Modelisation of background error statistics in ALADIN, and ARPEGE

Dominique Giard and Loïk Berre
Météo-France/CNRM/GMAP
42 av Coriolis, 31057 Toulouse CEDEX 1, France
dominique.giard@meteo.fr, loik.berre@meteo.fr

A. Computing background error statistics

A.1 Introduction

Background error statistics are assumed to describe the errors induced by the forecast model, i.e. both the amplification of initial errors along forecast and the fact that the model itself is far from perfect. They are usually computed on large sets of differences between model forecasts valid at the same time.

A.2 "NMC" method

(Claude Fischer, Maria Siroka, Wafaa Sadiki, Maria Monteiro, Gergö Bölöni, Andras Horanyi)

Description

The idea is to compute differences between two forecasts issued from the same assimilation suite, but with a 24h delay between the two corresponding analyses : **[** $P_{j+1}(T) - P_j(T+\Delta t)$ **]** is computed over a large range of situations. Usually, for global models, the difference between a 36 h and a 12 h forecasts is considered, i.e. T =12 h and Δt =24 h.

For limited-area models (LAM), where the involved space and time scales are smaller and the choice of coupling is an additional degree of freedom, variants may be considered.

"Lagged" versus "standard" NMC background error statistics

The difference between the so-called "lagged" and "standard" statistics is based on 3 requirements :
- which scales are to be analysed in the LAM ?
- which coupling during the forecasts used in the computation of statistics ?
- which coupling in the assimilation process ?

Standard statistics :
- to be used when all (resolved) scales are to be analysed in the LAM
- $P_{j+1}(t)$ for ALADIN is computed using $P_{j+1}(t)$ for ARPEGE (t : 0 → 12h)
- *consistency is kept between the coupled and coupling models for each forecast, and initial differences are analysis increments that have evolved in time.*

→ a standard coupling method may be used in assimilation

Lagged statistics :
- to be used when the largest scales are considered as already analysed by the coupling model
- $P_{j+1}(t)$ for ALADIN is computed using $P_j(t$ +24 h$)$ for ARPEGE (t : 0 → 12h)
- *consistency is kept between the initial and lateral boundary conditions used by the two forevcasts (provided by the same, old, coupling forecast), and initial differences are differences between ALADIN and ARPEGE forecasts.*

→ a combination of assimilation with blending may be required to recover large scales from the coupling model

Lagged statistics are better suited to ALADIN domains, of small size and high resolution (when compared to the HIRLAM ones used for data assimilation).

Impact of the range and date of forecasts

[$P_{j+d}(T) - P_j(T+\Delta t)$ **]** was computed for ALADIN-Hungary for various combinations of T and Δt **:**

T =+6h, +12h, +18h, +24h, +30h, +36h, +42h

Δt= +6h, +12h, +18h, +24h, +30h, +36h, +42h

$T + \Delta t \leq 48\text{h}$ d=.25×*int*(Δt / 6h)

The impact of changing the range of forecasts is not so much, as illustrated hereafter. The main difference is between standard (full line) and lagged statistics, with the variance maxima shifted towards smaller scales (Fig. 1).

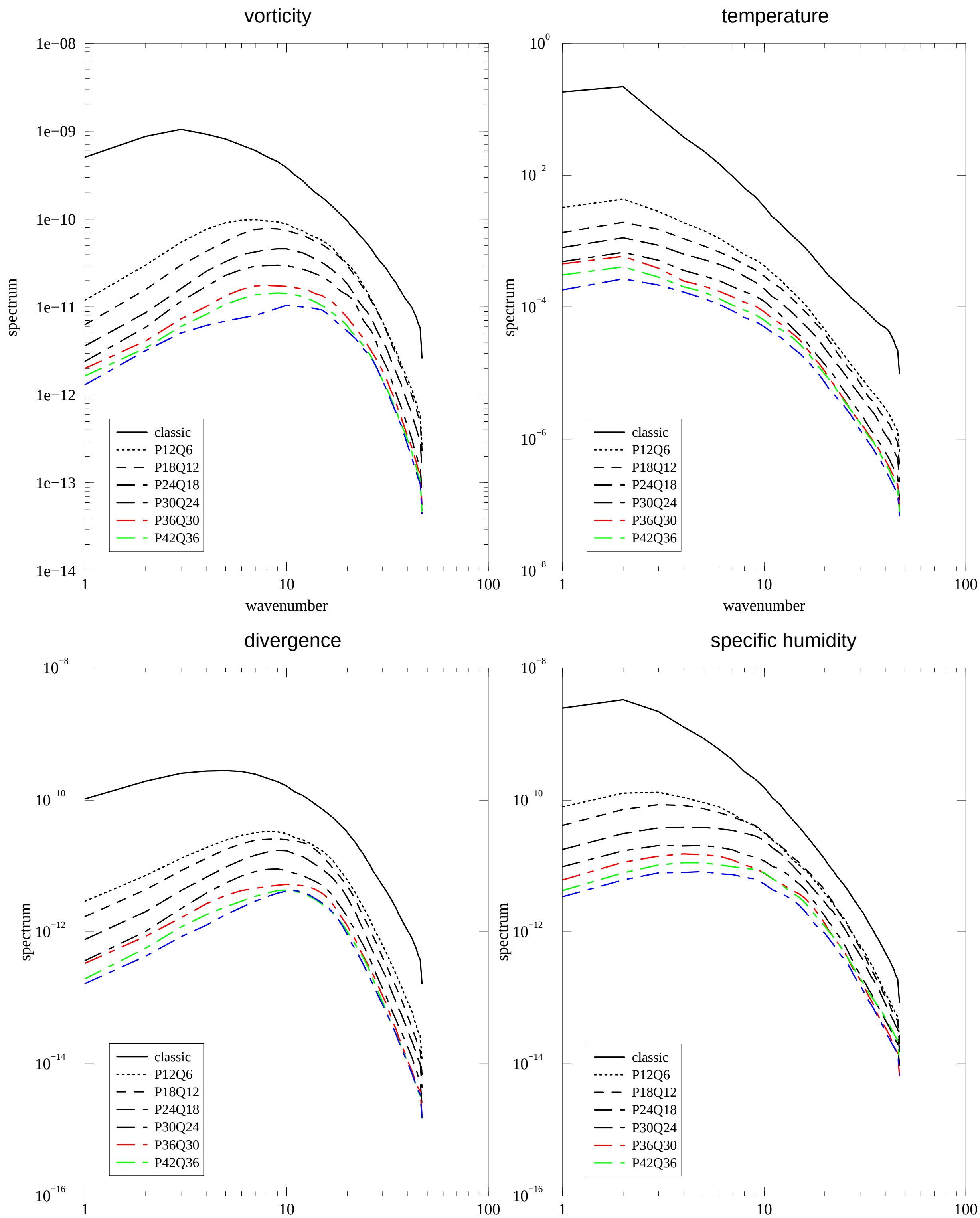


Figure 1 : Impact of the range of forecast *(T)* for a delay *(*Δt*)* of 6h between forecasts
ALADIN-HU model, variances spectra at level 13

Results for ALADIN

Lagged statistics were computed for several ALADIN domains, of various location, size and resolution : LACE, France, Hungary, Morocco, NORAF. But they show rather little sensitivity, both to the domain and the range of forecasts. The only exception is for ALADIN-NORAF, covering a very large domain (SW corner : 0°, -36° ; NE corner : 43°, 56°) with a rather low resolution (31 km), where the impact of the Hadley circulation appears on *T-Div* and *q-Div* correlations.

A.3 "Analysis ensemble" method

(Loïk Berre, Margarida Belo Pereira)

Principle

Independent assimilation experiments, called elements, are run with a 6 h cycle. For each element and at each analysis step, observations are perturbed randomly. The perturbation is chosen of Gaussian shape, with a zero mean and a standard deviation equal to that chosen for observation errors in the assimilation scheme.

Errors propagate along the subsequent 6 h forecasts : $e_b(t+6\text{h}) = M{\cdot}e_a(t)$. After a few days, differences between the 6 h forecasts (background) from independent elements are computed, by pairs. One obtains the background error : e_b but also the analysis error : $e_a = e_b + K\,(e_o - H{\cdot}e_b)$, from the observation errors e_o, the model M, the observation operator H and the gain matrix K used in the computation of elements.

This takes into account only the amplification / propagation of errors by the model. To include modelling errors, several models may be combined.

Experiments

Experiments were performed only with the global model ARPEGE so far, which avoids considering the numerous coupling issues. First 3d-var then 4d-var assimilation was used. Three datasets were evaluated :
- using only ARPEGE-Métropole (stretched and tilted grid),
- using only ARPEGE-Tropiques (regular grid),
- using differences between these two models, once "post-processed" on the same intermediate regular grid,

and the first charts of variance and length-scales plotted (Fig. 3). Averages are computed over one month, for each assimilation cycle (00, 06, 12, 18 h UTC)

First results

The standard deviations are lower than those obtained with the NMC method, especially for the stretched geometry (with a ratio of 2 instead of 1.4). They increase when combining the two models, i.e. when considering also the impact of resolution on background errors(Fig. 2).

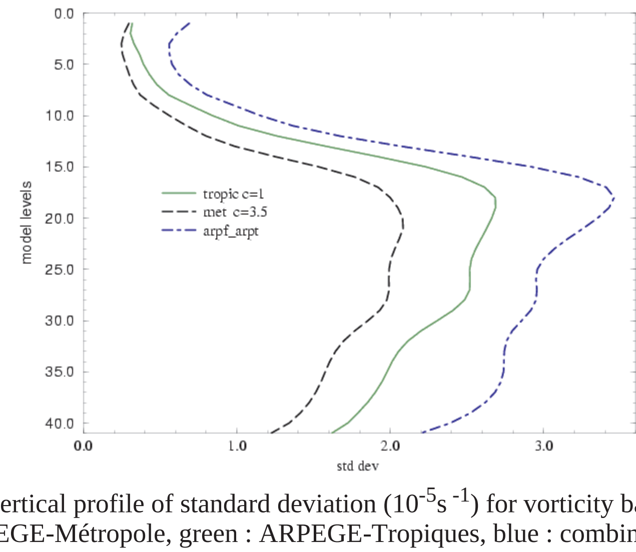


Figure 2 : Vertical profile of standard deviation ($10^{-5}s^{-1}$) for vorticity background error
(black : ARPEGE-Métropole, green : ARPEGE-Tropiques, blue : combining both models)

Compared with the NMC method, the ensemble method provides smaller scale statistics, which is visible in the error variance spectra and in the (larger) degree of ageostrophy (Fig. 3). It seems possible to study the respective effects of the analysis and of the atmospheric processes on the time evolution of the errors.

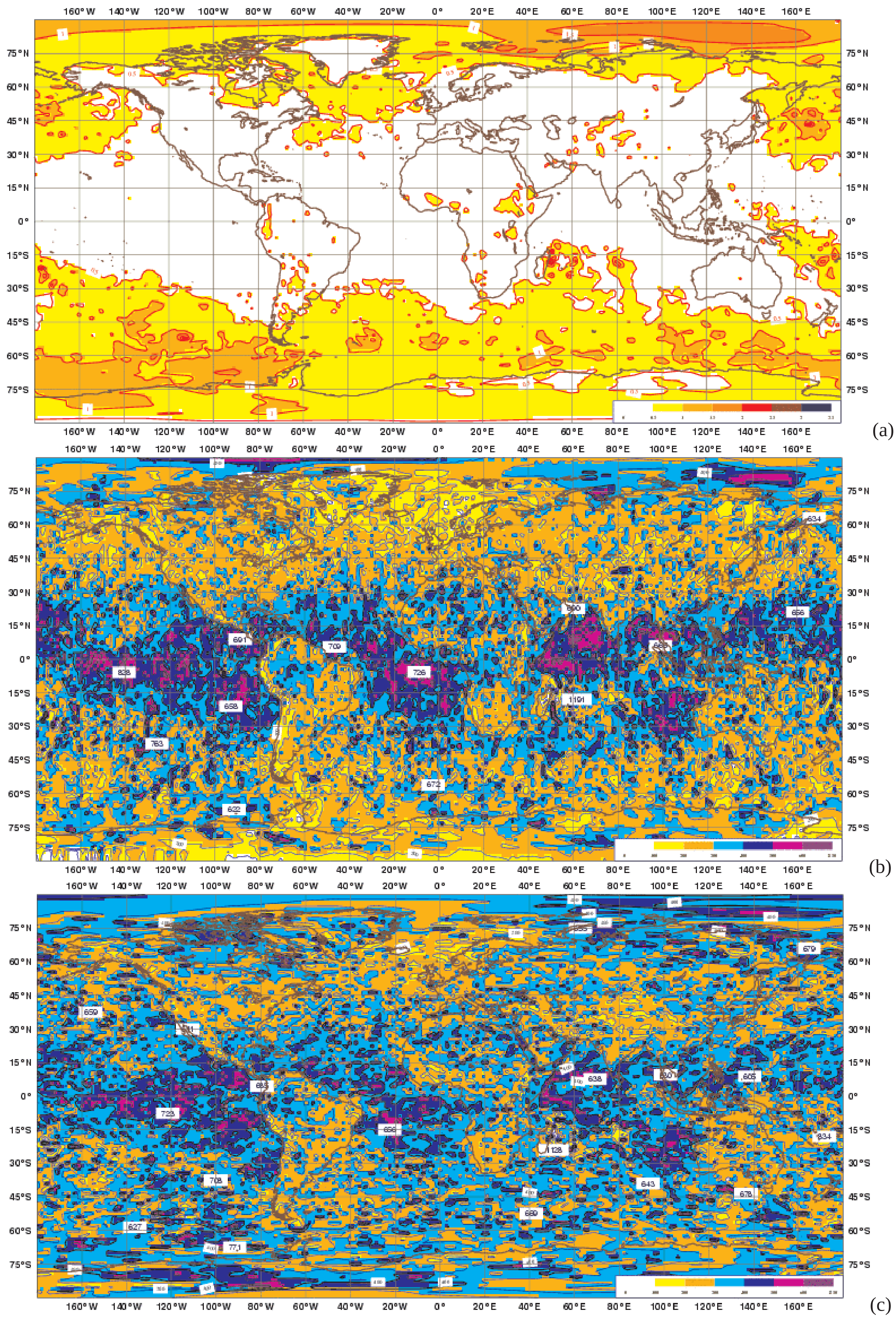


Figure 3 : Background error statistics for surface pressure in ARPEGE-Tropiques, estimated with an ensemble approach (a) standard deviation (hPa), (b) zonal length-scale (km), (c) meridional length-scale (km).

It is possible to use a small number of ensemble members (e.g. two !), as in the NMC method, in order to estimate the time and space averaged covariances in a cheap way, with the full operational model. Conversely, the ensemble approach offers the possibility to use a larger number of members, if we want to study in the future flow-dependent features, possibly using a simplified version of the assimilation and model system.

The first impact studies (two 15-day assimilation experiments) indicate a positive impact, over Europe and the Tropics, of the ensemble covariances, compared with those provided by the NMC method.

B. Diagnostic studies

(Loïk Berre, Mohammed Raouindi, Simona Stefanescu, Alex Deckmyn, Margarida Belo Pereira)

B.1 Diagnosis of horizontal (in latitude or longitude) variations

Several tools may be used to evaluate the properties of background error statistics (horizontal scales, isotropy, ...) and their representation by different modelisations of the B-matrix. The following diagnostics are used for ALADIN, and ARPEGE whenever possible :

- Estimation of standard-deviation (σ) or variance (σ^2) maps **in gridpoint space**

$$\sigma^2(X,Y)(T) = \overline{\Delta T}^{\,2}(X,Y) = \frac{1}{N}\sum_{t=1}^{N} \Delta T^{\,2}(X,Y,\ t)$$

- Estimation of (e.g. meridional) length-scale maps **in gridpoint space**,
  <u>assuming</u> that the local auto-correlation function is flat at the origin (dx=dy=0), which can be expected knowing that the correlation decreases with distance on both sides of the origin, and considering that the function is continuous at the origin.

$$L_y^{\,2}(X,Y)(T) = \sigma^2(X,Y)(T)\ /\ \left[\sigma^2(X,Y)\left(\frac{\partial T}{\partial y}\right) - \left(\frac{\partial \sigma}{\partial y}(X,Y)(T)\right)^2\right]$$

- Estimation of variance maps **in spectral space for a LAM** (bi-Fourier transforms)

$$\sigma^2(X,Y)(T) = \sum_m \sum_n A_{m,n} e^{i2\pi(mX+nY)} \quad \text{with} \quad A_{m,n} = \sum_{m'} \sum_{n'} \overline{\Delta T_{m+m',\ n+n'}\ \Delta T^*_{m',\ n'}}$$

- Estimation of (e.g. meridional) length-scale maps **in spectral space, for a LAM**,
  <u>WITHOUT assuming</u> that the local auto-correlation function is flat at the origin.

$$L_y^{\,2}(X,Y)(T) = -\sigma^2(X,Y)(T)\ /\ \left[\frac{\partial^2 cov}{\partial y^2}(X,Y)(T)\right]_{(dx=dy=0)}$$

$$\text{with} \quad \left[\frac{\partial^2 cov}{\partial y^2}(X,Y)(T)\right]_0 = \sum_m \sum_n B_{m,n} e^{i2\pi(mX+nY)} \quad \text{and} \quad B_{m,n} = -\sum_{m'} \sum_{n'} \overline{n'^2 T_{m+m',\ n+n'} T^*_{m',n'}}$$

The next step is the computation of the preferred direction for length-scales.

B.2 Analysis of charts of standard deviation and length-scales

Charts of standard deviations and length scales have been computed in gridpoint space for ARPEGE (Métropole and Tropiques), to evaluate the "analysis ensemble" method (Fig. 3).

In ALADIN, computations were performed using a version of ALADIN-Morocco with a rather low resolution (25 km), in order to evaluate :

- the variations in latitude and longitude, and the validity of the frequently used hypothesis of isotropy (here using standard NMC statistics) (Fig. 4);
- the impact of the approximations made to simplify gridpoint computations (comparing fields computed in gridpoint and in spectral spaces) (Fig. 5),
- the skill of new Jb formulations (Fig. 7).

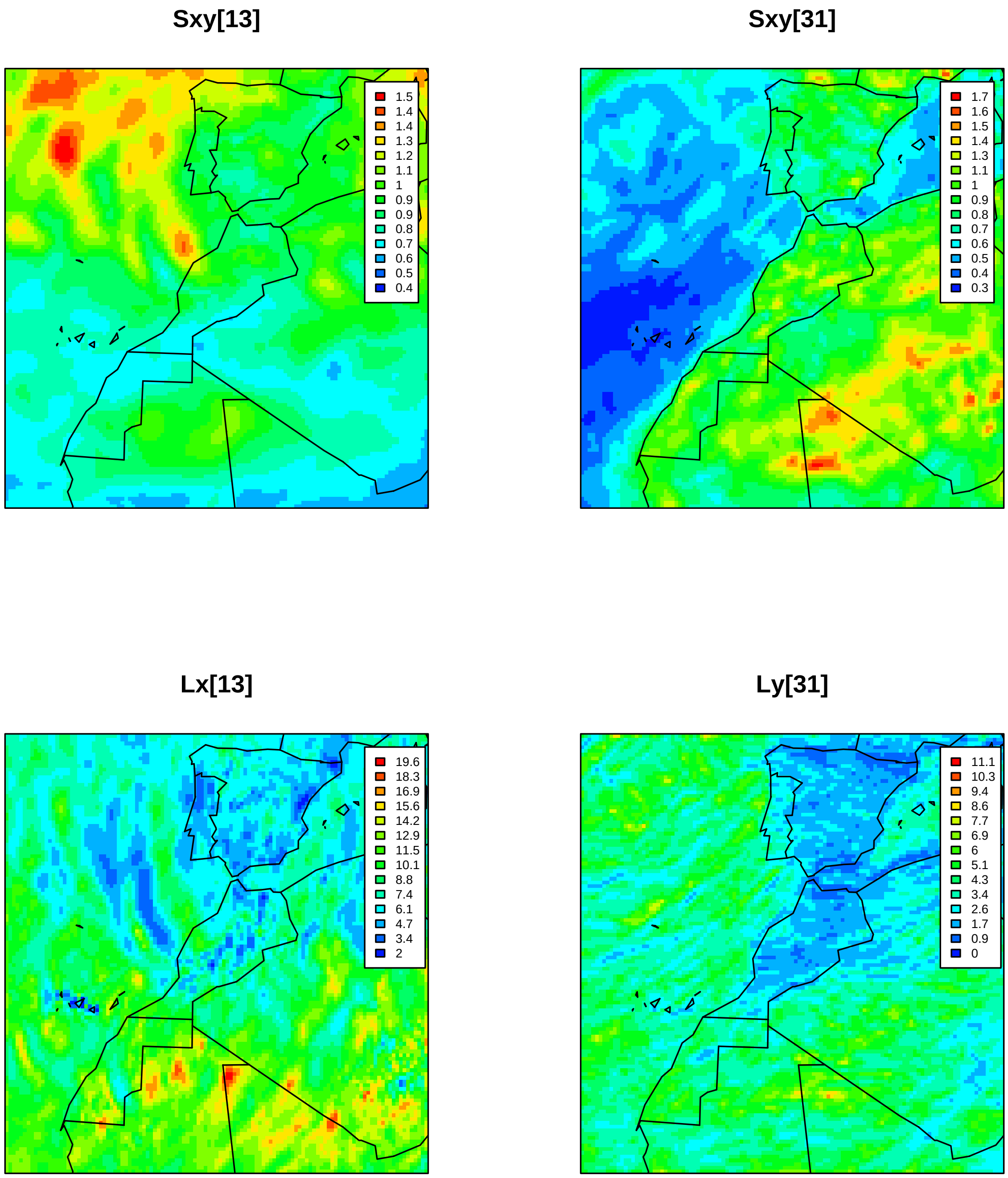


Figure 4 : Gridpoint estimations for ALADIN-Morocco :
Sxy[13]: standard deviations at level 13 ; Sxy[31]: standard deviations at level 31 ;
Lx[13]: zonal length-scales at level 13 ; Ly[31]: meridional length-scales at level 31.

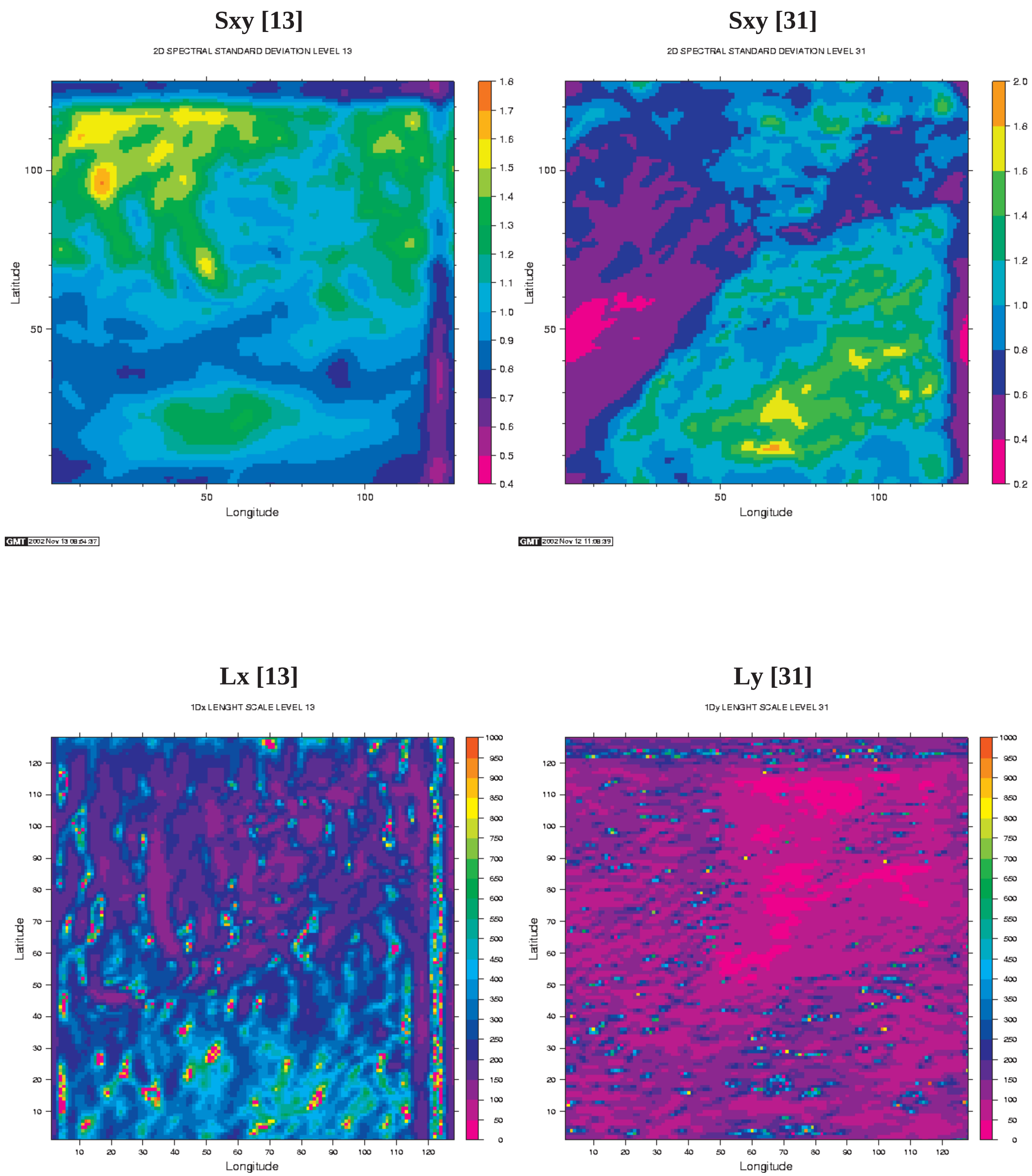


Figure 5 : Spectral estimations for ALADIN-Morocco :
Sxy[13]: standard deviations at level 13 ; Sxy[31]: standard deviations at level 31 ;
Lx[13]: zonal length-scales at level 13 ; Ly[31]: meridional length-scales at level 31.

B.3 Summary of the diagnostic results

The variance and length-scale maps are simple tools to explore the ensemble covariances, and evaluate the properties of different Jb formulations. The calculation is relatively easy both in gridpoint and spectral space.

The agreement between the length-scale maps derived from the gridpoint and spectral approaches suggests that the assumption of flatness of the correlation function at the origin is rather good.

The predominant horizontal variations may be summarized as : the free troposhere is much influenced by the tropics/mid-latitude contrast, while the low levels are more influenced by the land/sea contrast (and/or the density of observations).

C. Modelling the background cost-function (B matrix)

C.1 From a diagonal to a block-diagonal (2d) formulation

(Mohammed Raouindi, Simona Stefanescu, Loïk Berre)

The validity of the hypothesis of isotropy, which enables the use of a simple 2d-diagonal **B** matrix in the formulation of the Jb cost function, was reconsidered, evaluating off-diagonal meridional (resp. zonal) terms, e.g. computing :

$$\overline{\Delta T_{nm} \times \Delta T^{*}_{n'm}}^{(m)}$$

Such an evaluation is rather easy in ALADIN, because of the spectral bi-Fourier representation of fields. Tests were performed for the ALADIN-Morocco domain mentioned previously, examining first the variations with latitude. The impact of truncations (for *n-n'*) from 5 to 42 (maximum, i.e. using the full matrix) was evaluated. Figure 6 shows the impact of truncation on the mean zonal correlation function :

$$cor\,(y,z) = (\ \mathbf{B}_{gp} \cdot \delta_{xy} = (sp \rightarrow gp) \cdot \mathbf{B}_{sp} \cdot (gp \rightarrow sp) \cdot \delta_{xy}\ ), \text{ averaged over x.}$$

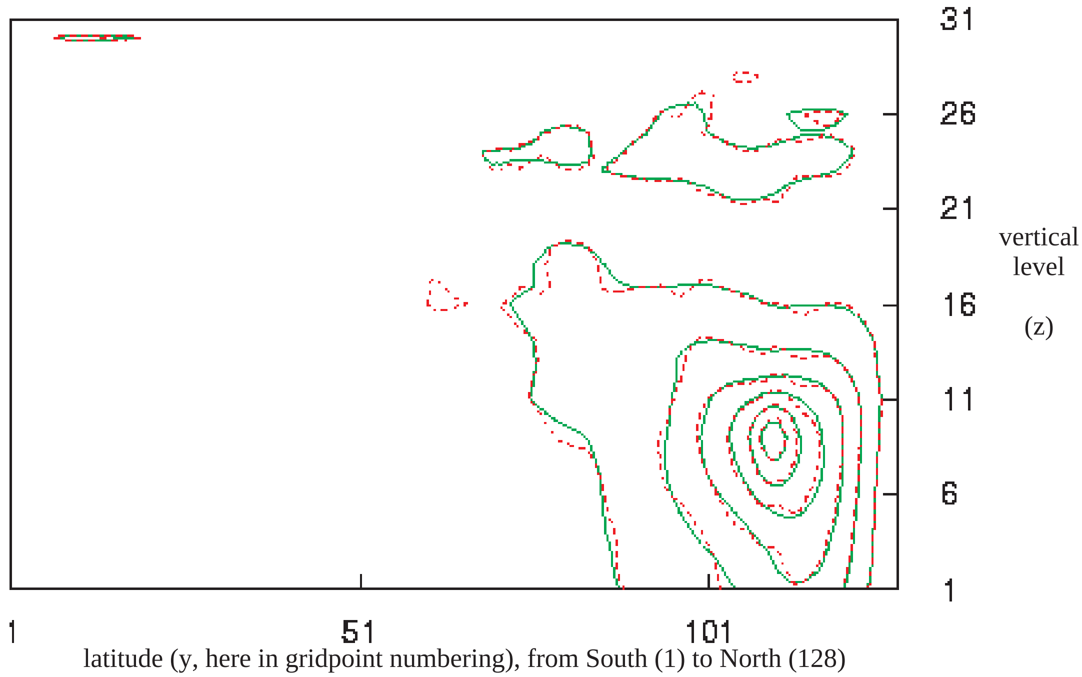


Figure 6 : Description of the temperature correlation function for ALADIN-Morocco. The extension zone id 11-points wide (118-128). Red, dotted, line : full **B** matrix ; green, full, line : block-diagonal **B** matrix with truncation 10 for wavenumber difference. First isoline : 0.85, distance between isolignes : 0.15.

The main features to be underlined are :

- the importance of meridional variations, with dependency on height,
- the quick decay of correlations in the extension zone,
- a quite good representation of the main variations using a small (10) truncation.

Zonal variations were also investigated, using 1d ($\Delta T_{m,n} \rightarrow \Delta T_{m,y}$, as for the first investigations of meridional dependencies), but also more general 2d spectral transforms of the forecast differences.

C.2 Wavelet representation

(Alex Deckmyn, Loïk Berre)

Method

Forecast errors ($\Delta T_{m,n}$) are projected and statistics computed in a wavelet space, where the **B** matrix is assumed to be diagonal.

Wavelets are considered since they allow a "position × scale" representation. Meyer's functions are used in ALADIN, because of their similarity with a Fourier representation. Alex Deckmyn extended their formulation to any number of gridpoints (i.e. even not of the form 2^N), which allows their use for any domain.

First results

The first experiments were performed using the same ALADIN-Morocco domain and standard NMC statistics (i.e; including all the scales described by the model). Figure 7 shows the geographical variations of the resulting variances and length-scales.

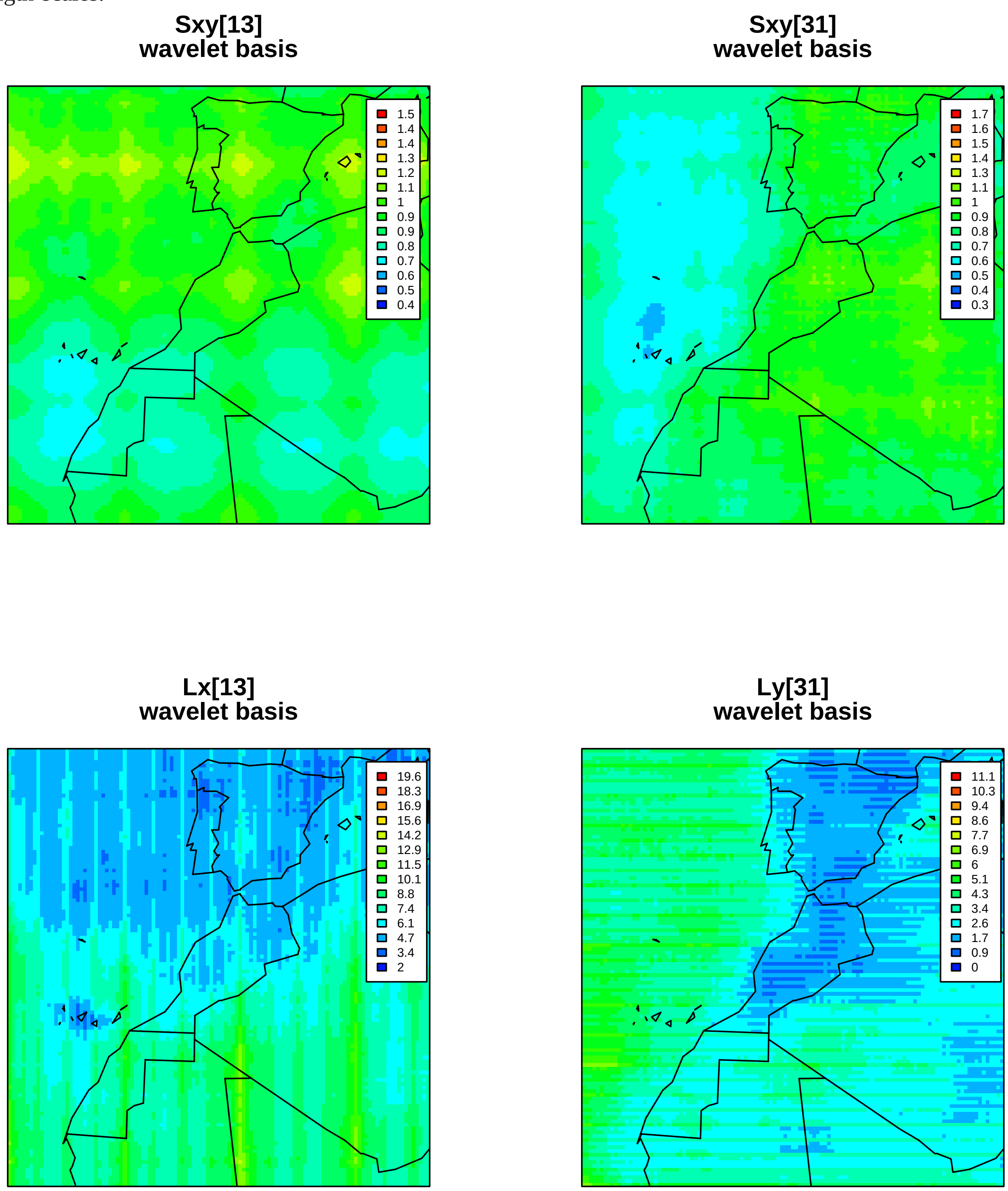


Figure 7 : Same as Figure 4, but when using a diagonal wavelet representation

Here are some possible early lessons of the wavelet experimentation :

- This method is simple and not expensive.
- A wavelet diagonal approach seems to allow the representation of some geographical variations in the variances and in the length-scales of the correlations.
- An influence of the wavelet discretization grid on the implied variance maps could be diagnosed, in addition to a tendency to underestimate on the average the length-scales.
- These two features may indicate a possible influence of the diagonal assumption (in the wavelet space) on the (imperfect) representation of some (more or less) large-scale correlations. However this problem may be alleviated when considering lagged NMC statistics.
- This interpretation tends to be supported by the better behaviour of the wavelet-implied estimates at low levels, where small-scale features become relatively more important.
- There are fewer details in the wavelet-implied variance and length-scale maps: the extreme values tend to be smoothed. This may be consistent with the implicit representation of cross-correlations between "close" wave numbers (i.e. which contribute to the same wavelet mode). It tends to suggest that the main potential of the wavelets may be mostly the representation of some "smooth/large-scale" variations of the small-scale statistics.

C.3 Biperiodicity problems

(Claude Fischer, Vincent Guidard, Loïk Berre, Gergö Bölöni)

Problems were exhibited in the framework of ALADIN-France 3d-var : use of "lagged" NMC statistics, assumption of horizontal homogeneity and isotropy, multivariate approach, and a large (~ 3000 km wide) domain with a resolution around 10 km.

Figure 8 illustrates what can happen when there are observations only over a part of the domain. In this experiment, a wide (1000 km) band of observations in the Southern part of the domain is used. The 3d-var analysis generates strong increments where observations are available, but the corrections don’t vanish within the thin (~ 100 km wide) extension zone, and "re-enter" the domain through the Northern boundary.

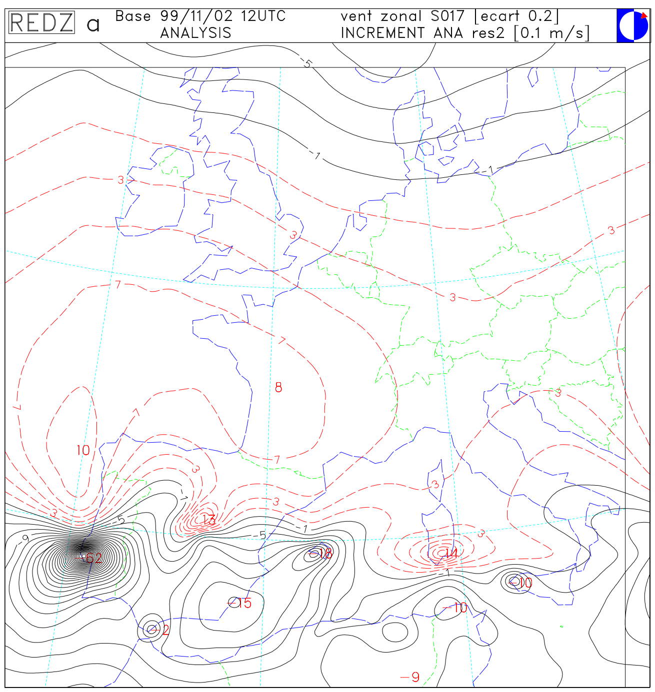
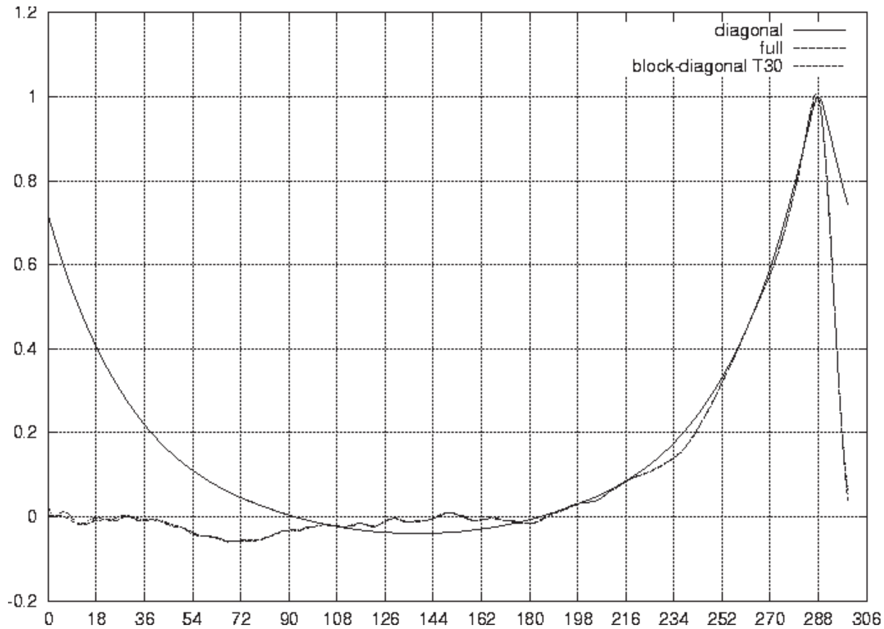


Figure 8 : 3d-var analysis increments for zonal wind at level 17 in ALADIN-France, for an experiment using observations only in the Southern part of the domain. Isolignes every 0.2 m/s.

Several cures were considered :

- rejecting the observations which are too close to the border ⇒ NO SUCCESS !
- zeroing the σ_b in a rim zone around the domain ⇒ NO SUCCESS !
- increasing the extension zone (hence the cost of the model !) ⇒ TO BE TESTED

• use of a block-diagonal approach, i.e. no longer assuming horizontal isotropy, at least because of the impact of lateral boundaries (see Fig. 9)) ⇒ PROMISING !



latitude (here in gridpoint numbering), from South (left) to North (right)

Figure 9 : Description of the temperature correlation function for ALADIN-France using different **B** matrix. Full line : diagonal, dashed line : full, dotted : block-diagonal (truncation 30 for wavenumber difference).

When the assumption of isotropy is relaxed, lateral values at the border of the C+I domain are preserved, and biperiodicity is ensured through an adjustment within the extension zone only.

• use of compactly supported functions to describe correlations ⇒ IN TEST

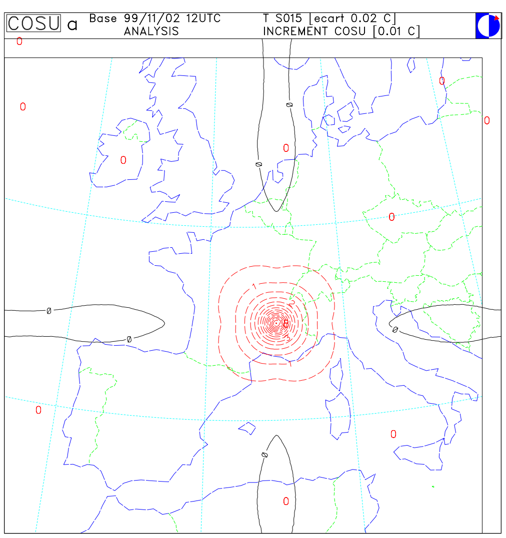
• combinations of solutions ...

<u>C.4 Use of compactly supported functions</u>

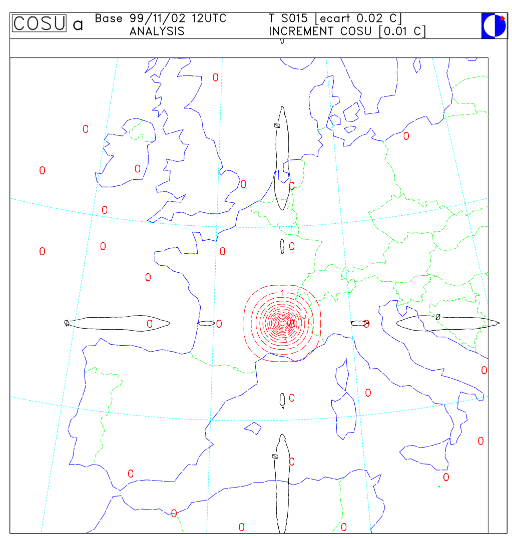
(Claude Fischer, Vincent Guidard, Loïk Berre)

The idea is to assume that auto-correlations and cross-correlations are close to zero at some distance of the observation. A bi-Fourier procedure that enforces a compact support for the correlation functions (smoothly zeroing from a given distance) has been implemented, and the first impact studies started.

The first tests were performed in a simplified framework : ALADIN-France, univariate approach, single observation of temperature (at 500 hPa), evaluation of the impact on the temperature increment at model level 15.



NOT compactly supported autocorrelations



compactly supported autocorrelations