Data Assimilation by Field Alignment, Testing the Theory

Carlos Geijo, AEMET.

1 Introduction

This HIRLAM Newsletters contribution presents first tests done with a position errors correction algorithm. The algorithm, which appeared in the literature a few years ago[1] (REM hence forth), tackles the problem of shifts in position between features in model fields and observations by enlarging the analysis control space with a set of smooth 2D vector fields which give the displacements of the model fields which minimize the position errors. This adding of new control variables in the DA problem is formulated in terms of Bayesian theory, and it leads to the definition of a new "cost function" or "objective" which represents explicitly both amplitude and position errors. This scheme can work with deterministic or ensemble approaches alike. The tests presented in this communication have however been conducted only in a deterministic context with the mesoscale NWP system HARMONIE . The tests used model fields as surrogates of observations, and this renders the results, which in themselves are satisfactory, only tentative.

2 Brief description of the method

The detrimental impact of position errors on an analysis cannot be overemphasized. They introduce biases which violate important assumptions in statistical interpolation methods. It is also very complicated to account for this kind of errors with commonly used variational methods (3D-Var), even considering an ensemble approach (REM), because the sources of position errors are multiple and difficult to characterize. The key step in the method described here is to explicitly represent and minimize position errors. This is achieved by enlarging the set of control variables {X} to include the displacements **{ q }** that correct for position errors. The new set of control variables then becomes **{X,q}** with **q** defined at each node of the grid.

Similarly to ordinary 3D-Var, we can express the inference of the model state conditioned to observations **Y** as (up to a normalization) : $P(\mathbf{X,q} \mid \mathbf{Y}) \alpha P(\mathbf{Y} \mid \mathbf{X,q}) P(\mathbf{X^f} \mid \mathbf{q}) P(\mathbf{q})$, where the first factor is known as "data likelihood", the second is the "amplitude prior" and the third is the "displacement prior".

The $P(Xf \mid q)$ term represents the forecast error distribution conditioned to the displacement. It is assumed that this pdf is gaussian in the position-corrected space, even if it isn´t in the uncorrected space. This statement is expressed as (eq 11 in REM):

$$P(\mathbf{X^f} \mid \mathbf{q}) \alpha \det(B(\mathbf{q}))^{-1/2} \exp -1/2 \{ (\mathbf{X(r-q)- X^f (r-q)})^T B(\mathbf{q})^{-1} (\mathbf{X(r-q)- Xf (r-q)}) \}$$

where the state **X(r-q)** and the forecast or first-guess $\mathbf{X^f (r-q)}$ are defined on the deformed grid (hence the **r-q** notation) where the position errors have been removed. The notation B(**q**) underlines that the

[1] Ravela,S ;Emanuel K. and McLaughlin D. Physica D 230 (2007) 127-145 available online at www.sciencedirect.com

[2] *Hirlam – Aladin Research on mesoscale MOdelling Nwp In Europe,* HARMONIE system is the non-hydrostatic mesoscale model code and system ensuing from the collaboration of ALADIN and HIRLAM consortia within the IFS framework

forecast error covariance B on the original grid has to be remapped to the deformed grid and then the second moment of P($\mathbf{X}^f|\mathbf{q}$) must clearly depend on the deformation applied. The other "prior" P($\mathbf{q}$) brings in constraints imposed on the q field, expected properties it must satisfy. The choice in REM is P($\mathbf{q}$) α exp { – L($\mathbf{q}$) } where

$2\ L(\mathbf{q}) = w_1 \sum_j | \text{grad}\ \mathbf{q} |_j^2 + w_2 \sum_j (\text{div}\ \mathbf{q})_j^2$; j $\varepsilon\ \Omega$, Ω computational grid

and w_1 and w_2 are free parameters that weight the gradient and the divergence penalty terms, respectively. This is a smoothness constraint. Finally, the "data likelihood" term, under the Gaussian assumption, takes the 3D-Var form but with the proviso that the model state is defined on the deformed grid.

With these definitions of probabilities we arrive at a "field alignment cost function" JFA (eq 14 in REM) whose minimization is difficult. The dependence of B on $\mathbf{q}$ makes it particularly awkward to solve.

To overcome these difficulties, REM propose two approaches. The first is named "one-step algorithm" and it requires an ensemble of forecast states. The name refers to the fact that the minimization proceeds in the whole control space **{X,q}**, that is, amplitudes and positions are assimilated simultaneously. This scheme however would scale poorly with ensemble size (REM). The second approach is the "sequential solution". This approach can function in ensemble and in deterministic systems alike. The name refers to the fact that the stationary point for J_{FA} (to be more precise an approximation to it (J_{FA}^*) where the dependency of B on $\mathbf{q}$ is ignored when computing grad JFA and in this way avoids the computation of derivatives of higher order terms) is reached by solving first the "alignment equation":

$\text{grad}_{\mathbf{q}}\ J_{FA}^*{}_{|\mathbf{X}=\mathbf{X}^f} = 0$ which is solved for $\mathbf{q}$, with solution $\mathbf{Q}$

and second, the "amplitude equation":

$\text{grad}_{\mathbf{X}}\ J_{FA}^*{}_{|\mathbf{q}=\mathbf{Q}} = 0$ which is solved for $\mathbf{X}$

In the ensemble framework, once the members have been aligned, they can be used to compute a covariance B_Q which can be utilized after EnKF or alternative schemes such as "square-root formulations" (REM). In the deterministic framework, aligned first-guess fields can be used to craft covariance error matrices much in the same way as is done with uncorrected fields.

One important observation is that the "alignment equation" does not depend on B. This is because $\mathbf{X}$ is fixed to $\mathbf{X}^f$ when solving it for $\mathbf{q}$. With the choice for L indicated above, this equation reads:

$w_1 [\ \Delta\mathbf{q}\ ]_j + w_2 [\ \text{grad}\ (\text{div}\ \mathbf{q})\ ]_j = [\ (\text{grad}\ \mathbf{X}^f{}_{\mathbf{r-q}})^T\ H^T\ R^{-1}\ (\ \mathbf{Y} - H\ \mathbf{X}^f{}_{\mathbf{r-q}})\ ]_j$; j = x,y

with R the obs error covariance and H the obs operator. The rhs forcing term (which stems from the "data likelihood" P($\mathbf{Y}$ | $\mathbf{X,q}$) factor) must be calculated on the deformed grid, and this makes the equation non-linear, to be solved iteratively with adequate boundary conditions. It is found that spectral methods in an extended domain work well for $\mathbf{q}_{\mathbf{n}} = 0$ as LBCs. Another observation is that there is not a natural length scale for $\mathbf{q}$. It is a free parameter together with the relative weight of the gradient and divergence penalties (w_1/w_2). In practice this scale must be adjusted to avoid displacements out of the computational grid, or in other words, deformations of the domain boundaries.

3 Description of the Tests

3.1 Preparation of the data

The method has been tested in an idealized situation. To what extent the results obtained are relevant to more realistic situations is a question left for future work. We start out from meteorological fields generated with HARMONIE 36h1.3 running at 2.5 km horizontal resolution and 60 hybrid levels ("HIRLAM levels") over a domain which contains the Iberian Peninsula, Balearic Islands and nearby territories (see fig 1). One set of fields ("the truth") is the ECMWF +6H forecast valid for 28 January 2011 at 00 UTC interpolated to the "IBERIA_2.5" domain by using the software that prepares the initial conditions for a standard HARMONIE-AROME run without surface or upper-air analysis ("downscaling run"). Another set of fields is the ECMWF +15H forecast valid for 27 January 2011 at 21 UTC. These fields are interpolated to the "IBERIA_2.5" domain in the same way. All the necessary fields were converted to grid point space by running HARMONIE in "full-pos" mode.

The differences between the fields of the first set and the second set, considered both as valid for the same time (28th 00 UTC), provide us with a convenient sample of "errors" that include position errors arising (mainly, but not only) from timing errors. Note that the two sets correspond to different IFS cycles (27th 06 UTC and 27th 18 UTC respectively). Maybe it is opportune to mention here that some authors list among the sources of position errors: inadequate data and model resolution, incorrect model physics, and simplifying dynamical/statistical assumptions (see REM for more info).

The mslp field (fig 1) is dominated by a stationary high on the western part of the domain and by an eastwards moving and deepening low on the eastern part. The experiment will focus on this eastern area (here forth AoI "area of interest"), which is defined 256 x 256 grid points big (the whole "IBERIA_2.5 domain is 576 x 480 points, EZ included) and centred over the low in question. The mean pressure tendency during 3 hours here reaches values of about 1hPa / hour. The wind field at low levels (level 55) (fig 3) displays a clear vortex that drifts a distance of about 60 km north-east in the course of three hours and at the same time deepens about 3hPa and becomes more compact. The wind speed differences in the AoI (level 55) (fig 4 left) range between +13 m/s and - 4 m/s, reflecting the intensification of the low (differences are "obs – fg", i.e., 28th 00 UTC – 27th 21 UTC). The differences in the T and q (specific humidity, not to interpret as the **q** in the previous section) fields suggest a southwest-northeast wave propagation pattern (fig 5 left and fig 6 left, respectively). Values (level 60) are in the range [+2.5, -2] K for T and [+1, -1.2] gr/Kgr for q.

It is convenient to take a look too at the results at higher levels. We quickly realize that the difference fields bears no similarity with that seen lower levels. For instance, for the wind speed difference at level 20 the values are between +10 and -10 m/s approx, more evenly scattered over the area, and there is no trace of the vortex in the wind fields (not shown). We realize too that the selection of AoI for the lower levels might be not adequate for the higher levels. In these tests this area has been held fixed for all levels.

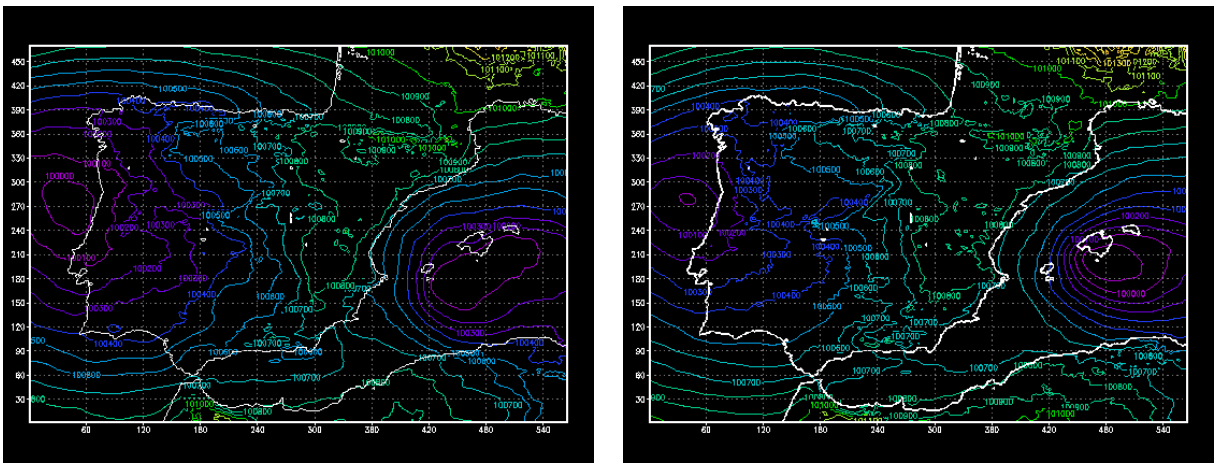

Figure 1: mslp fields over the HARMONIE IBERIA_2.5 domain used in these tests. They are ECMWF forecasts interpolated at 2.5 km. On the left we have a +15H forecast with valid time 27th January 2011 at 21 UTC. On the right a +6H forecast valid for 28th January at 00 UTC is displayed. The difference therefore corresponds to 3 hours evolution of the mslp parameter or, if taken both as valid for the same time, to a timing error of 3 hours. The mean pressure tendency over the area of interest is about 1 hPa / hour.

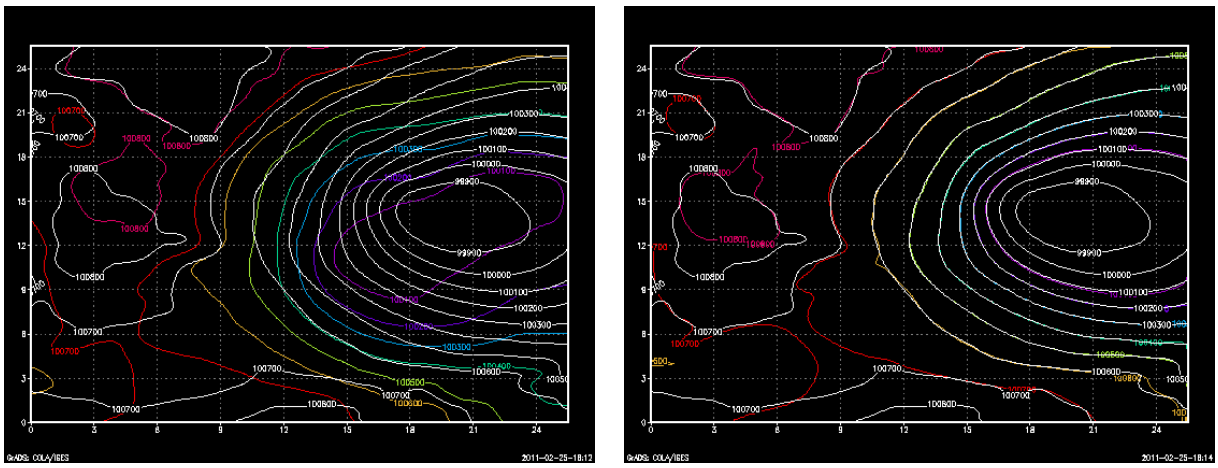

Figure 2: The position correction on the mslp fields. Left "fg" and "obs" superimposed before the alignment. Right: after the alignment. Note that the area displayed corresponds now to a sub-domain centred over the eastern low seen in fig 1.This sub-domain is 256 x 256 points (i.e., about 640 km) big and is referred to as "area of interest" AoI.

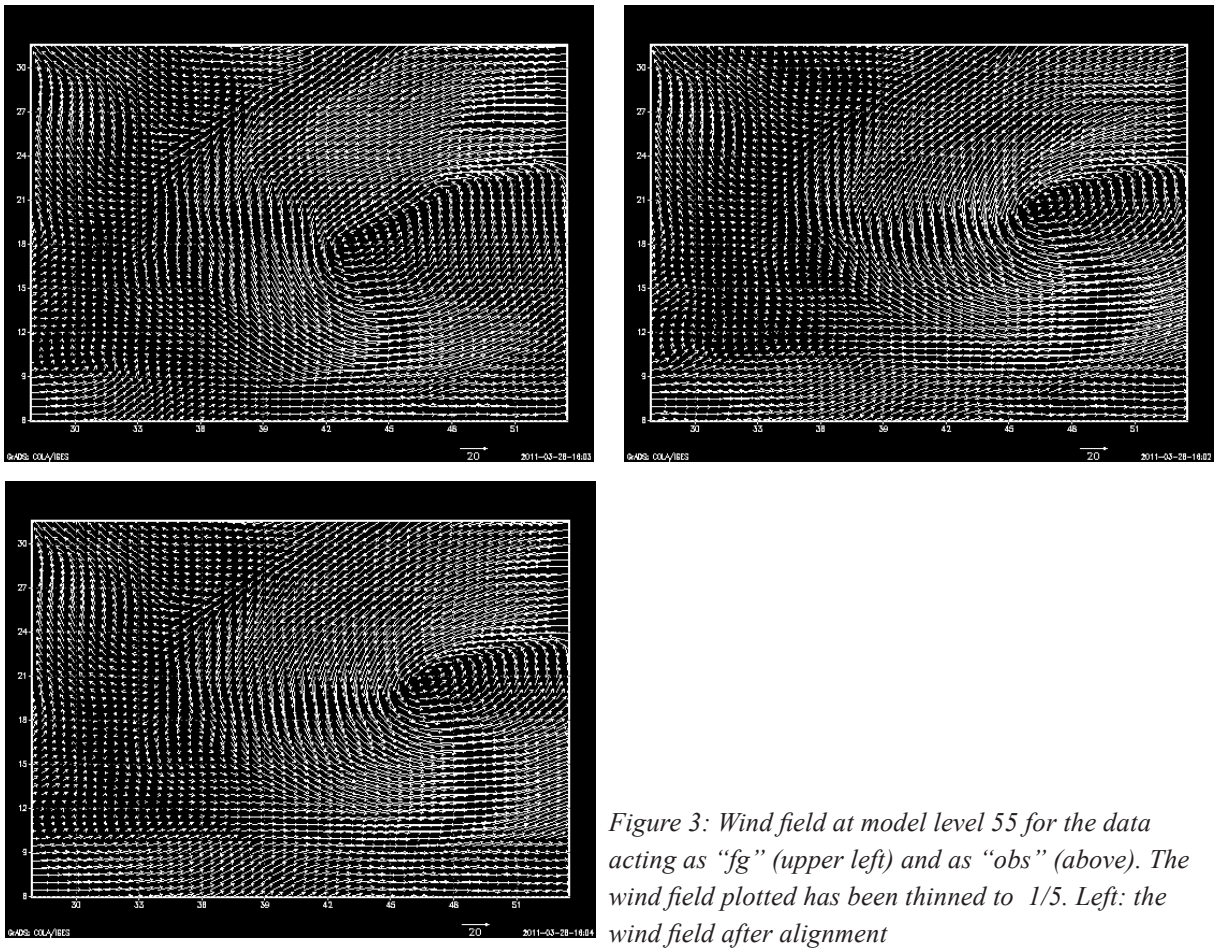

Figure 3: Wind field at model level 55 for the data acting as "fg" (upper left) and as "obs" (above). The wind field plotted has been thinned to 1/5. Left: the wind field after alignment

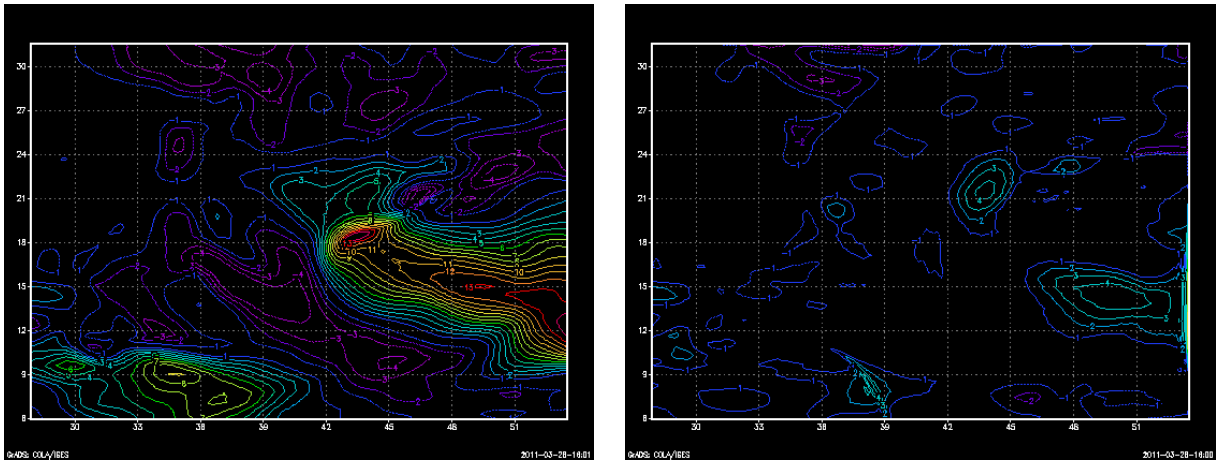

Figure 4: Left: Wind speed differences ("obs - fg") at model level 55. Right: differences after the alignment. Units m / s.

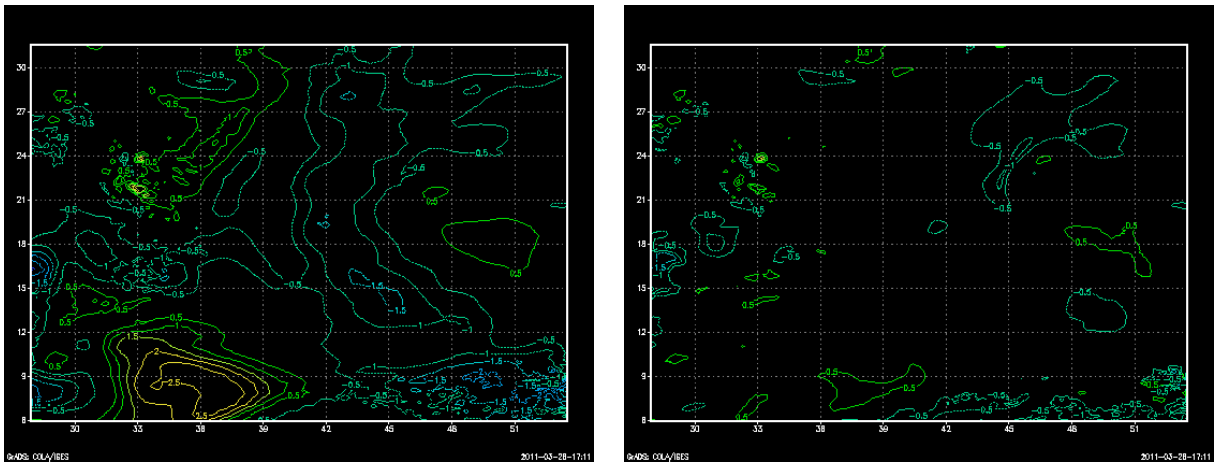

Figure 5: Left: T (K) differences ("obs – fg") at level 60. Right: differences after the alignment

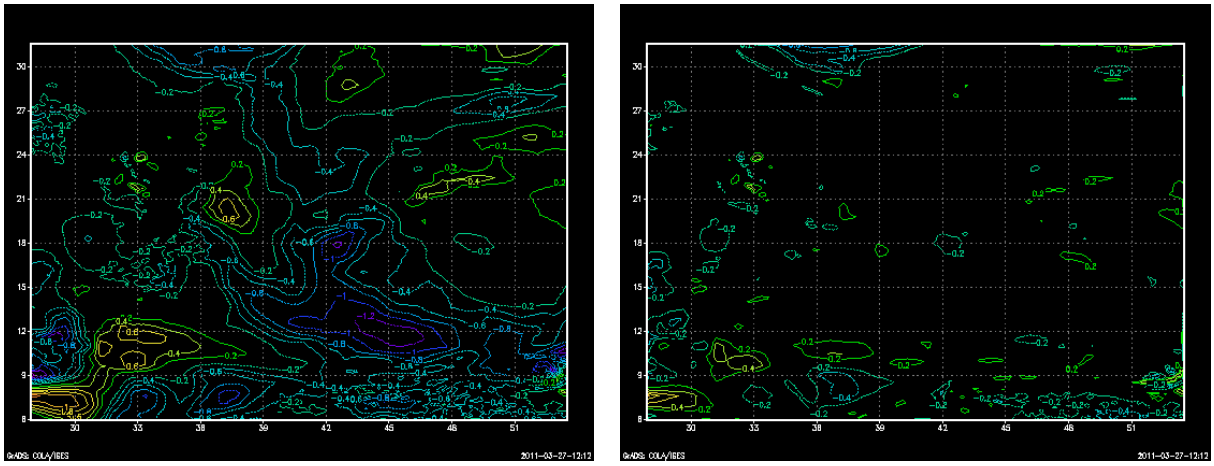

Figure 6: Left. q (gr/Kgr) differences ("obs – fg") at level 60. Right: differences after alignment.

3.2 Set of experiments and its realization

The tests have been designed to check the impact of the FA (Field Alignment) correction on the quality of forecasts. The work is executed according to a simple plan. The experiment consists of three runs. They all are +12 H forecasts, starting at 28th 00 UTC, with the same LBCs (3-hourly update cycle). The first run is the control run (CNTL) and it is generated with the first set of fields described in the previous section ("preparation of the data") as initial conditions. The second run ("SHFT") takes as initial conditions the second set of fields of the previous section. The upper-air fields are 3D-Var analysed in order to check to what extent a conventional analysis can handle the problem of position errors in this particular case. The observations, however, are not real observations but bogus observations from the fields acting here as "truth". The third run ("ALGN") takes the same initial conditions as the second run but, previously to the 3D-Var analysis, an alignment to correct for position errors is done. After this FA step, the upper fields are analysed as in the second run. No surface analysis is done in any case.

Working with field values as observation surrogates makes it easy to achieve a good sampling and have good control on the location of the observations. A fairly dense network of "stations" was defined over the AoI and surrounding zones (fig 7, left). Pseudo in-situ observations for ps, u, v, T and q were generated for each location marked on this figure. The sampling in the vertical is

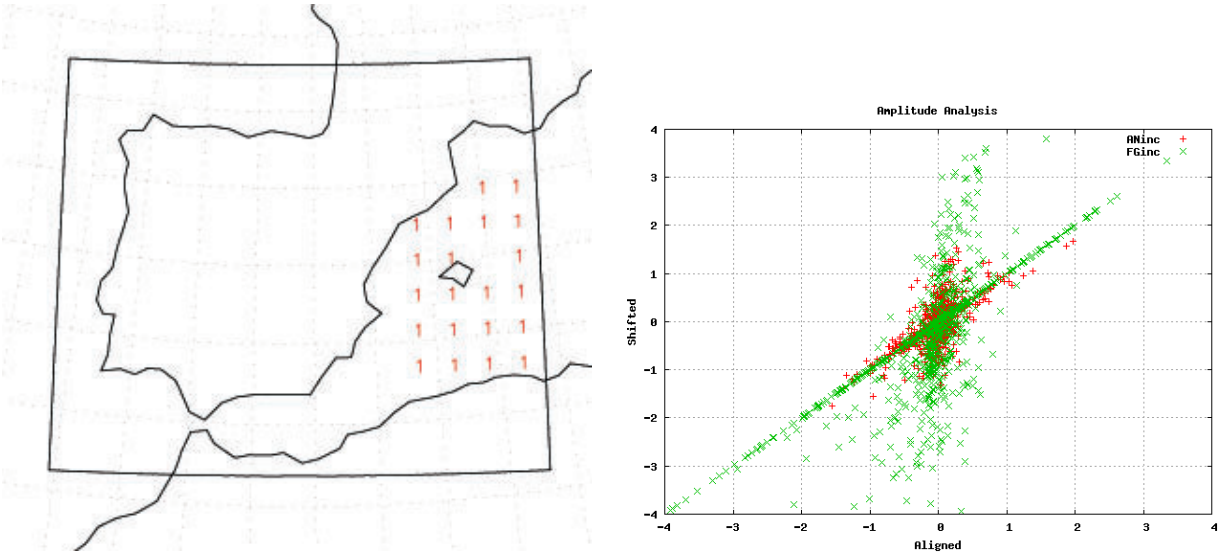

Figure 7: Left: Location of the "pseudo-stations" where profiles with 10 levels of "pseudo-obs" of u, v, T ,ps, and q are generated in order to carry out the amplitude correction step. Right: scatter plot "SHFT"(vertical) vs. "ALGN" (horizontal) for the σo normalized fg increments (green) and analysis residues (red).

10 levels evenly distributed between the surface and 200 hPa approximately. This made a set of 816 pseudo-observations in total. At this point one might raise the following objection. In order to perform the kind of FA correction done in these tests in a realistic situation, a much denser network of observations still would be required. To this objection one may argue that the FA correction can utilize observation sources that delineate acceptably well continuous patterns related to weather phenomena but that are not so good in terms of measuring absolute parameter values. That is, this technique could take advantage of synergies between remote sensing observations (good in pattern location) and in-situ observations (more accurate parameter measurements). One may as well try to go around this obstacle by using techniques that generate patterns from (dense) in-situ networks.

The alignment is done following the method outlined above in section 2. The method turned out to work very well. The fields used in these tests are all in model hybrid levels. Small scale structures induced by orography are clearly visible in these fields. This represents a handicap for the alignment algorithm because the alignment forcing term becomes very noisy. Moreover, it is clear that the modulation of some fields like, for instance, surface pressure, by orography, should not be displaced. For the ps field, this problem can be conveniently handled by using the mslp field as a proxy or substitute, that is, the alignment algorithm is applied on the mslp field rather than on the ps field. However, the scheme requires not only a solution to the alignment equation, it is also necessary to deform continuously the field to be aligned. This is done by using bi-cubic interpolation which is also sensitive to this small scale orography-induced structure. It was found that a more convenient way to work with the ps field is to extract the orographic component at the beginning of the process, align the mslp field and finally add the orographic correction on the solution. If * stands for "aligned", then this can be summarized as: $p_s^* - p_s = mslp^* - mslp$. The result for this case is excellent, as one can see in fig 2. The model isobars have been moved to match almost exactly the "observations". Note that the correction also works well over land (western and southern parts of the area, and the Balearic Islands in the centre). The low in the aligned field is shallower than it should be and this also illustrates perfectly the need for a correction in amplitude after the alignment. An issue that requires further investigation is how to ameliorate the discontinuities at the borders between the area where the alignment is carried out (i.e. AoI) and the rest of the model domain. These discontinuities are

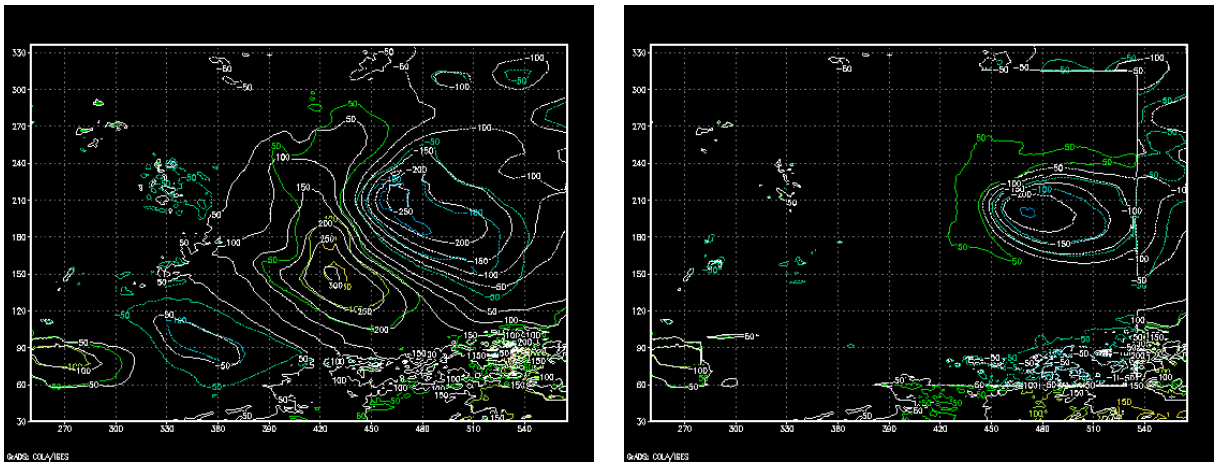

Figure 8: mslp fg increments before (white lines) and after (colour lines) the conventional 3D-Var assimilation for the "shifted" experiment (left) and "aligned" experiment (right). The area plotted comprises the "area of interest" and adjacent zones. The 3D-Var analysis for the "shifted" experiment does not change the phase of the wave-like pattern, only its amplitude. The correction in the "aligned" experiment is not optimally calibrated. Note the discontinuities at the borders of the AoI on the right plot.

clearly visible on the eastern part of the area (see fig 8 right). Quite surprisingly they did not have a clear detrimental impact on the forecasts with the aligned fields in these experiments.

The first tests where only the p_s field was submitted to the position errors correction process gave negative results because of a "rejection problem". The +1H forecasts from the "SHFT" and "ALGN" experiments were nearly identical. Then it was decided to process all remaining analyzed fields (u,v,T,q, i.e. 240 fields). It is of course unrealistic to pretend that one will be able to align all the fields in a real situation, but on the other hand, it is felt that surely positive results can be achieved with a reduced set of fields and that corrections at neighbouring levels will be correlated. All this however remains to be tested. In all these cases there is not a suitable "proxy" to overcome the difficulties originated by the orography. The adopted solution was to apply a low-pass filter to these fields and use these smoother fields as "proxies". The design of this filter was done quite carefully by comparing the spectra of the p_s and mslp fields over the AoI. It turned out that a zero-phase 51 point long filter with stop band at 32 Km works well for these experiments. In the same notation as for the ps field above, we have: $\chi^* - \chi = \mathrm{flt}(\chi)^* - \mathrm{flt}(\chi)$. Clearly, by proceeding in this way we leave out scales that surely should be displaced also. Implicitly it is assumed that they will be somehow dragged along with the larger scales with respect to which they are ancillary.

As was the case for the p_s field, the alignment algorithm worked satisfactorily for all the other processed fields. Figures 3, 4, 5 and 6 illustrate some cases for each analysis variable. The difference between the fields after the alignment (displayed on the right) is a small fraction of the difference prior to it (displayed on the left). Therefore we conclude that the FA algorithm succeeds in removing the position component of the error in all these cases.

Before the intial conditions for the SHFT and ALGN experiments are ready, we must still perform conventional 3D-Var analyses to account for the amplitude errors. This was done as outlined above. Some "ad-hoc" filters currently active in the system like the "reduction zone" to minimize wrap-around effects were switched off. The focus is on the impact on the forecast over the AoI and nearby zones and spurious increments on the opposite western part of the domain are of not concern here. The screening did not remove a single pseudo-observation, all 816 fall within tolerance limits.

The plot on the right of figure 7 is a scatter plot for the fg increments and analysis residues for both experiments. The green crosses correspond to fg increments. The points on the diagonal are associated with locations outside of the area where the FA correction was done, and therefore are equal in both cases. The other branch of the plot is almost vertical, reflecting the significant reduction in size of the fg increments by the FA step. The red crosses correspond to the analysis residues. Now the difference between both experiments is much reduced. Note however that in the analysis for the FA'd fields the calibration is not optimal because we did not generate a new set of estimates for the "sigmas" accounting for the minimization of position errors. Had we considered this before, the analysis residues for the ALGN experiment would have been smaller.

We see that, not surprisingly, the 3D-Var algorithm does not perform very well in correcting the position of the low. The "dipole-like" pattern in the mslp fg increment field is still apparent in the analysis although is somewhat flattened. Figure 8 (left) indicates that the "dipole-strength" change is from 5.5hPa to 2.5hPa approximately. Also the deepening of the low in the mslp fg increment field for the ALGN experiment is not intense enough (figure 8 right). As a consequence of the absence of a correction of the weight of the observation forcing in the analysis of the aligned field, the low appears still too shallow by about 1.5hPa.

3.3 Results of the experiments

The forecasts produced from both experiments SHFT and ALGN can be compared with the forecasts of the control run and in this way gauge the impact of the FA step. The results are very positive. During the first hours of the integration the impact on "weather parameters" like precipitation and wind near the surface is apparent. The impact dilutes afterwards, in this case after 4-5 hours in the wind field and 2-3 hours in the precipitation field. The range of persistence of the impact in these experiments is surely conditioned, not only by the size of the AoI, but also by the proximity of it to the eastern border of the HARMONIE IBERIA_2.5 domain and the circumstance that the cyclonic circulation causes in-flow from this side. Because of this the LBCs (which are identical for all experiments) take over relatively soon.

Figure 9 is a panel of 8 plots which show the impact on the wind speed at 30 meters above the ground during the first 4 hours of the forecast. Each row corresponds to a specific forecast range, at the top we have +1H and the last at the bottom is for +4H. The left column displays CNTL - SHFT differences and the right column CNTL – ALGN differences. We note that the "error" over the AoI in the aligned experiment for +1H is very small and that there is no indication of the presence of noise that could have been caused by imbalances in the initial conditions. We do see in +2H and after a suspicious wind streak stretching from Mallorca to the Menorca islands, but it is present too in the SHFT experiment and therefore we must conclude that it is not a spurious pattern originated by the FA correction, it must have come from outside of the AoI. The wind speed in SHFT is severely underestimated (up to more than 10 m/s) during the first hours of the forecast over a big area to the southwest of Mallorca island . The error pattern shape suggests that the flow in the CNTL experiment is enhanced by a prominent orographic feature on the northwest of the island: "Serra Tramuntana" (1400m amsl). In the SHFT experiment the shallower and displaced low induces a weaker circulation over the island and the orographic channelling is not so intense.

The last figure, figure 10, is another panel with similar layout but in this case for the parameter "simulated radar reflectivity". This parameter can be generated by the post-processing package of HARMONIE. It is calculated from the model concentrations for different hydrometeors (rain, graupel, snow, etc…). The fields plotted correspond to a level of 30 m above ground and are good proxies for precipitation. The units however cannot be easily translated into precipitation intensity. Differences are as before CNTL - SHFT (left) and CNTL – ALGN (right). We observe that the

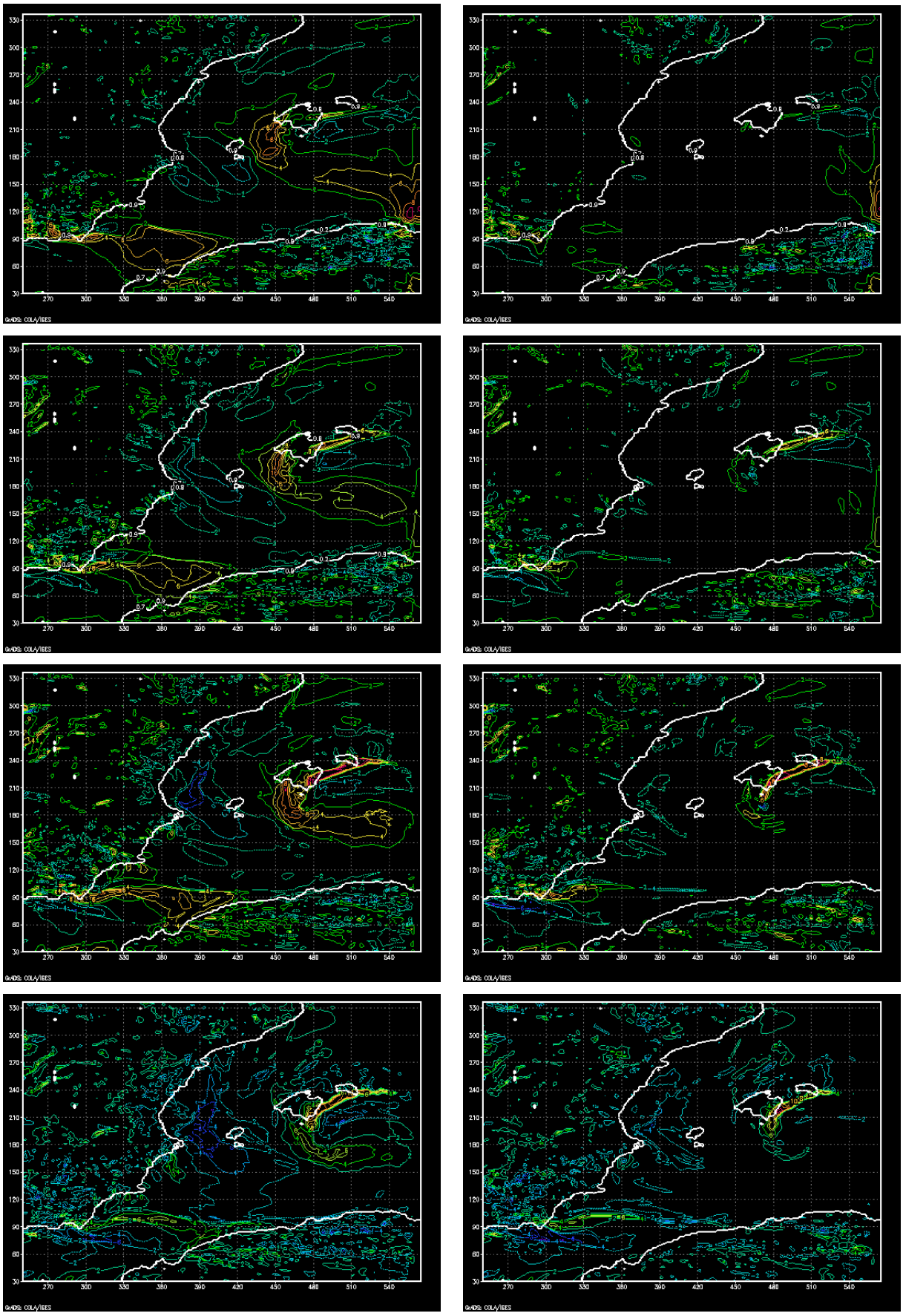

Figure 9: Test results on wind speed at 30 m above ground. First row +1H, last row +4H. Left shifted experiment, right aligned experiment. See text for details.

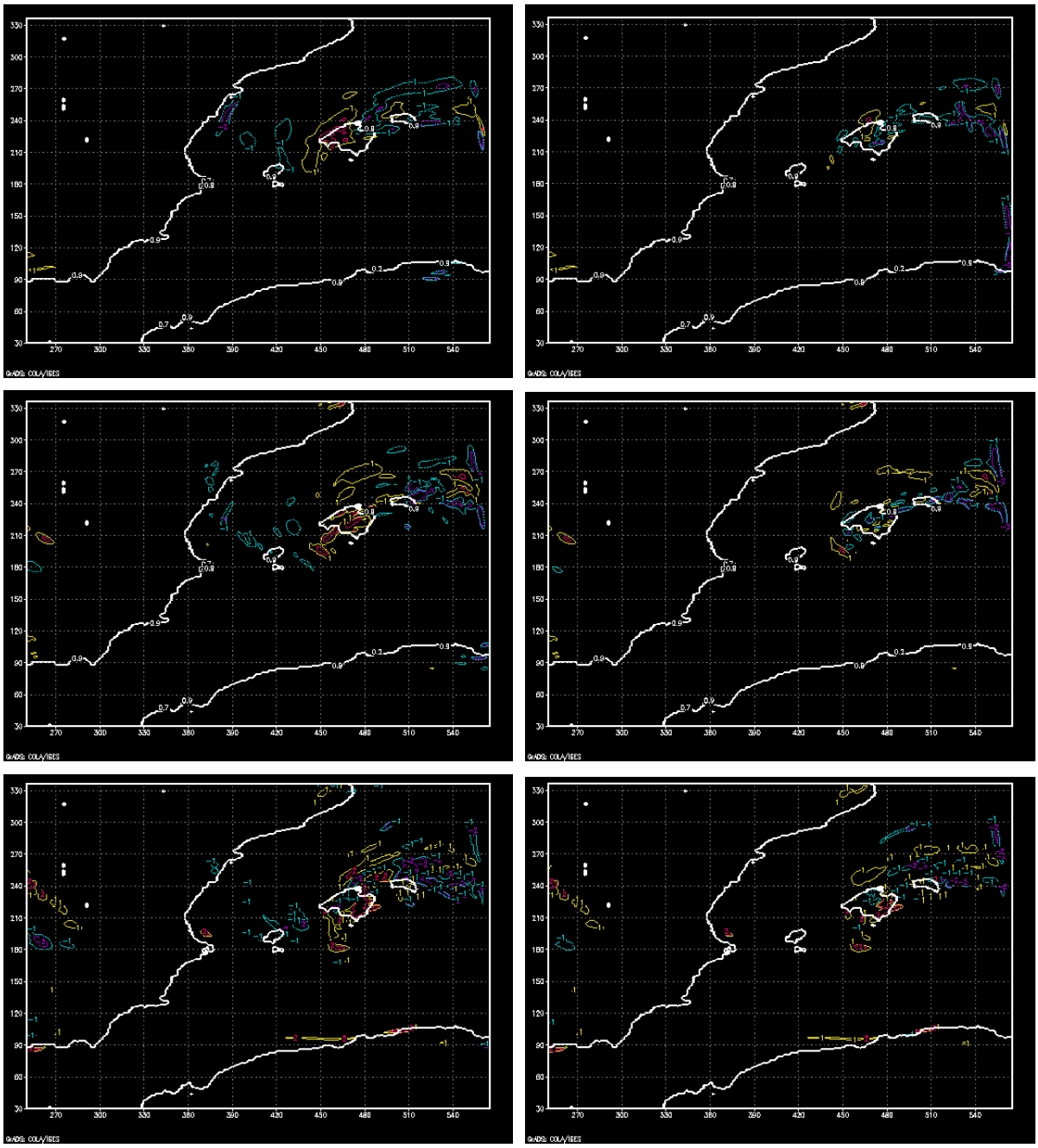

Figure 10: Test results on simulated radar reflectivity at 30 m above ground. First row snap-shot +1H, last row +3H. Left column: shifted exp errors. Right column: aligned exp errors. Red +2 Orange +1 Blue -1 Purple -2). (arbitrary units). See text for comments.

"snapshot" +1H (first row) indicates that precipitation is clearly underestimated in the SHFT experiment over the whole island of Mallorca while in the ALGN experiment the contours show both signs and smaller values. The error field for the ALGN experiment has a random aspect in contrast with that from the SHFT experiment. The situation at the "snapshot" +2H is very similar, while one hour later (+3H) we see that the errors from both experiments look equally erratic, an indication that the FA impact on the precipitation forecast has already vanished. We can therefore conclude that the timing error in the precipitation forecast over the AoI is significantly reduced in the ALG experiment when compared with the SHFT. A more precise quantitative evaluation of errors in rainfall has not been done.

4 Conclusions

The algorithm for correction of position errors briefly described in the first part of this article has been tested with model fields used as surrogates for observations. The test consisted of a set of runs with HARMONIE 36h1.3 at a resolution of 2.5 km and 60 vertical levels on a single case characterized by moderate weather activity over the western Mediterranean. The algorithm passed the tests successfully. The field alignment method proved to be effective in removing the position errors and improved significantly the forecasts as compared with other forecasts where the errors in the initial conditions were analysed only with a conventional 3D-Var algorithm.

During the realization of these experiments some ideas and methods were tested. Mathematical methods to solve the alignment equation, global or local alignment, the pre-processing of the data to handle orography, and the utilization of proxies for the alignment of the wind field (wind speed, divergence, vorticity) are some examples. The experience gained will be useful in the future.

Future work will focus on testing the methods with real data. A number of difficulties are foreseen as model fields, in many respects, only vaguely resemble observations. Problems related to the quality of the data, incomplete data, lack of correspondence between model and observations, etc. will have to be analysed. Nonetheless, it is felt that this approach has great potential and the encouraging results presented in this communication support this expectation.

Acknowledgment

Many thanks to Tilly Driesenaar who took the work of improving the readability of this HIRLAM Newsletter.