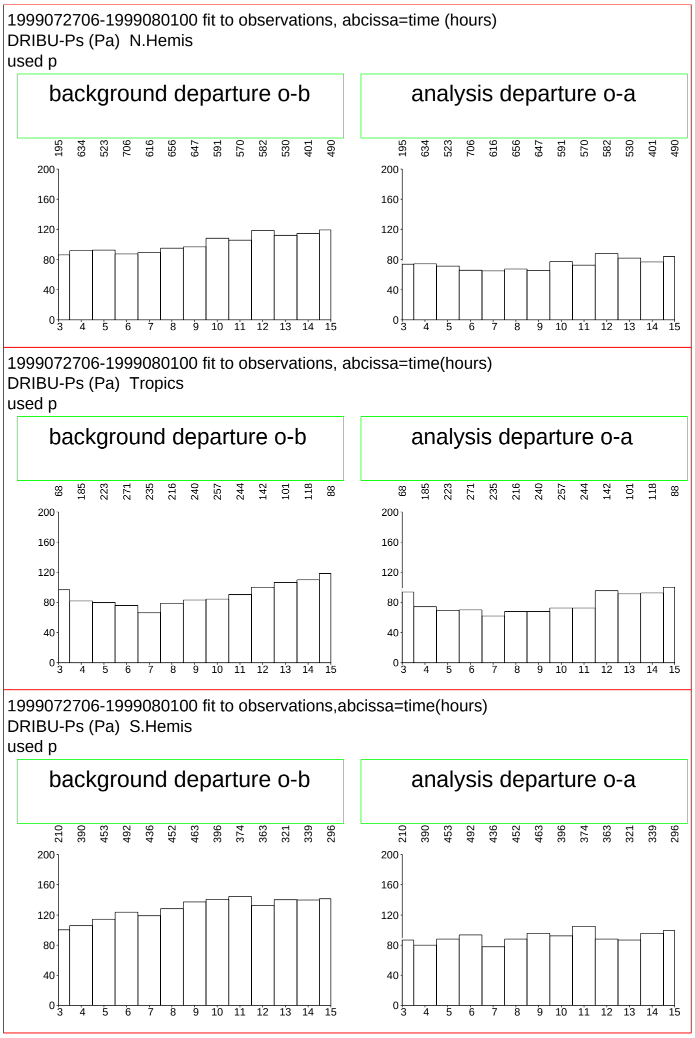

Fig 5: Distribution in time of the rms average differences between the 12-hourly 4D-Var analyses (right), background (left) and the observations, as a function of time, for three different areas. The abscissa is UTC time, using the same binning as 4D-Var. The ordinate is the rms fit in Pascals. The rows of rotated numbers indicate the population of each bin. Drifting buoy surface pressures have been used because they exhibit little time dependency linked to the observing system. The plot reveals that background departures increase with time, as do forecast errors with range, and analysis departures are almost flat inside the 12-hour 4D-Var window. The exception is the tropical area, where there are inconsistencies in the analysis of the tidal signal on surface pressure. This suggests that 4D-Var internal model errors can be neglected in the current ECMWF system for extratropcial purposes.

Diagnoses of the relative inconsistencies between the increments δH_{lr} and δH_{hr} have been done for all observations used by 4D-Var. Ideally these differences should be 0%, if they exceed e.g. 50% it means that the 4D-Var increment is distorted by 50% in terms of the observed values. It is found that:

- observations at low levels, near the tropopause and humidity observations are most affected;
- radiance observations are least affected;
- inconsistencies grow linearly as a function of time inside the 4D-Var window and can be considerable (of the order of 100%) near the end of the 12-hourly 4D-Var window;

- inconsistencies are noticeably reduced when the inner loop model horizontal resolution is increased (e.g. from T63L50 to T106L50), but not when the vertical resolution is increased (from L50 to L60).

These results deserve further investigation; they suggest that the incremental problems may be reduced when the inner loop model is made more realistic by improving the horizontal resolution, but they may also increase if more non-linear phenomena start to be resolved (e.g. the PBL processes in the L60 system). A meteorological illustration is provided in Fig. 6: in an active cyclogenesis, the inner loop model is so unrealistic that the incremental analysis produces high-resolution increments that are completely different to those implied by 4D-Var (which are supposed to be statistically optimal). Here, the problem is not so much a lack of resolution in the inner loops as the discrepancy between inner and outer loop models.

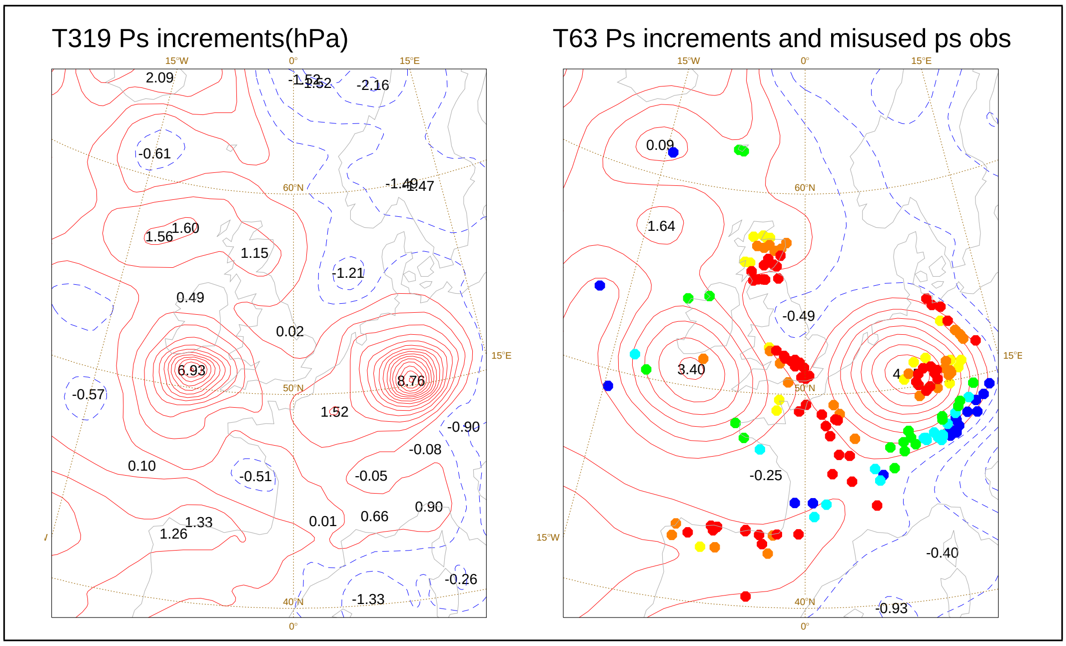

Fig 6: Surface pressure increments as seen by the 12-hourly 4D-Var first inner loop at high (left panel, T319) and low resolution (right panel, T63), 9 hours into the 4D-Var window, for an analysis of the 1999 first Christmas storm (window between 3 and 15UTC on 26 December 1999). The isolines and numbers depict the increments, in hPa. The coloured dots indicate used surface pressure observations for which the inner and outer loop increments disagree by more than 100%, meaning that the second inner loop is going to move the high-resolution analysis further away from the observations than the background. As a result, the 4D-Var analyses of the storm at 12UTC were poor with both 6- and 12-hourly cycling.

4.3 Incremental quality control

Inconsistencies between the inner- and outer-loop increments show which observations are badly used by the incremental algorithms: this information can be used to avoid using them in the analysis. Intuitively, if the high-resolution model reacts to the increments in a way that is very different from what was intended by 4D-Var, it will be equivalent to adding noise to the analysis, which will tend to degrade it. There is no guarantee that inconsistencies are actually caused by the observations that suffer most from it, but since 4D-Var tends to react locally to observations, it is very likely. It makes sense to withdraw these observations from the analysis. Accordingly, the following quality control step has been designed. At the end of the first 4D-Var incremental inner loop, a new quality control procedure is applied when the first low-resolution increment is added to the high-resolution background:

vi) For each observed datum, if the increment at low resolution δH_{lr} is greater than a predefined threshold (absolute value greater than 50% of the assumed observation standard error), then the relative difference between the high- and the low-resolution increment is computed: $d = |\delta H_{hr}/\delta H_{lr} - 1|$;

vii) If d is greater than 100%, then the assigned observation error is increased using a smooth function of d, such that the observation weight becomes negligible for values of d greater than a predefined threshold.

viii) the new (increased) observation errors are used in the second 4D-Var incremental inner loop, so that the final analysis effectively will ignore observations with a large d.

The thresholds have been tuned in a conservative way, so that very few observations are rejected. This is enough to improve the diagnoses of incremental inconsistency, but not enough to have a measurable impact on the forecast scores. The intention is to reject most data in the most problematic cases, while retaining the observations that provide useful information and are correctly used by the analysis.

The net meteorological effect of this change is to use about 1% less of all conventional observations (almost no satellite observations are rejected). The background and analysis fits to used observations are improved in a long assimilation, notably with respect to low-level and SSM/I observations of humidity. The high- and the low-resolution increments in the second 4D-Var inner loop are more consistent, notably with respect to radiosonde temperatures (which was expected since these observations are generally available late in the 4D-Var window).

4.4 Improvements to the linearization trajectory

In the incremental 4D-Var, the low-resolution model is linearized in the vicinity of a low-resolution, nonlinear forecast started from the background state[1]. Consequently, the quality of the linearization is affected by discrepancies between the high- and low-resolution model runs and between their initial states. This is more damaging in 12- than in 6-hourly cycling because the low-resolution model runs for twice longer. The following improvements have been implemented in the trajectory computation:

- change of the low-resolution nonlinear model physics to use the same physics package as the high-resolution model. Previously, an obsolete physics package (diagnostic clouds instead of prognostic cloud scheme) was used. The physics team has confirmed that this resulted in substantial changes to the low-resolution forecast fields.
- provision of realistic cloud fields to the initial state (the background) of the low-resolution nonlinear model, by interpolating the background cloud fields.
- improvement of the interpolation procedure from high to low resolution by using the high quality “full-pos” postprocessing package. Among other features it prevents overshooting in the interpolation of surface fields and negative values of specific humidity or ozone.

It has been confirmed that these changes provided more realistic initial fields to 4D-Var as well as improvements to the trajectory fields. The impact on the forecast scores of a data assimilation suite was not measurable.

---

[1] It would be more correct to carry out the linearization in the vicinity of the forecast that 4D-Var is aiming to correct, i.e. the high-resolution forecast itself. This feature is currently under development.

4.5 *“Late cycling” of 4D-Var*

By construction, incremental 4D-Var requires its control variable, the background constraint and resolution changes to be defined at the start of the time window. This feature of 4D-Var (linked to the non-invertibility in time of the forecast model) can be regarded as a drawback: in the presence of imperfections such as the incremental inconsistencies, 4D-Var will tend to provide the best fit to the oldest observations. In a real-time data assimilation system, one would actually prefer to make the best use of the latest observations.

An exception is the preparation of the final analysis. The inner loop 4D-Var produces an optimal trajectory of increments to fit all the observations. One is free to choose the time at which the trajectory is added to the forecast, as illustrated schematically in Fig. 7. This influences the analysis fit to observations. By selecting the nominal analysis times (1200 and 0000), one gets a better fit to the radiosondes at these times. The downside is that the spinup created by adding the increments starts at that same time, so that it will take longer to settle in the forecast[1] than if the analyses were prepared at the start of the window. Data assimilation experiments with this change only (apart from 12-hourly cycling) did not show any measurable impact on forecast scores.

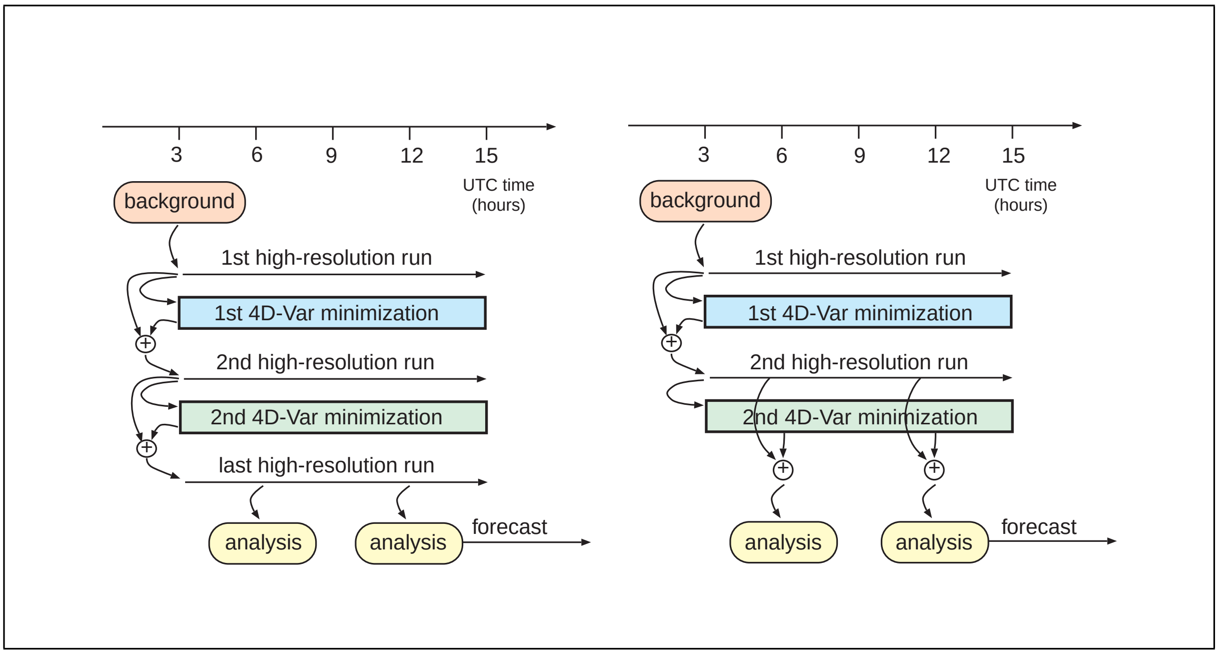

Fig 7: Schematic depiction of the preparation of the analysis in the previous incremental formulation (left) and with the new ‘late cycling’ formulation (right). The incremental 4D-Var involves two low-resolution 4D-Var minimizations that compute increments using observation departures from high-resolution ‘runs’ (non-linear forecasts compared to observations). In the previous formulation, the increments were added at the beginning of the time window (here, the 3 to 15UTC window as an example, the procedure is identical for the 15 to 3UTC window). Now, with “Late Start” the increments are added to the penultimate trajectory later, at the nominal analysis times.

---

[1] The most concerning kind of spinup in the ECMWF model forecasts is on the precipitation, which lasts for more than one day, and is only marginally affected by this change.

4.6 Consistent cycling of cloud fields

There is no 4D-Var cloud analysis to speak of. In the previous system, the analysis cloud fields were kept equal to the background cloud fields provided by the previous forecast: they were not updated for consistency with the analysed fields of wind, temperature and specific humidity. In the 12-hourly cycling, it would mean that the analysis cloud fields are the product of inconsistent model runs over nine hours since the beginning of the 4D-Var window (compared to three hours with 6-hourly cycling). This weakness is not believed to have a notable impact on the forecast performance of the current system, but it may be a problem with the future plans for assimilating more and more observations related to clouds and humidity.

A better initialization of clouds has been implemented, by which the cloud fields are taken from the latest available high-resolution forecast in the incremental procedure (which is the one at the end of the first inner loop). At no extra cost, it provides cloud fields that are more consistent with the other variables in the analysis. It also provides the hooks for a future cloud analysis system by which the high-resolution cloud field could be analysed in parallel with the low-resolution 4D-Var runs.

5. Pre- and post-operational validation

Data assimilation experiments have been carried out with the T319L60/T63L60 system over four periods of two to four weeks each, with all the above features included. The impact on the medium-range forecast scores is neutral on average for all areas, parameters and verification techniques (verification against operational analysis, own analysis and radiosondes), including Europe. Over Europe, all scores that showed a detrimental average impact have been subjected to a T-test to confirm that the impact was not statistically significant. The only consistent and significant impact was found on short-range wind scores against the own analysis of each experiment, which are beneficial and suggest improved consistency inside the data assimilation system. The proposed explanation is that 12-hourly 4D-Var analyses are better dynamically balanced than 6-hourly 4D-Var analyses. Several subsets of the 12-hourly 4D-Var package of modifications have been tested individually:

- the doubling of the 4D-Var period on its own
- the modifications to the cloud, low-resolution physics and incremental interpolations in the 12-hourly system
- the late cycling of 12-hourly 4D-Var
- the incremental quality control
- the use of four incremental inner loops instead of two; this was withdrawn from the package because of its cost

Each test has been run over a minimum of four weeks and showed a non-measurable impact on the forecast scores. The impact on the use of observations has been assessed too. The salient features are:

- an increased distance of the background to observations in the second half of the 12-hour 4D-Var window (see also Annex C). This is a normal consequence of the use of a longer assimilation cycle. The background fit to observations in the first half is generally unaffected, with the exception of SSM/I TCWV which are much improved; this is interpreted as an improvement in the assimilation of humidity.
- a generally increased distance of the analysis to observations (in 12-hourly 4D-Var compared to 6-hourly 4D-Var). This is a consequence of the use of more flow-dependent structure functions which tend to be broader in space than the ones in 6-hourly 4D-Var. It is also caused by the constraint of fitting twice as

many observations in a single analysis, so that it is less likely that observations are overfitted. Hence, the increased distance to observations is not necessarily a bad sign.

- an improvement of the analysis fit to observations when the incremental quality control is activated.
- no significant change in the number of observations used or rejected. In an operational context one expects 12-hourly 4D-Var to use slightly more observations in the first half of the window thanks to a longer cutoff time; this could not be assessed in experimental mode.

A study of the numerical behaviour of the minimization did not find any sign of a degraded preconditioning or speed of convergence of the minimization problems.

The corresponding e-suite was run for 72 days (from 1 July 2000) and did not show any significant negative impact on hemispheric scores, which led to the operational implementation of 12-hourly 4D-Var in September 2000. The performance in this summer period was however, significantly negative for Europe in the medium range. The length of the 4D-Var window was therefore further reviewed in the context of the higher-resolution system that was implemented in November 2000. The scores displayed in Annex B show that the change from 6- to 12-hourly 4D-Var assimilation technique has a positive impact on the T511/T159 system (also over Europe), which confirms that there are interactions between model characteristics and the performance of a long-window 4D-Var.

6. Discussion: the future evolution of 4D-Var

The increase of the cycling period of 4D-Var is in line with the original long-term plans for the development of 3D- and 4D-Var as part of the IFS project. The development strategy for data assimilation has been continuously evolving over the years with the accumulation of new theoretical and experimental knowledge.

The conclusion is that an increase in length of the 4D-Var window is a good way to improve the structure functions by making them more flow dependent at no extra cost (although, in an operational context, the required peak computer power is doubled), and by providing a better multivariate coupling between the observed variables and the model fields. The limit to this improvement will come from the underlying hypotheses of incremental 4D-Var: a good consistency between the linearized inner-loop and non-linear outer-loop models, and small model errors. This paper has shown that the most serious limitation today seems to be the lack of consistency between the inner and outer loop models. This hypothesis is supported by the evidence of internal inconsistencies in the use of observations by 4D-Var, which grow with the length of the window, and by the inability to get a benefit from the use of more frequent high-resolution updates in the incremental algorithm (this was already the case with 6-hourly 4D-Var).

In the incremental algorithm, some obvious inconsistencies arise from the use of different atmospheric models in the inner and outer loops. The inner loop model resolution and physical realism should therefore be improved as much as possible. The trajectory used for linearization should be as consistent as possible. To fully understand the problems with the incremental technique, it would be necessary to develop a non-incremental version of 4D-Var for use as a reference. It would help to understand precisely the problems that are caused by model inconsistencies and linearization errors.

In the long run, even if the incremental technique can be improved, 4D-Var will always suffer from problems with the linearization of increments. Future improvements in the quality of the assimilation system (e.g.

through the use of better observations and models) should reduce the amplitude of analysis increments, which will solve some of the linearization issues. On the other hand, the trend in NWP is to resolve smaller scales as well as increasingly complex atmospheric phenomena (PBL physics and cloud microphysics) which may be too nonlinear for a 4D-Var algorithm to handle.

In conclusion, the main principle behind the long-term improvement of 4D-Var is going to be the physical realism of all its components: inner loop model resolution and physics, accuracy of the linearization, quality of the background term J_b, efficient use of the observations, physical initialization. The length of the 4D-Var time window is a useful degree of freedom that can be tuned to achieve the best possible compromise between rich, flow-dependent structure functions, and the limitations brought by: linearization, the incremental technique, model errors and practical constraints of the operational production. Since these factors are constantly evolving, it will be necessary to review the length of the 4D-Var window from time to time.

7. Conclusion

Compared to the previous operational system, the most significant component of the 12-hourly 4D-Var package of modifications is the doubling of the 4D-Var window and data assimilation period. It is supposed to improve the quality of the analyses by the use of more realistic structure functions. The technical change costs nothing, except some strain on the real-time schedule of the operational production. The scientific change has shown some good potential in earlier versions of the system, but the forecast impact is very sensitive to changes in the model itself. When the 12-hourly 4D-Var package was made operational, the impact on the forecast scores was about neutral on the forecast scores, with improved forecast consistency, and a reduced analysis fit to observations. This provides a good basis on which to test and develop further improvements to the 4D-Var system. In Annex B we shall see that in more recent experiments using the high resolution forecasting system (T511/T159) 12-hourly 4D-Var has outperformed the 6-hourly system.

Some changes to the assimilation have been developed in order to improve on various issues, which are becoming more important with the new, longer window, in relation with approximations in the incremental formulation and the assimilation of cloud fields. These will be pursued. Further developments to improve the incremental technique are underway.

Acknowledgements

We are grateful to the members of the Data Assimilation section for their contributions towards the 12-hourly 4D-Var system, and to Anthony Hollingsworth and Adrian Simmons for comments on the manuscript. Christian Jakob suggested the modifications to the low-resolution physics package, and Pedro Viterbo highlighted the problem with interpolation of surface fields. We thank Els Kooij-Connally for editing and formatting the manuscript.

References

Andersson, E., M. Fisher, R. Munro and McNally, A., 2000: Diagnosis of background errors for observed quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Quart. J. Roy. Met. Soc.*, **126**, 1455-1472.

Courtier, P., Thépaut, J.N. and Hollingsworth, A., 1994: A strategy for operational implementation of 4D-VAR, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120,** 1367-1388.

Derber, J. and Bouttier, F., 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A,** 195-222

Fisher, M., 1998: Development of a simplified Kalman filter. ECMWF Tech. Memo. 260.

Fisher, M., 1999: Background error statistics derived from an ensemble of analyses. RD internal Memorandum, 3 Sept 1999, ref.R43.2/MF/79.

Fisher, M. and E. Andersson, 2001: Developments in 4D-Var and Kalman Filtering. ECMWF Tech.Memo.no.347.

Klinker, E., Rabier, F., Kelly, G. and Mahfouf, J.F., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.*, **126,** 1191-1216

Mahfouf, J.F. and Rabier, F., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part II: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.,* **126,** 1171-1190.

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

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

Talagrand, O., 1999: A posteriori evaluation and verification of analysis and assimilation algorithms. Proc. ECMWF workshop on “Diagnosis of data assimilation systems”, 17-28. Available from ECMWF, Shinfield Park, Reading RG2 9AX UK

Thépaut, J.N., Hoffman, R., and Courtier, P., 1993: Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.,* **121,** 3393-3414.

Annex A: The performance of 12-hourly 4D-Var in the T511/159 system, and the choice of timestep

by M. Miller and M. Hortal

The high resolution system T511/T159 was successfully introduced into operations on 21 November 2000. During its development a number of choices/decisions were made with regard to the precise configuration of the system such as the size of the timesteps, magnitude of horizontal ‘diffusion’ and precise form of orography. Furthermore, prior to its implementation there was concern as to the overall performance of the 12hour 4D-Var T319T319/T63 system implemented on Sept 12th 2000 and how this would impact the high resolution system. Here, we briefly summarize the results of the substantial amount of experimentation carried out on two of the most critical of these issues: firstly the choice of timestep as this has a direct impact on the cost of the high resolution system and secondly the performance of the T511/TL159 system in 6hr versus 12hr cycling.

1. The timestep

The T319L60 system used in operations prior to the resolution upgrade used a timestep of 20 minutes for both the high resolution component of the assimilation and the 10-day forecast. As the Semi-Lagrangian dynamics does not impose a strict limitation on the upper value of the timestep the choice is determined partly by the need for accuracy/stability in the physical parametrizations, but also by pragmatism.

In the early stages of testing T511 values of 10, 15 and 20 minutes were tried and 15 minutes was adopted for the main development programme. In the run-up to implementation the sensitivity to timestep was revisited in the context of 12hr 4D-Var and Cycle 23R2. Two periods of experimentation were run; from 26th Nov 1999 to 31st Dec 2000 and 9th Aug 2000 to 30th Sept 2000 with the timestep increased to 20 minutes in both the assimilation and forecast. Fig. 1 summarizes the objective scores for the 89 cases scored against the operational analyses as it is the medium-range impact that is presented here. It is clear that there is an overall degradation with the longer timestep although there are some areal and seasonal variations (not shown) such that, for example, Europe is very degraded in winter but slightly positive in summer. The scatter plots, (not shown) do not show dramatic levels of significance but support the general conclusion. Inspection of the statistics of fits to observations showed nothing of note. The potential for further experimentation is discussed later.

2. 12 hour versus 6 hour cycling in 4D-Var

Unfortunately the pre-operational E-Suite for 12hr T319/T63 did not perform as well as the comparable 6hr version, however preliminary results at that time indicated that no such problem appeared to exist with the T511/T159 system. If this proved to be supported by a more substantial set of experimentation then both the long-term viability of 12hr cycling and the importance of the inner loop resolution would be confirmed. The same periods for experimentation as those above were used with Cycle 23R3 to give a 90 case sample. Fig. 2 shows the objective scores for the Northern Hemisphere and Europe. As can be seen this larger sample fully supports the 12hr system at these resolutions and at least for this sample, the European area is particularly

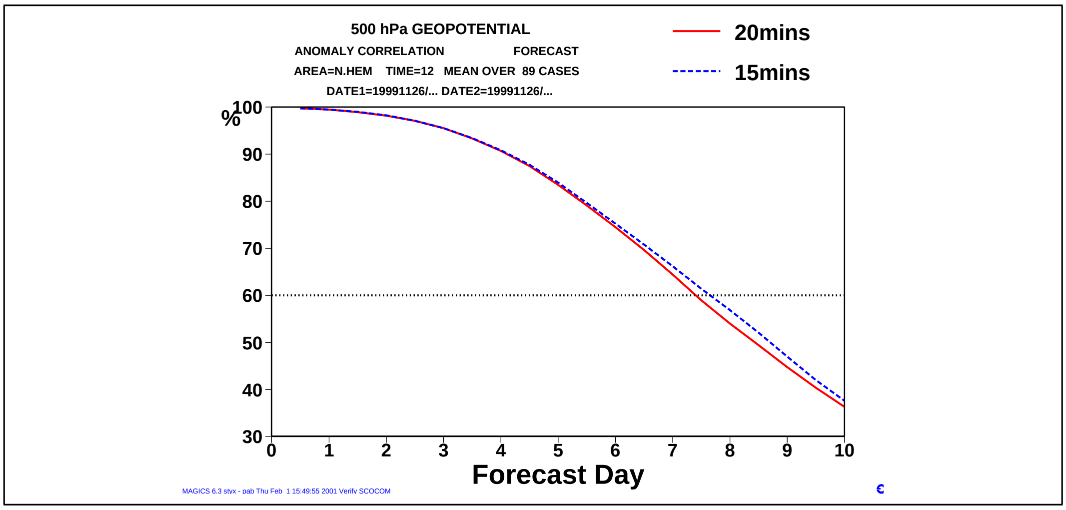

Fig 1: Northern Hemisphere forecast performance at T511/T159 with 20 and 15 min timesteps, averaged over 89 cases.

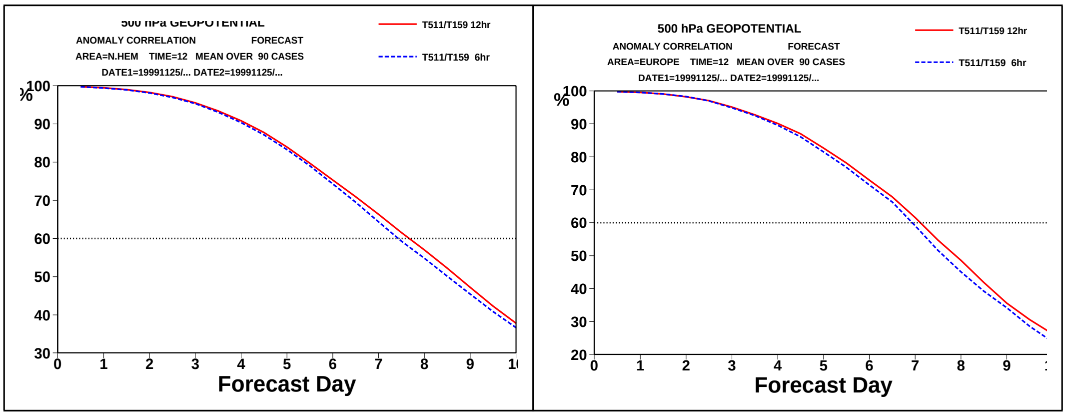

Fig 2: Impact of doubling the 4D-Var cycling period on the forecast scores (N. Hem on the left and Europe on the right), in average over 90 cases, in the context of the T511/T159 high-resolution system introduced in operations in November 2000.

improved. There is relatively little seasonal difference (not shown) but it is noteworthy that the 1000hPa scores for E Asia show the largest positive impact (significant at 98% at D+3, and 99.8% at D+6) which may point to diurnal sensitivities in the length of cycling window.

3. Remarks

The current choice of timestep (15 mins) provides more accurate forecasts than a longer step (and presumably a more stable system) at an increased cost. Whether a further shortening of the step, eg. to 10 mins., would provide any further benefits, has not been extensively tested. This would be computationally expensive and might not be considered cost-effective. Nevertheless some testing should be done in the future, although other ongoing developments in the numerics of both the dynamics and physics will need to be considered in this context. The tests here could also be extended to include hybrids such as using 15mins in the assimilation and 20 mins in the 10-day forecast etc., although this added complexity is not particularly desirable. This study has not addressed the question of the timestep in the lower resolution inner loops of 4D-Var which in principle has a major cost impact. This is currently under consideration although preliminary results do not indicate the prospects of significant savings.

The experimentation presented here clearly indicates that the current 12hr 4D-Var system is preferable to using 6 hourly cycling. Additional experimentation with 12 hr 4D-Var and T319/T159, albeit for only a subset of the dates, shows that higher resolution inner loops contribute most to this improvement (Fig. 3). The prospect of further improvements (at a price) using T255 inner loops is the subject of current study.

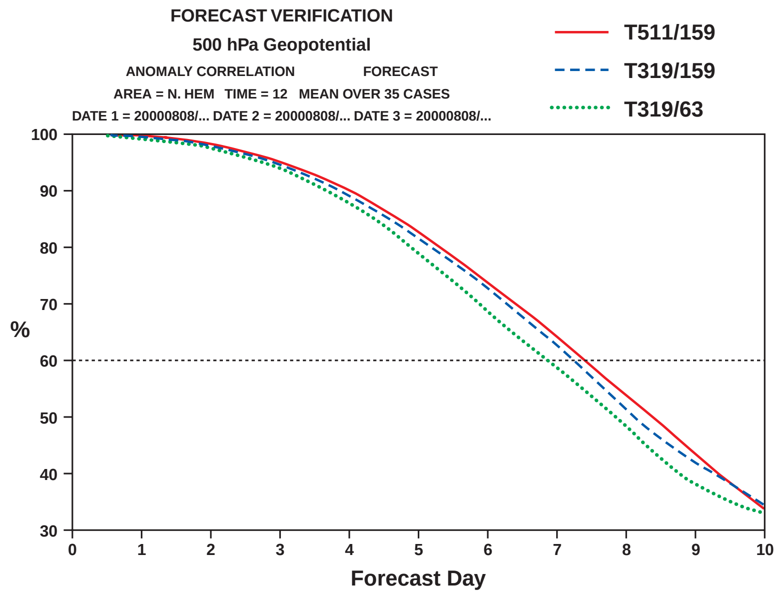

Fig 3: Impact of inner- and outer-loop resolution increase in the context of 12-hourly 4D-Var, in terms of Northern Hemisphere 500 hPa geopotential anomaly correlation, averaged over 36 cases:20000912-12UTC. The three lines represent T319/T63 (dotted green - operational from September to November 2000), T319/T159 (dashed blue) and T511/T159 (operational from November 2000)

None of the above sets of experimentation had any marked tropical impacts although both the 15 minutes timestep and the 12 hr 4D-Var forecasts gave slightly improved tropical objective scores. This is in addition to the systematic improvements in the Tropics due to the T511 resolution per se.

Annex B: The impact of resolution on fits to observations, in 6- and 12-hourly assimilation

***by* L. Isaksen**

For the period 8 August to 31 August 2000 a range of 4D-Var assimilation experiments that investigate the impact of 6h versus 12h cycling and the impact of inner and outer loop resolution are available. For these experiments the first guess and analysis fit to observations has been investigated. It is of particular interest to investigate how resolution affects the forecast errors of the trajectory and the inner loop model integration used during the assimilation. This has been studied by checking how AIREP data fits the first guess fields and analysis fields for each hourly timeslot during the 6h and 12h assimilation period. AIREP data is useful for this task because observations are available continuously throughout the day and it has consistent high quality.

The figure shows the average normalised cost function value as a function of time. The average normalised cost function is similar to the variance by definition, i.e. a quadratic quantity, so the curves have steeper slopes than similar RMSE curves would have. The normalised cost function is used because it is then possible to combine the statistics for both temperature and wind measurements into one number.

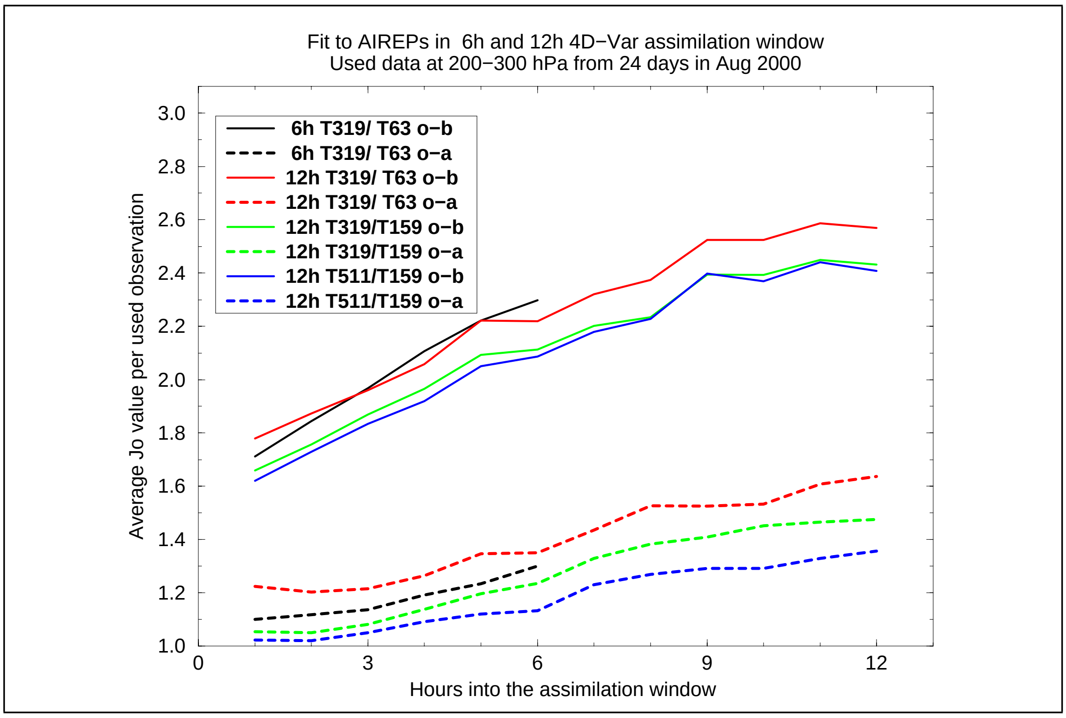

Fig 4: Average normalised cost function for background (o-b) and analysis departures (o-a) for AIREP wind and temperature data plotted as a function of time (hours) within the 6 and 12-hourly assimilation window

The solid lines show the observation-minus-background (o-b) values for 6h 4D-Var T319/T63, and for 12h 4D-Var in three different resolution configurations (T319/T63, T319/T159 and T511/T159).

The 6h 4D-Var T319/T63 o-b fit (solid black line) is similar to the first 6 hours of the 12h 4D-Var T319/T63 o-b (solid red line). During the last 6 forecast hours the 12h 4D-Var T319/T63 o-b fit degrades, as one would expect. The dashed black and red lines represent the o-a (observation - high resolution analysis) fit for the two assimilation experiments. It is seen that the 6h 4D-Var T319/T63 o-a fit is better than for the 12h 4D-Var T319/T63. The analysis near the end of the assimilation window is of worse quality than the analysis at the first part of the assimilation window. This is most pronounced for 12h 4D-Var T319/T63. Because the first guess fields during the last part of the 12h 4D-Var T319/T63 window are clearly of worse quality than for the similar 6h 4D-Var, the 12h 4D-Var has as more difficult task to produce an analysis that consistently fits the observations from the 12h period.

The forecast errors are most important in 12h 4D-Var. They affect both the calculation of innovations in the outer loop and the minimisation performed in the inner loop. To investigate the impact, two additional 12h 4D-Var assimilations were run with higher resolution inner loop (T159) plus an assimilation with higher resolution inner (T159) and outer loop (T511). The curves are respectively green and blue. The solid curves show that the first guess fit is much better for both these higher resolution assimilations than for T319/T63. The T511 outer loop gives a slightly better first guess fit than T319 outer loop. The dashed curves show that the analysis fit to the T319/T159 is better than both 6h and 12h T319/T63 assimilations. The 12h T511/T159 analysis clearly fitting the data better than the 12h T319/T159 assimilation. Especially during the last part of the 12h assimilation window the T511/T159 is improved, resulting in a more uniform use of the data during the 12h window.

4. Remarks

The current choice of timestep (15 mins) provides more accurate forecasts than a longer step (and presumably a more stable system) at an increased cost. Whether a further shortening of the step, eg. to 10 mins., would provide any further benefits, has not been extensively tested. This would be computationally expensive and might not be considered cost-effective. Nevertheless some testing should be done in the future, although other ongoing developments in the numerics of both the dynamics and physics will need to be considered in this context. The tests here could also be extended to include hybrids such as using 15mins in the assimilation and 20 mins in the 10-day forecast etc., although this added complexity is not particularly desirable. This study has not addressed the question of the timestep in the lower resolution inner loops of 4D-Var which in principle has a major cost impact. This is currently under consideration although preliminary results do not indicate the prospects of significant savings.

The experimentation presented here clearly indicates that the current 12hr 4D-Var system is preferable to using 6 hourly cycling. Additional experimentation with 12 hr 4D-Var and T319/T159, albeit for only a subset of the dates, shows that higher resolution inner loops contribute most to this improvement (Fig. 3). The prospect of further improvements (at a price) using T255 inner loops is the subject of current study.

None of the above sets of experimentation had any marked tropical impacts although both the 15 minutes timestep and the 12 hr 4D-Var forecasts gave slightly improved tropical objective scores. This is in addition to the systematic improvements in the Tropics due to the T511 resolution per se.

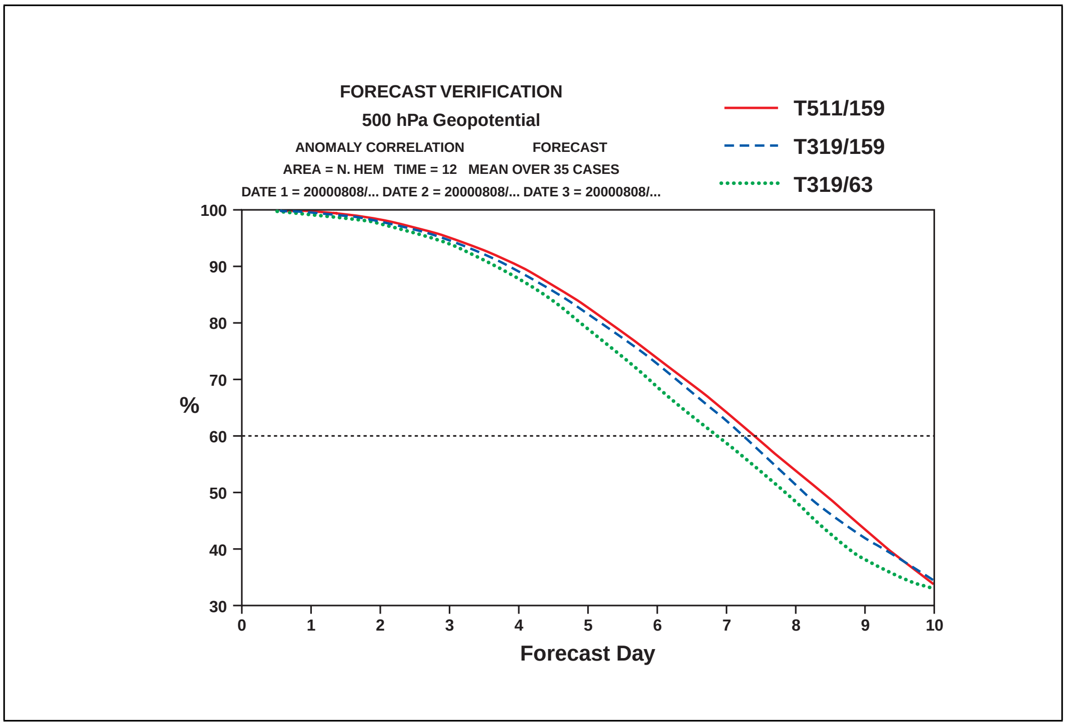

Fig 5: Impact of inner- and outer-loop resolution increase in the context of 12-hourly 4D-Var, in terms of Northern Hemisphere 500 hPa geopotential anomaly correlation, averaged over 36 cases:20000912-12UTC. The three lines represent T319/T63 (dotted green - operational from September to November 2000), T319/T159 (dashed blue) and T511/T159 (operational from November 2000)