Estudios de impacto de observaciones satelitales con el algoritmo de asimilación HIRLAM-4DVar

SATELLITE OBSERVATION IMPACT STUDIES WITH THE HIRLAM 4D-VAR ASSIMILATION ALGORITHM

C. Geijo(1), B.Amstrup(2), J.de Vries(3),N. Gustafsson, P.Dahlgren(4), O.Vignes(5), E.Whelan(6) y X.Yang(2)

(1) Agencia Estatal de Meteorología, AEMET (España), (2) Instituto Meteorológico Danés, DMI. (3) Real Instituto Meteorológico de los Países Bajos, KNMI. (4) Instituto Meteorológico e Hidrológico de Suecia, SMHI. (5) Servicio Meteorológico Noruego (Met.no). (6) Servicio Meteorológico de Irlanda (Met Éireann).

SUMMARY

We have carried out a number of observation impact studies with the aim to determine the individual and joint contributions of different space-based observing systems on the skill of short-range forecasts (up to 48 hours) over the Northern Atlantic and European regions. These studies are an essential part of the current HIRLAM data assimilation plan to include in the near future a bigger set of satellite observations than currently used in the operational runs of the different meteorological services within the HIRLAM consortium. In our experiments with data from up to 8 different satellites and the HIRLAM 4D-Var scheme running with a 6-hours long assimilation window, we have detected a significant positive impact on the forecast skill. The satellite observations not only help to better characterize the initial state, they reduce the mean size of increments and so contribute to filter numerical noise generated by the advection scheme in the assimilation algorithm. Analysing the geographical distribution of this impact, we have found some dependency on the type of observation in line with expectations, but we have ascertained too a good level of synergism among them. We recognise the need to consider more experiments with longer periods before more sound conclusions can be drawn and we are working on them.

Introduction

Nowadays it is a well established fact that use of satellite observations has been a major factor in the verified improvement of weather forecast quality. On the other hand, their utilization is hindered by difficulties related to their bigger complexity as compared to other meteorological (conventional) observations that have been assimilated for a long time. This is maybe the first reason why their assimilation in NWP centres provided with, in relative terms, fewer resources is behind schedule. In the HIRLAM data assimilation group we think that the situation must go improving and this work has been motivated with the aim at intensifying the operational use of satellite observations at the meteorological centres within the HIRLAM consortium.

1. Experiments description

1.1 Used Satellite Data

The satellite data used in these experiments are: **(a)** temperature and humidity microwave sounders AMSU-A/B and MHS onboard the satellites NOAA-15, 16, 17 and 18. For the NOAA-17 only AMSU-B since AMSU-A failed in October 2003. **(b)** Wind data (AMV) from SEVIRI (M-9) and MODIS (EOSI-II) imagery, and **(c)** near sea surface winds from SEAWINDS scatterometer embarked on QuikScat.

(a) Radiances measured by AMSU-A/B and MHS are converted to temperature and humidity analysis control variables with the use of the radiation transfer algorithm RTTOV-8. This software, developed by NWP-SAF, includes a sea-surface emissivity model (FASTEM). The emissivity of the other two types of surface considered here: land and ice (or Lithosphere and Cryosphere), is set to fixed pre-defined values. Because the highest level of the HIRLAM-60L model is 10 hPa, the AMSU-A "stratospheric channels" (11 to 14) were disregarded. In order to minimize errors stemming from our insufficient knowledge of the emissivity of the surface in the range of the GHz in which these instruments work, those channels with biggest contributions from the surface (i.e., 1 to 3) were equally discarded. For surface types different from open sea, and for the same reason, channels 4 to 6 (land) and 4 to 5 (ice) were not assimilated. For humidity sounders AMSU-B and MHS, the upper top at 10hPa is not a problem (the tiny amount of water vapour above this level plays no role in NWP) but it is still a reason of concern the inaccurate knowledge of surface microwave emission properties. Out of the five channels that these instruments have, just those from 3 to 5 were actually assimilated and this only over sea surface. Although all the discarded channels do not force the analysis some of them are used in quality control checks that filter erroneous data or measurements gained under circumstances impossible to be correctly simulated by RTTOV-8 (e.g. precipitation or thick cloudiness).

Since the EARS-ATOVs infrastructure was put in place on both shores of the Northern Atlantic, ATOVs data at that spatial scale can be easily received routinely through the broadcast service EUMETCAST that is operated by EUMETSAT. The timeliness for these data is adequate for regional models that have cut-off times as short as two-three hours. If the direct reception via local equipment is preferred, the timeliness is still better but at the expense of a reduced spatial coverage. In this case it is necessary also to carry out some important pre-processing locally.

(b) As it is well known, the “Atmospheric Motion Vectors” generated from SEVIRI images (SEVIRI-AMV) measure the mean wind over a time interval of one hour at a spatial resolution of about 70 Km. In the case of the so-called “High Resolution Vectors” this scale becomes somewhat smaller: 32 Km. The time resolution can also vary. For example, in the product “Rapid Scan AMVs” the averaging time interval shortens to ten of minutes or so. The specific characteristics of each type of vector, that is, space-time resolution, imagery used for the tracking, kind of tracer employed, kind of height assignment algorithm used, etc… can be easily red off the BUFR messages in which these data are distributed. The AMVs utilized in these experiments were the “standard” vectors produced by MPEF facility and hourly broadcasted via GTS and EUMETCAST.

MODIS-AMV data included in our experiments were generated at NOAA-NESDIS by using MODIS images gained in successive overlapping passes of Aqua and Terra satellites over the North Pole (also known as ES-I and EOS-II). By alternating images from one and the other spacecraft, it is possible to produce wind fields north of 60N with a periodicity of about 45 minutes. These vectors measure the mean wind at a time scale of two hours or so (three consecutive images are required) and at a spatial scale of about 60 Km, that is, comparable to SEVIRI-AMV data. The availability on time of these data for NWP regional models is more difficult to achieve than for AMV from geo-stationary orbit. The reason is the requirement of more complex on-ground infrastructure because these data are gained from polar orbit. During the last years important progress has been made in this crucial point and the timeliness is nowadays, thanks to the beginning of operations of the DB service (Direct Broadcast Service) in 2007, less than three hours. It is however necessary to warn at this point, that these experiments did not use DB data, but NOAA-NESDIS data which has a somewhat longer latency time. For more information on this issue, consult http://stratus.ssec.wisc.edu or the IWWG (International Winds Working Group) web-site http://cimss.ssec.wisc.edu/iwwg/iwwg.html. In the latter case more information on geo-stationary AMV data and wind scatterometer data is also available.

(c) SEAWINDS echoes scattered off the sea surface by roughness at the centimetre scale can be converted to wind at 10 meters above the surface. This conversion is what the SDP facility (SEAWINDS Data Processor) located at KNMI in the Netherlands does. These wind data are distributed also through EUMETCAST, although the data can be received with a shorter time delay if the ftp server that the KNMI has installed for this purpose is used. The baseline product has a spatial resolution of 25 Km. However, we have experimented with a product at coarser resolution (100 Km) because it has better S/N ratio. The frequency of availability for these data is in principle that of a Sun-synchronous orbit (i.e., about 90 minutes) but actually this is only possible for high latitude areas. Over the area of interest for this work, the spatial coverage and the time frequency vary through the day as determined by the QuikScat orbit parameters. This situation could be very much improved if other scatterometers already taking data were used as well (e.g., ASCAT).

SEAWINDS data were assimilated without bias correction and without first-guess check. The quality control procedure for these wind data relays heavily on internal processing carried out by SDP. It detects situations prone to measurement error (precipitation, presence of sea-ice, etc…) and flags them accordingly. For SEAWINDS, the Bayesian algorithm VarQC is very important because the wind direction ambiguity removal problem is solved at this stage in the assimilation. The wind direction ambiguity arises because SEAWINDS can not determine uniquely this parameter. Depending on the position of the measurement cell in the SEAWINDS swath geometry, SDP can return up to four directions or ambiguities. SDP provides too with a “solution probability” parameter which is utilized by VarQC during the analysis to resolve the ambiguity.

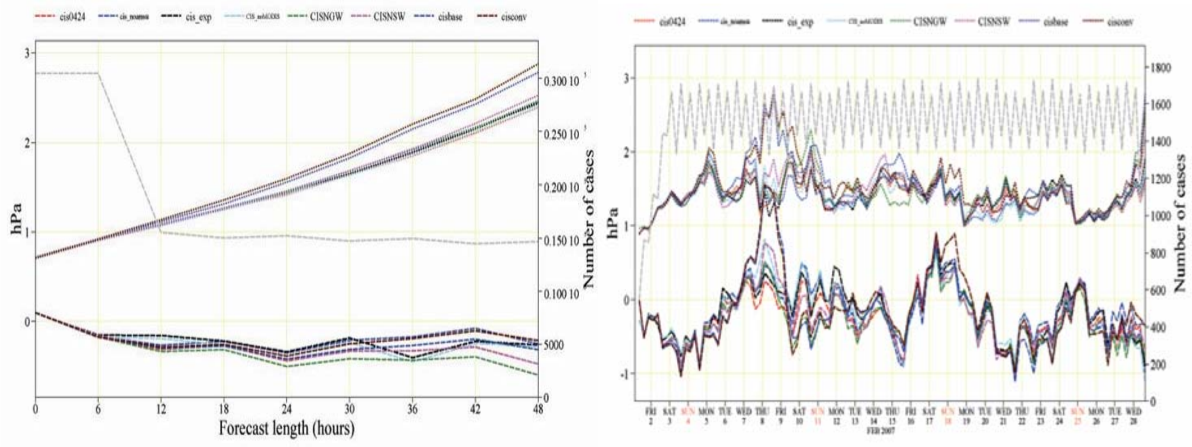

Figure1 Ps parameter verification for the EWGLAM surface station list in February 2007. Left error growth curves, right error time series.

.

1.2 The DA Algorithm

The data assimilation algorithm used in these experiments is the HIRLAM 4D-Var algorithm which has been developed by Nils Gustafsson et al. In some HIRLAM NWP centres (e.g., SMHI in Sweden), this algorithm is in routine operations since beginning of 2008. Very schematically, the algorithm produces an analysis by minimizing a "cost function" which consists of several terms:

$$J = J_b + \Sigma_i (J_o)_i + J_c$$

The J_b term is the "background constraint term". This term is needed to state the problem correctly. Given that the number of unknowns (the analysis space dimension) is very much bigger than the number of data (the number of observations), it is required to introduce constraints or additional conditions in order to be able to get to a useful solution. J_b is built with statistical regressions among the analysis control variables and represented in the wave-number space in a totally similar way to the 3DVar algorithm J_b term. Σ_i $(J_o)_i$ is the "observation forcing term". The sub-index "i" refers to the time slot within the DA time window at which the corresponding set of observations belong. This term forces the first-guess with the information brought into the weather forecasting system by the observations. J_c is the "penalty term" that allows for a better balanced analysis by filtering out high frequency spurious oscillations ("gravity waves"). The control space consists of horizontal wind, air temperature, surface pressure and air humidity (q), all of them at the beginning of the DA time window.

At difference with an algorithm like 3D-Var, 4D-Var propagates in time the observation increments or innovations through the data assimilation window by using the "adjoint technique". The process is a multi-loop iterative one. An "inner-loop" searches the minimum of the J function by using the tangent and adjoint (TL/AD) models. The integrations of these models are carried out with simplified physics (e.g., no representation of the hydrological cycle) and at coarser space-time resolution than the forecast model. An "outer-loop" updates the baseline trajectory calculated with the forecast model at the beginning of the process and after the inner loop is completed. The initial configuration that we tried consisted of one single iteration in the outer loop (i.e., the baseline trajectory is refreshed just once at the beginning of the process with the most recent boundary conditions) and 60 iterations at 1/3 x-t resolution in the inner loop. In later trials, we checked other configurations (see below). The computation time required by this algorithm with this configuration is about the same as a 48 hours forecast, that is, about 4-5 times as much as that required by a 3D-Var algorithm.

1.3 Experiments Configurations

The spatial domain chosen for these experiments comprises the Northern Atlantic, Europe and parts of Northern Africa and the Northern Polar Cap (fig2). The spatial resolution of the forecasts is 0.15 degrees and the vertical is resolved in 60 levels. The model is hydrostatic. The boundary conditions were taken from forecasts produced by the global ECMWF deterministic model and refreshed every three hours. Predictions with 48 leading time were launched twice a day at 00 and 12 UTC and shorter predictions of just six hours leading time at 06 and 18 UTC in order not to break the assimilation-forecast six hours cycle. The DA time window is six hours long, from -3h30 minutes to +2h30 minutes with respect to this analysis nominal time.

The control experiment uses conventional observations together with a reduced number of satellite observations, namely, AMSU-A data over

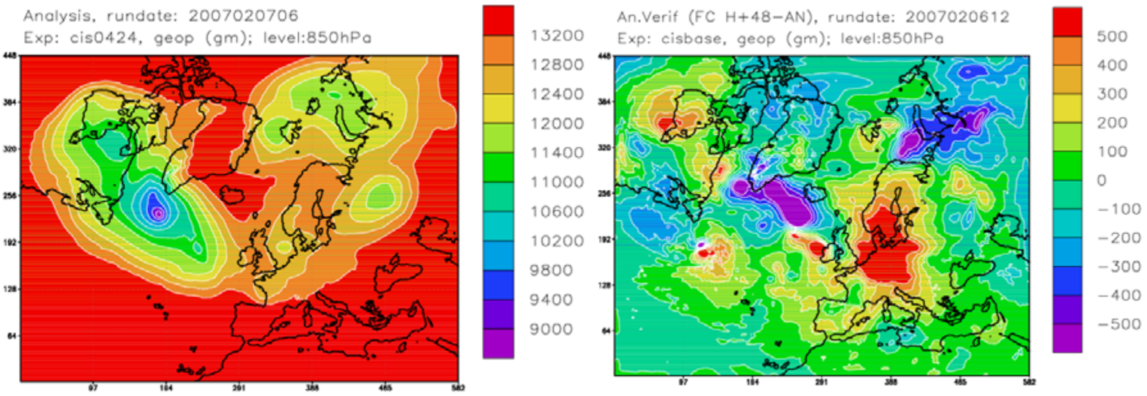

Figure2 Φ analysis at 850 hPa on 2007020706 for the "all-included" experiment (left) and difference map with respect to the control after an integration 48 hours long. Two small scale structures visible off the coasts of Ireland and Denmark, which were not present in the control, are responsible for the spectacular growth that can be seen on the figure on the right (see text for more details).

open sea from NOAA-15 and NOAA-16. This is the default configuration for the HIRLAM 7.2 version and lower. The experimental runs use the same set of conventional observations as the control and all the satellite observations presented above. It is referred to as "all included".

Initially two periods of time were chosen to experiment with. One typical "summer case" (August 2006) and another "winter case" (February 2007). In the later case we found a very significant positive impact in the experiment (see fig1 and below for the discussion on the results) and it was then decided to extend the tests by including several denial experiments, that is, experiments in which one observation is withheld while the rest are assimilated normally, in order to identify the role played by each type of satellite data in this case. This means that five different denial experiments were carried out: "no-AMSU-A", "no-AMSU-B/MHS", "no MODIS-AMV", "no SEVIRI-AMV", and finally "no SEAWINDS".

2. Results Presentation and Discussion

In this section we will focus for more clarity and brevity on parameters related to the mass-field, that is, surface pressure and geo-potential height. In a hydrostatic model as the one used in this work, these parameters are calculated by means of sums on the vertical levels as indicated in the "tendency equation" for Ps, and the "hydrostatic equation" for Φ parameter. This can explain why the verification results for these parameters in experiments of this kind look usually more consistent and less noisy than those for other, single level, parameters (e.g. wind at 500 hPa).

In the first place **(a)** results at "European scale" are presented. These were obtained from observations coming from the EWGLAM list of stations during February 2007. At this spatial scale, the number of sonde ascents during one month is large enough to extract significant results for the upper-air verification. Second **(b)** results at a regional scale are presented. In this case only surface verification results are discussed. Third **(c)** we finish with the presentation of results obtained in a second batch of experiments (in this case only for control and all-included) that were carried out after numerical instabilities in the analyses were detected which pointed out the convenience of choosing different algorithm settings.

(a) Surface and Upper-Air Verification at European Scale. Figure 1 on the left shows Ps verification results for February 2007 for the whole set of EWGLAM surface stations. It can be seen very clearly that the error curve for the control experiment (and another curve which assimilated just conventional observations) grows at a faster rate than that corresponding to the experiments in which satellite observations were assimilated, jointly ("all-included") or discriminating among them ("denial experiments"). The difference at +48 Hours is about 0.5 hPa, or in relative terms about 18% of the total error in the control. Each point on these curves was calculated with about 15000 observations. We see that in this averaged verification it is virtually impossible to attribute the positive impact to any particular type of observation because all the curves are very close to each other, with maybe a very slight advantage for SEAWINDS data. We do not see here problems of interference among the observations, that is, the "all-included" line is not above any of the curves for the different denial experiments. The positive impact is concentrated around days 8 and 9 of the month as it is shown on the error time series on the right graph of figure 1. Upper-air verification gives results (not shown) in good agreement with the surface results. The error time series for Φ at 850, 500 and 250 hPa display a pronounced impact centred around those days too.

A more detailed study of what occurred those dates reveals that, indeed, the initial conditions for control

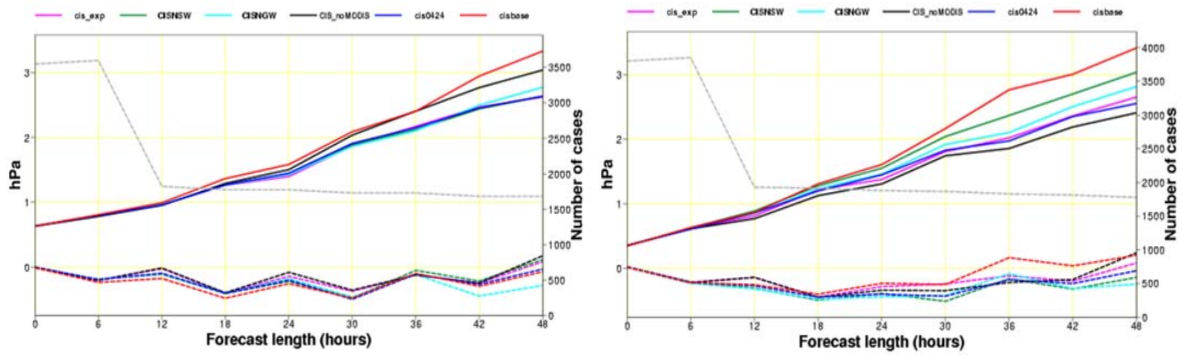

Figure3. Ps Error-forecast leading time curves for Finland (left) and for the Netherlands (right)

and experiments differ in small scale structures over the Atlantic off the coasts of Ireland and Denmark. These differences undergo explosive growth in 48 hours (figure2). It seems then reasonable to hypothesize that the intensive use of satellite observations allowed the analysis to specify in bigger detail and accuracy the initial state and in consequence generate better forecasts. However, as it is discussed after, the positive effect of these observations seems to have materialized through a process somewhat more subtle than just the mere detection of unstable baroclinic structures which went unnoticed for the set of conventional observations due to their lower spatial and time resolutions.

(b) Surface Verification at Regional Scale. If we carry out Ps verification segmented by European regions, we find that it is possible to identify which type of observations has relatively bigger positive impact for each area. For instance, over Finland (fig3, left) we see that the black curve ("no MODIS-AMV") lies very near to the control curve, that is, these data explain most of the measured positive impact. The deep blue curve corresponding to the "all-included" experiment is below the rest, i.e. we confirm that there is synergy among the different types of observations. If we now travel a bit to the South, to the Netherlands (fig3, right), we encounter that the error-curves graph has a different aspect. The green line ("no SEAWINDS") is now the one which contributes most to the positive impact, followed by the SEVIRI-AMV data. In this plot it is possible to see as well that the MODIS-AMV data have had a *negative* impact contribution, that is, we lose the synergy in this case.

This local negative effect of MODIS-AMV data, which is detected in the results for Ireland and UK areas as well and much less pronouncedly in the upper-air verification (not shown), and the overall weak impact of ATOVs data, must be attributed to a non-optimal calibration of the algorithm. There are different techniques to adaptively tune the assimilation algorithm. These techniques are based on the agreement between actual and expected values of several important assimilation parameters like prescribed errors or "sigmas", or the cost function or some of its terms. Following this methodology it was possible to correct the deficiencies in the algorithm calibration.

It is found that when one is concerned with just one single obs type in observation use experiments the magnitude of the effect produced by fine tuning is usually "second order" when compared with the magnitude of the effect of assimilating or not that observation type. But as we see in these experiments, the fine tuning of the algorithm becomes more meaningful when a bigger set of observations is considered.

(c) Optimization of HIRLAM 4D-Var Configuration Settings. That most of the impact concentrates in a few events, as it is shown by figure1 (right), it is something unexpected. Pronounced failures in forecasts are not usual for systems that use conventional observations. On the other hand, the 4D-Var algorithm advects observation increments by means of semi-lagrangian schemes whose stability is sensitive to factors like the amplitude of these increments or the relation between space-time resolutions. It is clear that this dynamical aspect of the algorithm makes it more complex than "static" algorithms like OI or 3D-Var.

A detailed analysis of the origin of the differences that appear on figure 2 unveiled the presence of noise in the analyses of day 5 of the month (not shown) as their cause. This is about 48 hours before those differences on figure 2 showed up. The auto-regressive nature of the analysis-forecast process makes it easy to understand how this accumulation of errors can happen.

In light of this finding, it was decided to carry out a second batch of experiments with a algorithm configuration more effective in filtering out this sort

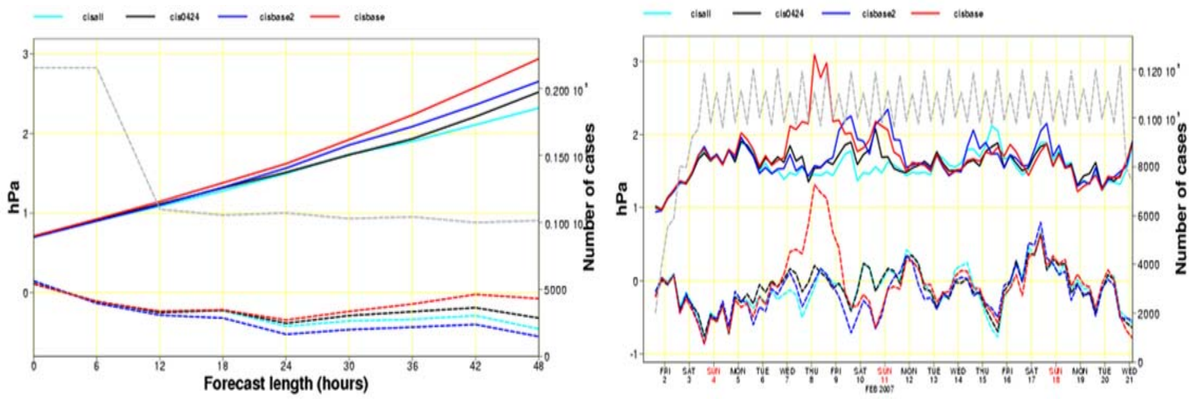

Figure4 Error growth curves (left) and error time series (right) for the experiments as commented in the text: “(c) Optimization of HIRLAM 4D-Var Configuration Settings”. See text for the details

of noise. In particular, it was decided to substitute the one single “outer-loop” configuration at 45 Km for a dual configuration with increasing resolutions 90 and 45 Km respectively.

The verification results can be found in figure 4. On the left we can see that the dual configuration improves results for both the control and the “al-included” cases, actually more for the control (compare red and deep blue curves) and somewhat less for the “all included” (compare black and cyan curves). It is important to realize that the use of satellite observations is clearly beneficial in any configuration: single outer-loop (red and black) and dual (deep blue and cyan). In the logic that the dual configuration removes the generation of noise, the difference between these two lines (deep blue and cyan) would give the size of the improvement attributable only to satellite observations. This is maybe slightly smaller than the detected in the first batch of experiments yet very important. Even more, the positive contribution of the observations is bigger than the one achieved by solely changing the algorithm configuration (deep blue and black). The right of figure 4 shows the time series for the errors of the four experiments. Having in mind the previous discussion, it can not come as a surprise that there is no trace of the pronounced failure on days 8 and 9 in any of the experiments that used dual configuration. The positive impact is now more evenly distributed in time.

3. Conclusions

In this work we have experimented with 4D-Var assimilation in limited area of a relatively wide class of satellite observations. Although the period of experimentation has been short, we have been able to confirm a significant positive impact on the quality of forecasts that use these observations as compared with those that do not include them in their analyses. By testing with different algorithm configurations, we have come to the conclusion that this positive impact is generated through two different ways. One, what we could call “direct way”, the atmospheric state is better specified thanks to many more observations. Two, say the “indirect way”, the 4D-Var algorithm is less prone to generate noise when the density and frequency of observations is increased. This is so surely because in the scenario of satellite observations being assimilated the mean size of the increments is reduced.

When the results are looked at an “European scale” no observation shows a clear advantage over the others, but when the results are analyzed at regional scale it is found a certain dependency of the size of the impact with observation type which, on the other hand, is in total agreement to what one can expect. In most cases observations contribute coherently to the positive impact. In those situations where we did not find synergy, we have checked that better fine tuning of the algorithm can reduce most of the interferences between observations. Other aspect that is worth recalling is the importance of experimenting with atmospheric situations characterized by its dynamism. The lack of significant results over the Iberian Peninsula is most probably due to the fact that the “typical winter period” chosen was not very interesting from the meteorological point of view over this area.

The period of experimentation is probably too short to draw sound conclusions. However, the results obtained so far should encourage the continuation of this work. In particular, extension of the experiments to other periods and may be to more observation types are options with an interesting outlook.