Some developments at ECMWF during 2005

M. Hortal
(ECMWF)

1. Introduction

Three are the most important developments having found its way into the operational system in 2005 or early 2006, namely

- A new moist PBL scheme

- A wavelet formalism to compute and apply the background cost function J_b in the 4D-Var assimilation system

- An increase in the horizontal resolution to T_L799 and an increase in the number of vertical levels to 91

On top of that, the first tests at the horizontal resolution of T_L2047, in which the distance between grid points in the Gaussian grid is ~10km, have been run and archived, and a very limited number of diagnostics applied to the runs.

2. The new moist PBL scheme

(M. Köhler 2005: A new PBL parameterization to improve low clouds, ECMWF RD Tech Memo 518)

We have developed a new PBL parameterization that unifies and improves the treatment of clouds in the PBL.

In the previous ECMWF model the boundary layer scheme is formulated in dry variables (dry static energy and specific humidity) and the cloud generation relies on the shallow convection scheme, which is not adequate for the stratocumulus regime. The result is that stratocumulus is systematically underestimated and that too much solar radiation reaches the surface.

Given the problems with that scheme in stratocumulus regions it was decided that an entirely new unified treatment of boundary layer clouds was needed. The main ingredients deemed necessary were: (i) moist conserved variables used throughout the PBL parameterization, (ii) a combined mass-flux/K-diffusion solver, (iii) a treatment of cloud variability and (iv) a treatment of the transition between stratocumulus and shallow convection with typically high and low cloud cover respectively.

The concept behind the combined mass-flux/K-diffusion approach is to describe the strong large-scale organized updraughts with mass fluxes and the remaining small-scale turbulent part with diffusion.

For the description of cloud within the PBL a total water variance framework was chosen with a generation term from down-gradient transports and a decay time scale of boundary layer height divided by updraught velocity.
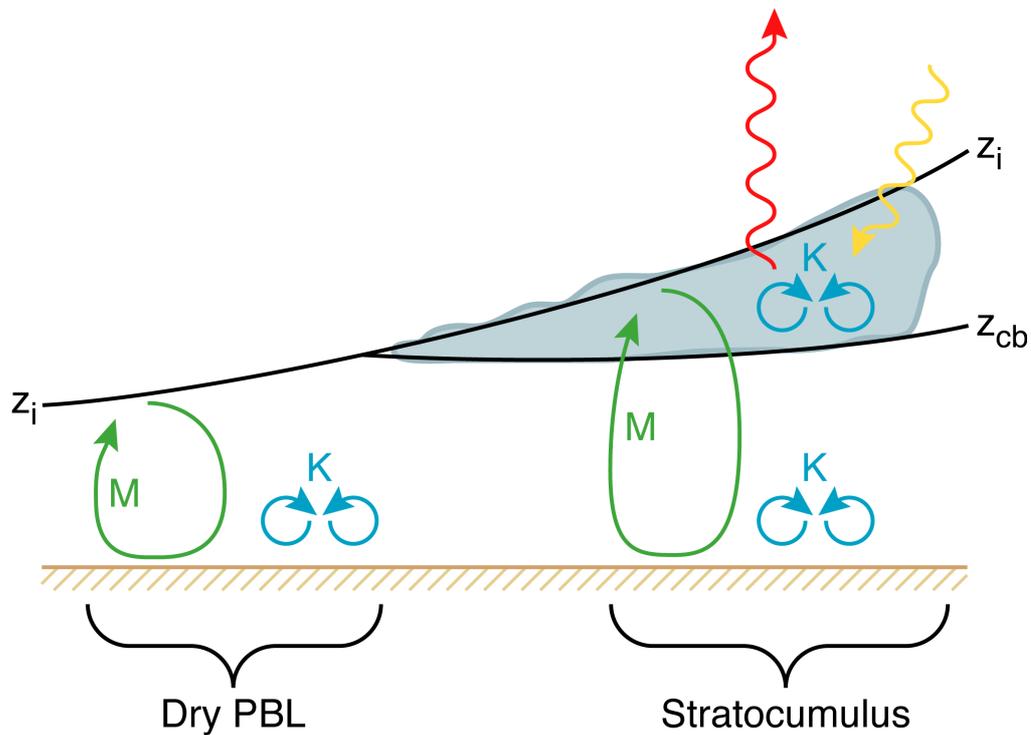
Figure 1: Transition of a dry PBL to stratocumulus. The parameterization of the associated convective transports in the new scheme are illustrated in turquoise for the diffusion component and in green for the mass flux component.

Figure 1 illustrates the transition between a dry PBL and a stratocumulus topped PBL, where the cloud base z_{cb} is above or below the inversion height z_i respectively. M denotes the PBL penetrative mass flux term. K describes the surface driven turbulent eddies, while in the stratocumulus regime there is an additional K term based on cloud top cooling initiated turbulence. Terrestrial radiative cooling and solar heating near cloud top are indicated with red and yellow waves respectively.

a Low Cloud Cover impact of new PBL
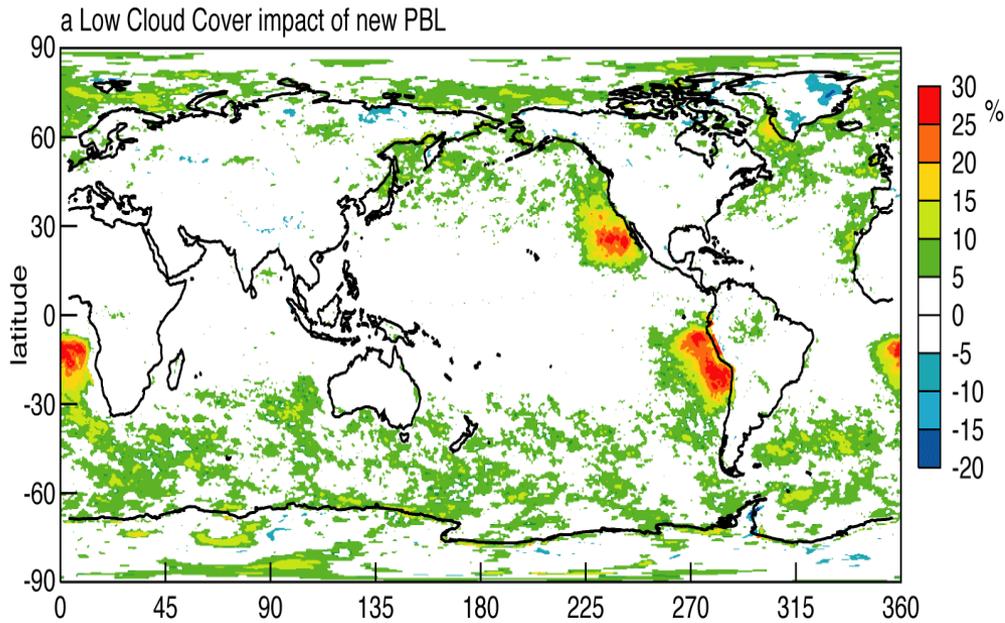
Figure 2: Low cloud cover difference for analysis/forecast experiments with and without the new PBL scheme. The figure displays the impact of the new MK-PBL scheme. 9.5 and 10 day forecasts from six test periods are used covering 7-27 July 2001, 9-22 October 2001, 1 February - 1 March 2004, 1-31 July 2004, 1-31 October 2004, 3-15 December 2004. This amounts to 140 days.
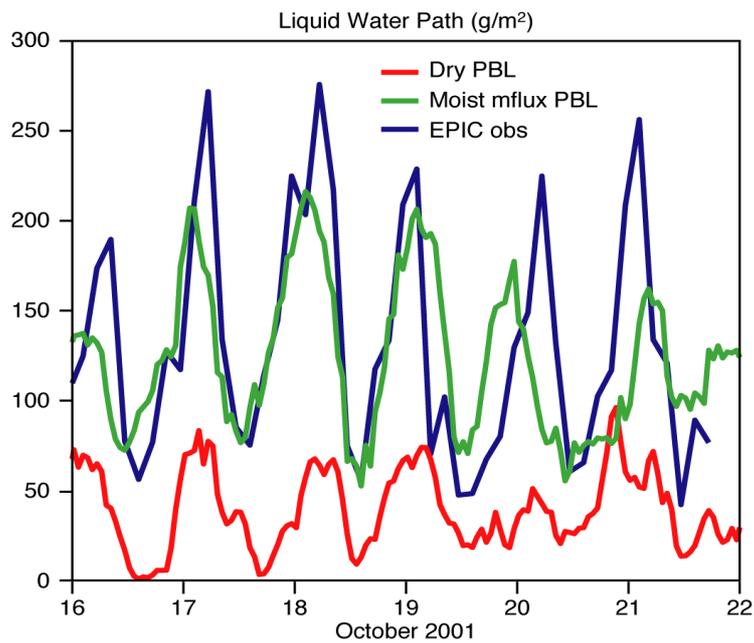
Figure 3: Liquid water path evolution during the Eastern Pacific Investigation of Climate (EPIC). It included a stationary observation period at 85West, 20 South during 16-21 October 2001. Here we show the observed liquid water path (blue) and the 3-hourly forecast data using the old (red) and the new MK-PBL (green). The day 1,2 and 3 forecasts were averaged according to verifying time to obtain a smooth curve.

3. Wavelet J_b

(M. Fisher: Background error covariance modelling. ECMWF seminar proceedings. September 2003)

The modelling and specification of the covariance matrix of background error are important elements in any data assimilation system, since it is primarily the background error covariance matrix that determines how information from observations is spread to nearby grid-points and levels of the assimilating model, that allows observations of the wind field to be used to gather information about the mass field, and *vice versa*.

The state vector of a typical analysis system for NWP has a dimension of around 10^6. Consequently, the background error covariance matrix contains roughly 10^{12} elements. This is too large to be stored in the memory of current computers.

Using the spectral technique we ca arrive at a block-diagonal matrix with one block for each total wavenumber n, and for each variable. Each block has dimension Nlevels×Nlevels, where Nlevels is the number of model levels, and represents the vertical correlation for a particular variable and wavenumber.

The advantage of the spectral approach is that it reduces the horizontal correlation matrix to a diagonal matrix. The disadvantage is that, by assuming the correlations to be equivalent to a convolution, the resulting correlations are homogeneous and isotropic.

At the opposite end of the spectrum, it is possible to specify vertical and horizontal correlations in the grid-space of the model, as a function of horizontal location. This approach allows full spatial resolution, but provides no spectral resolution. In particular, with this approach, the same vertical correlation matrix is applied, regardless of the horizontal scale of the features involved.

Clearly, a compromise between the two extremes of spatial and spectral resolution is required. It is well know that wavelet methods allow simultaneous resolution of spatial and spectral features, making them attractive for use in background covariance modelling.

Wavelet Jb defines the analysis' "change-of-variable" and the background cost function as:

$$\mathbf{x} - \mathbf{x}_b = \sum_{j=1}^{K} \mathbf{\Psi}_j \otimes \mathbf{V}_j \boldsymbol{\chi}_j \quad \text{and} \quad J_b = \frac{1}{2} \sum_{j=1}^{K} \boldsymbol{\chi}_j^{\mathrm{T}} \boldsymbol{\chi}_j$$

where
- index j denotes "scale" (wavenumber band).
- the functions $\mathbf{\Psi}_j$ are localized spatially, and in wavenumber.
- the V_j are block-diagonal, with vertical covariance matrix blocks, and one block per gridpoint.

Convolution with $\mathbf{\Psi}_j$ limits the influence of $\mathbf{V}j$ to a band of wavenumbers and to nearby gridpoints. The result is a covariance model that allows both spatial and spectral variation of covariances.

The ability of the wavelet Jb formulation to produce inhomogeneous horizontal correlations is demonstrated in Figure 5, which shows the effective horizontal structure functions for points over north America and over the equatorial Pacific. The length scale for horizontal correlation is clearly larger in the equatorial Pacific than over North America. The centre and right panels show the effective wavenumber-averaged vertical correlation matrix for a point over North America and the corresponding matrix for a point in the equatorial Pacific Ocean. Clear differences in boundary-layer structure and tropopause height are apparent.
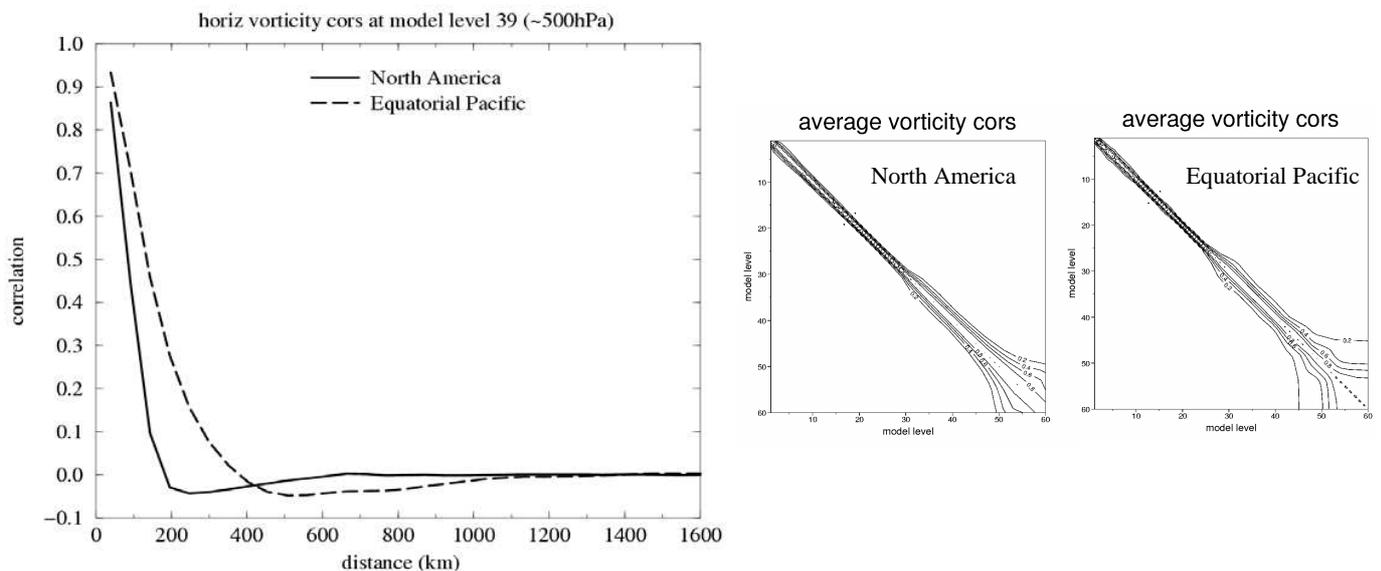
Fig 4 Effective horizontal structure functions and effective wavenumber-averaged vertical correlation matrix for vorticity for a point over North America and a point in the equatorial Pacific.

4. Increase in vertical and horizontal resolutions

Fig 6 shows the distribution of vertical levels in the old (60 level) and new (91 level) operational distributions. The new distribution is due to become operational in February 2006. The vertical resolution is increased everywhere but mainly near the tropical tropopause where it is doubled. The top of the model is also raised from 0.1 to 0.01 hPa.

The horizontal resolution will be increased at the same time from $T_L 511$ to $T_L 799$, a factor of more than 50%. That means that the number of points per level in the corresponding reduced Gaussian grid is a factor of 2.4 larger in the new resolution than in the old resolution. The largest expected increase in computational cost in the forecast model are the Legendre transforms, for which the cost increases as the cube of the spectral resolution, while the rest of the computations in the model scale with the number of grid points which is roughly proportional to the square of the spectral

truncation. The cost of the radiation parameterization will be relatively less expensive as the full radiation computations are called only every 3 hours, which means a smaller frequency in terms of number of time steps as the time step of the T_L799 will be smaller that the one used at T_L511 (12 minutes for the higher resolution and 15 minutes for the lower).
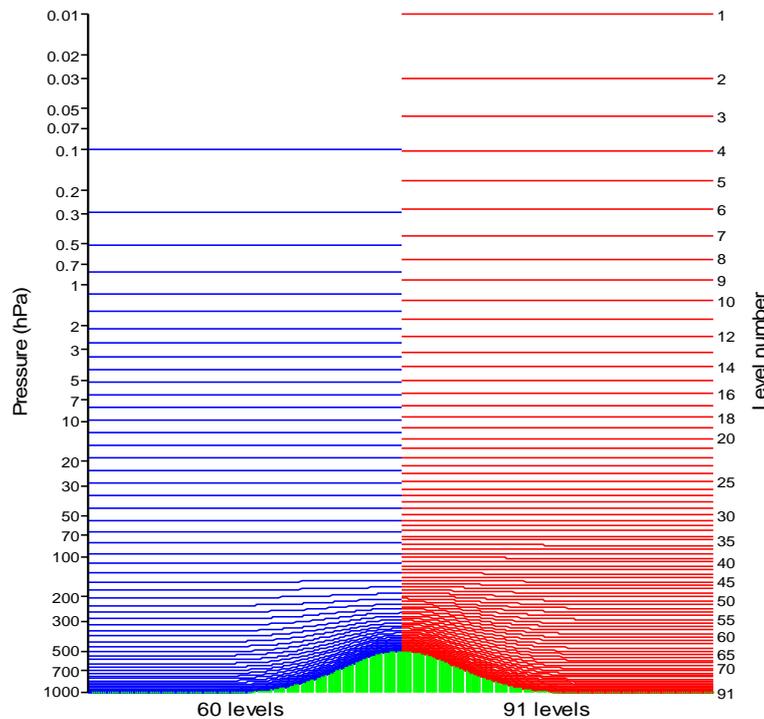
Fig. 5 Distribution of vertical levels in the old (60) and in the new (91) vertical resolutions.

Corresponding to the new horizontal resolution, the resolution in the second minimization of the 4D-Var assimilation has been increased to T_L255 from its previous value of T_L159.

After fixing some problems of dependency of the physical parameterizations with the vertical resolution and one instability in the T_L255 tangent-linear model coming from the way the reference trajectory was interpolated from the 799 to the 255 resolution, a large number of assimilation-forecast experiments have been run, in both research and pre-operational mode.

At the meeting a particular case was shown: the forecast of hurricane Katrina, which performed much better with the new than with the old resolution in all respects: position, intensity, structure and amount of rainfall.

All the mean verification scores of the set of cases run, except a warm bias in the upper tropical troposphere, which is increased with the new resolution, are significantly better in the experiments run with the new resolution that the corresponding ones using the old resolution. The new higher resolution will therefore be operationally implemented on February 1st 2006.

5. Runs at $T_L 1279$ and $T_L 2047$ resolution

The orography and other surface fields necessary to run the forecast model at a given resolution have been computed at the resolutions $T_L 1279$ and $T_L 2047$ using the reduced linear Gaussian grid. The number of points per level is 2,140,704 and 5,447,538 resp. in each of these two resolutions. A 36-hour forecast was run at these resolutions and its computational efficiency compared with the corresponding one run at the old operational (511) and new operational (799) resolutions. The results are displayed in Table 1 which shows that the ECMWF forecast model runs at a remarkable speed for a scalar computer of more that 9% of peak, except at $T_L 511$ where the start-up period is relatively more important for that short forecast.

Resolution & time-step	CPUs	Wall time for 1-day forecast (minutes)	Gflops on IBM p690+	% of Peak*
T511 L60 900 sec	128	7	81	8.4
T799 L91 720 sec	256	16	177	9.1
T1279 L60 450 sec	512	23	354	9.1
T2047 L60 300 sec	768	68	537	9.2

Table 1: Timing results in the IBM p690+ for the forecast model run at different horizontal resolutions. (*peak is 7.6 Gflops per PE for IBM p690+)
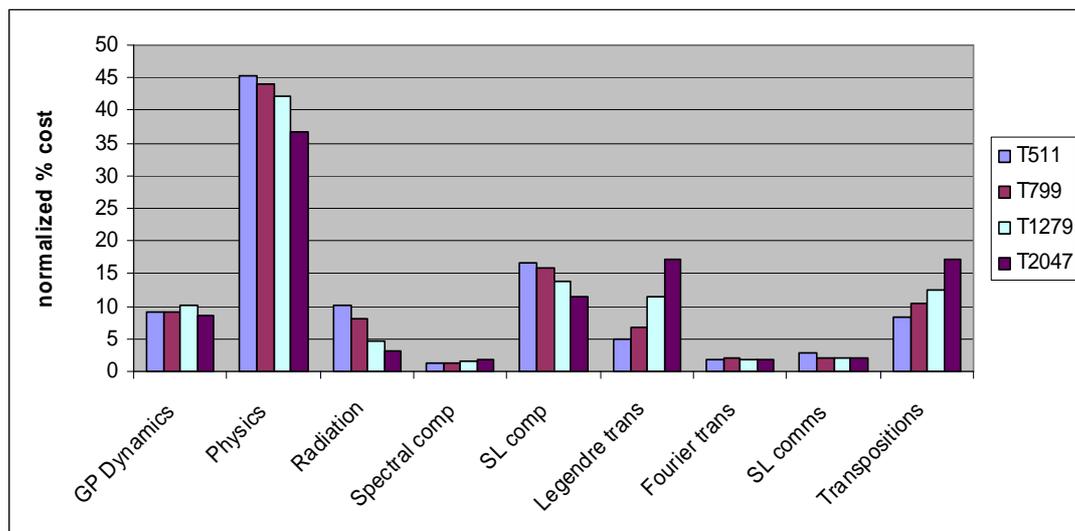
Fig 6 Distribution of the cost of the forecast model at four different horizontal resolutions

As expected, the Legendre transforms increase their relative cost with resolution but even at the highest resolution run (distance between Gaussian grid-points ~10km) they take less than 20% of the total time.

Another case was run at $T_L 2047$, starting at 7 September 2002 which corresponds to a heavy rainfall event in southern France. The correlation between the measured rainfall and the interpolated forecast to the measurement locations is shown to increase monotonically with resolution.